

申 报	系列：教师系列教学 为主型
	专业：计算机科学与 技术
	职称：教授

业绩成果材料

(申报人的业绩成果材料包括论文、科研项目、获奖以及其他成果等)

单 位 (二级单位) : 数学与信息学院

姓 名 王金凤

材料核对人:

单位盖章:

核对时间:

华南农业大学制

目 录

一、教学研究业绩

- 1.教学研究项目：关于教育部协同育人项目-基于鲲鹏的《操作系统》课程实践教学建设的立项通知及结题证书.....9
- 2.教学研究项目：关于教育部协同育人项目新一代软件工程教材建设的立项及结题证书..... 11
- 3.教学研究项目：关于广东省高校教学改革项目的立项通知及有关佐证材料..... 13
- 4.教学研究项目：关于广东省在线开放课程的立项通知及有关佐证材料.....16
- 5.教学研究项目：关于华南农业大学质量工程项目的立项通知及有关佐证材料..... 20
- 6.教学研究项目：关于华南农业大学一流课程的立项通知及有关佐证材料..... 23
- 7.教学研究项目：关于华南农业大学计算机系统类课程教研室的立项通知及有关佐证材料..... 27
- 8.教学研究项目：关于华南农业大学教学改革项目“翻转云课堂”在《操作系统》教学中的创新应用研究”的立项通知及有关佐证材料..... 29
- 9.教学研究项目：关于华南农业大学教学改革项目“基于创新能力培养的专业实践课中混合教学模式的探索与实践--以《操作系统》实践课为例”的立项通知及有关佐证材料.....30
- 10.教学研究项目：关于华南农业大学教学改革重点项目“工程认证背景下基于知识图谱的新工科一流专业建设和提升的研究与实践”的立项通知及有关佐证材料..... 32
- 11.教改论文：EduLGCL: Local-global contrastive learning model for education recommendation..... 33

12.教改论文：计算机类专业课程思政探索与实践—以操作系统课程为例.....	47
13.教改论文:工程认证背景下软件工程专业实训模式探索..	56
14.教学成果奖证书.....	64
15.国家级一流本科课程证书.....	64
16.广东省一流本科课程证书.....	65
17.教学比赛证书.....	65
18.主编教材：《操作系统原理-以 openEuler 为例》	66

二、科研项目

1.主持：关于风湿云平台（FSC）关键技术开发.....	71
2.主持：关于基于时空大数据的活鸡智能配送与调度系统关键技术研究及示范应用项目的合同.....	83
3.主持：关于基于大数据挖掘的农村耕地评价及智能决策系统研究项目的立项通知与示范项目的合同.....	98
4.主持：关于面向病毒基因大数据分析的若干问题及关键技术研究项目的立项通知.....	102
5.主持：关于面向多模态抗疫中医药功能大数据的“天盾”平台构建及示范应用项目的立项通知和合同.....	120
6.主持：关于面向温氏活鸡销售的智能调度模型研究与开发项目的合同.....	126
7.主持：关于恩平禽畜薯现代循环经济智慧示范园可行性报告项目的合同.....	136
8.主持：关于基于机器视觉的汽车 4S 店智能管理系统研发项目的合同.....	144
9.主参：高等农业院校中英联合办学项目材料.....	147
10.主参：关于面向多组学数据的高阶融合网络建模项目的立项通知.....	149

11.主参：关于渔业大数据系统标准化体系研究项目的立项通知.....	160
12.主参：关于基于移动式平台的水下场景立体重构与目标识别项目的立项通知.....	162

三、论文、著作等

1.检索证明.....	174
2.以第一作者发表本专业论文情况	
2.1. A calibrated SVM based on weighted smooth $GL(1/2)$ for Alzheimer's disease prediction.....	175
2.2. Parallel computing of fuzzy integrals: Performance and test	191
2.3. Fuzzy measure with regularization for gene selection and cancer prediction	214
2.4. Driver Fatigue Detection Using Improved Deep Learning and Personalized Framework	233
2.5. FF-GLAM-cs: a fusion framework based on GLAM with channel shuffle for speech emotion recognition.....	266
3.以通讯作者发表本专业论文情况	
3.1. EduLGCL: Local-global contrastive learning model for education recommendation.....	279
3.2. DFCCNet: A Dense Flock of Chickens Counting Network Based on Density Map Regression.....	293
3.3. DDTCN: Decomposed dimension time-domain convolutional neural network along spatial dimensions for multiple long-term series forecasting	313
3.4. A simple self-supervised learning framework with patch-based data augmentation in diagnosis of Alzheimer's	

disease	334
3.5. 基于聚类框架与局部感受野的实时人脸疲劳检测	347
3.6. 融入平滑组稀疏化的脑部 MRI 图像分类	354
3.7. 并行计算和稀疏存储在模糊积分上的应用	370
3.8. 基于蚁群算法求解 Choquet 模糊积分模型	378

四、科研成果

1. 知识产权

1.1. 专利授权证书：一种基于深度学习的司机状态实时监测方法.....	388
1.2. 专利授权证书：基于多任务深度学习的情感识别方法、系统及存储介质.....	390
1.3. 专利授权证书：一种基于多模态物联网的家禽养殖智慧监测方法及系统.....	392
1.4. 专利授权证书：一种基于大数据的企业外贸风险预警方法及系统.....	393
1.5. 软件著作权：基于微信小程序的会议室预约系统 V1.0	395
1.6. 软件著作权：基于微信小程序的华农智能导航系统 V1.0	396
1.7. 软件著作权：基于深度学习的鸡群跟踪计数系统[简称：鸡群计数系统]V1.0.....	397
1.8. 软件著作权：基于全局规划的禽类配送站点智能排班系统[简称：禽类配送站点智能排班系统]V1.0.....	398
1.9. 软件著作权：操作系统辅助教学系统 V1.0.....	399
1.10. 软件著作权：货物优化运输与排班系统 V1.0.....	400
1.11. 软件著作权：基于深度学习的人脸疲劳识别系统 V1.0	401

1.12. 软件著作权：美丽华农导航系统[简称：华农导游系统]V1.0	402
1.13. 软件著作权：基于深度学习的鱼群密度检测系统 V1.0	403
1.14. 软件著作权：基于深度学习的司机安全带检测系统 V1.0.....	404
1.15. 软件著作权：基于微信小程序的教室管理系统[简称：教室管理系统]V1.0.....	405
1.16. 软件著作权：基于区块链技术的校园公益平台系统[简称：校园公益平台]V1.0.....	406
1.17. 软件著作权：基于图卷积神经网络的学生课程个性化推荐系统[简称：触手课及]V1.0.....	407
1.18. 软件著作权：基于质点弹簧的血管手术模拟系统....	408
1.19. 软件著作权：农村耕地评价及智能决策系统.....	409
1.20. 软件著作权：自然环境特效渲染插件系统.....	410
1.21. 软件著作权：SuperSocket——基于手机 APP 管理的智能插座配套 APP.....	411
1.22. 软件著作权：基于深度学习的蝴蝶种类识别系统[简称：蝴蝶种类识别系统]V1.0.....	412
1.23. 软件著作权：基于神经网络的企业外贸汇率风险智能化决策系统[简称:企业外贸汇率风险智能化决策系统]V1.0	413
1.24. 软件著作权：汽车行业开放式创新 APP 系统 V1.0..	414
1.25. 软件著作权：基于大数据的渔业外贸数据分析平台[简称：渔业宝]V1.0.....	415
1.26. 软件著作权：面向大数据的汽车行业创新资源交易平台[简称:汽车行业创新资源交易平台]V1.0.....	416

1.27. 软件著作权：外贸事务专家在线咨询平台[简称:外贸在线咨询平台]V1.0.....	417
1.28. 软件著作权：企业外贸风险预警云平台[简称:外贸风险预警云平台]V1.0.....	418
1.29. 软件著作权：汽车行业创新资源交易大数据管理系统[简称:汽车资源交易管理系统]V1.0.....	419
1.30. 软件著作权：基于路径优化的旅游推荐系统[简称：旅游推荐系统]V1.0.....	420
1.31. 软件著作权：基于物联网数据分析的智慧养殖监测系统[简称：智慧养殖监测系统]V1.0.....	421


五、其他业绩

1.指导学生学科竞赛

1.1. 2024 年(第 7 届)“泰迪杯”数据分析技能赛本科及以上组一等奖.....	422
1.2. 2024 年(第 7 届)“泰迪杯”数据分析技能赛本科及以上组二等奖.....	423
1.3. 第十四届蓝桥杯全国软件和信息技术专业人才大赛全国总决赛 C/C++程序设计大学 B 组二等奖.....	424
1.4. 第十四届蓝桥杯全国软件和信息技术专业人才大赛全国总决赛 C/C++程序设计大学 B 组二等奖.....	425
1.5. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区 C/C++程序设计大学 A 组二等奖.....	426
1.6. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区 C/C++程序设计大学 A 组二等奖.....	427
1.7. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区 C/C++程序设计大学 B 组一等奖.....	428
1.8. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广	

东赛区 C/C++程序设计大学 B 组一等奖.....	429
1.9. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区 C/C++程序设计大学 B 组三等奖.....	430
1.10. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区 C/C++程序设计大学 B 组三等奖.....	431
1.11. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区 C/C++程序设计大学 B 组二等奖.....	432
2.个人荣誉	
2.1. 校级“教学名师”证书.....	433
2.2. 校级“庆祝建党 100 周年优秀共产党员”证书.....	434
2.3. 校级“教书育人先进个人”证书.....	434
2.4. 校级“优秀班主任”证书.....	435
2.5. 校级“华南农业大学第二届“双带头人”教师党支部书记素质能力大赛一等奖”证书.....	435
2.6. 院级“十佳工作者”证书.....	436
3.其他业绩	
3.1 全国党建工作样板支部.....	437
3.2 教育部内地与港澳大中小学师生交流计划大学生项目.....	443
3.3 全省党建工作样板支部.....	447
3.4 华南农业大学第二批“党建工作样板支部”.....	452
3.5 全省高校“双带头人”教师党支部书记“广东行”专项行动建设项目.....	456

1. 教学研究项目：关于教育部协同育人项目-基于鲲鹏的《操作系统》课程实践教学建设的立项通知及结题证书



中华人民共和国教育部

Ministry of Education of the People's Republic of China

当前位置: 首页 > 教育部司局网站 > 通知公告

教育部高等教育司关于公布有关企业支持的2019年第一批产学合作协同育人项目立项名单的函

各省、自治区、直辖市教育厅（教委），新疆生产建设兵团教育局，有关高等学校，有关企业：

为贯彻落实《国务院办公厅关于深化高等学校创新创业教育改革的实施意见》（国办发〔2015〕36号）和《国务院办公厅关于深化产教融合的若干意见》（国办发〔2017〕95号）精神，深化产教融合、校企合作，我司组织有关企业支持高校共同开展产学合作协同育人项目。根据《教育部高等教育司关于公布有关企业支持的产学合作协同育人项目申报指南（2019年第一批）的函》要求，有关高校积极组织师生向企业提交项目申请，有关企业对申报项目进行了遴选并向社会公示。现将立项项目汇总公布（见附件）。

有关高校要加强对项目的指导和管理，项目负责人要与相关企业加强联系，按照要求认真组织实施立项项目。有关企业要履行承诺，保证实际资金及软硬件投入按时到位，规范项目管理，保证项目顺利实施。

附件1：2019年第一批产学合作协同育人项目立项名单（按企业排序）

附件2：2019年第一批产学合作协同育人项目立项名单（按高校排序）

教育部高等教育司

2019年12月19日

项目编号	承担学校	公司名称	项目类型	项目名称	项目负责人
201901110005	华南农业大学	广州中望龙腾软件股份有限公司	新工科建设	面向新工科的机械类实践教学课程与实践	王惠雄
2019011738024	华南农业大学	华志林业有限公司	新工科建设	基于鲲鹏的《操作系统》课程实践教学建设	王金良
201901217013	华南理工大学	深圳国泰安教育技术有限公司	实践条件和实践基地建设	金融大数据与普惠金融实验室建设	刘仁和
201901240030	华南理工大学	天津丰软睿通教育信息技术有限公司	教学内容和课程体系改革	经济统计与算法综合实训的课证体系改革	田云 肖磊 陈海峰
201901254033	华南农业大学	武汉德商云计算股份有限公司	创新创业教育改革	农学院院创新创业实训平台建设	潘靖
201901013001	广东海洋大学	北京海威威光仪器技术有限公司	新工科建设	产教深度融合的光纤通信课程教学改革	王文华 吴伟娜
201901013016	广东海洋大学	北京海威威光仪器技术有限公司	实践条件和实践基地建设	以产学合作为抓手的人才培养实践基地建设	吴伟娜 王文华
201901035002	广东海洋大学	北京千禧互娱科技有限公司	创新创业教育改革	基于Python程序设计创新人才培养模式研究	陈入云
201901053003	广东海洋大学	北京现代中软软件开发有限公司	实践条件和实践基地建设	房地产业务与管理实践基地建设	喻清
201901071014	广东海洋大学	北京中软国际信马技术有限公司	实践条件和实践基地建设	基于中软国际的信管专业实践条件和实践基地建设	易伟明 郑慧红 肖敏
201901090004	广东海洋大学	广东力拓网络科技有限公司	教学内容和课程体系改革	广东海洋大学美育课程《音乐欣赏》移动教学课程体系建设	杜洁 李程 张常敏



教育部-华为产学合作协同育人项目 结题证书

项目批次：2019年第一批
 项目类别：新工科建设
 项目名称：基于鲲鹏的《操作系统》课程实践教学建设
 项目成员：王金凤，孙微微，张丽霞，张猜，肖磊
 承担单位：华南农业大学

经专家组评审，此项目已完成建设内容，予以结题，特发此证。

华为技术有限公司
二〇二二年四月十五日

科目收文明细账

部门：09000数学与信息学院 项目：(20002)产学合作协同育人项目-基于鲲鹏《操作系统》课程实践教学建设

科目信息：(99010301-)中央其他经费拨款- 起始年月：2019年01月

截止年月：2025年07月

单位：元

凭证日期	凭证编号	科目编号	科目名称	摘要	项目收入	项目支出	项目借款	项目还款
2020-05-21	(2)03052	99010301	中央其他经费拨款	华为技术有限公司转教学课题费入230062	30,000.00	0.00	0.00	0.00
2022-06-07	(2)01648	99010301	中央其他经费拨款	华为技术有限公司转教学课题费230062	20,000.00	0.00	0.00	0.00
负责人:王金凤				20002/产学合作协同育人项目-基于鲲鹏《操作系统》课程实践教学建设_合计	50,000.00	0.00	0.00	0.00
总计				总计	50,000.00	0.00	0.00	0.00

打印单位：华南农业大学

打印时间：2025-07-23 10:42:31

2.教学研究项目：关于教育部协同育人项目新一代软件工程教材建设的立项及结题证书



科目收文明细账

部门：(500)数学与信息学院 项目：(22002)产学合作协同育人项目-新一代软件工程规划教材-《操作系统》

科目信息：(61010203)其他类型项目来款- 起始年月：2022年01月

截止年月：2025年07月

单位：元

凭证日期	凭证编号	科目编号	科目名称	摘要	项目收入	项目支出	项目借款	项目还款
2023-10-31	(2)11729	61010203	其他类型项目来款	调整科目20220106#02364	50,000.00	0.00	0.00	0.00
负责人:王金凤				22002/产学合作协同育人项目-新一代软件工程规划教材-《操作系统》-合计	50,000.00	0.00	0.00	0.00
总计				总计	50,000.00	0.00	0.00	0.00

打印单位：华南农业大学

打印时间：2025-07-23 10:43:56

3. 教学研究项目：关于广东省高校教学改革项目的立项通知 及有关佐证材料

广东省教育厅

粤教高函〔2024〕9号

广东省教育厅关于公布 2023 年度广东省 本科高校教学质量与教学改革工程建设 项目立项名单的通知

各本科高校：

按照《广东省教育厅关于开展 2023 年度广东省本科高校教学质量与教学改革工程项目申报推荐工作的通知》等文件安排，经学校遴选推荐、省教育厅审核、公示等环节，现将 2023 年省本科高校质量工程建设项目立项名单予以公布，并就有关事项通知如下：

一、立项情况

确定立项建设省级实验教学示范中心 28 个、校企联合实验室 29 个、科产教融合实践教学基地 71 个、大学生社会实践教学基地 35 个、教师教学发展中心 5 个、课程教研室 167 个、现代产业学院 32 个、专项人才培养计划 62 项、高等教育教学改革项目 802 项。具体立项名单见附件。

二、项目管理

(一) 本次公布项目均为省质量工程建设项目，建设项目经学校组织建设、校内结题并通过省教育厅统一组织项目验收后，正式认定为省级项目。

(二) 项目正式实施前，请确保已对项目建设目标、建设举措、预期成果、建设进度安排等进行科学论证，论证专家应不少于5人，且至少有三分之一来自外校。论证后的目标、任务等将作为项目结题验收时的重要依据。

(三) 项目日常管理由学校主管部门负责，学校应统筹做好项目中期检查、校内结题验收等工作。校内结题时，邀请校外评审专家人数不得少于专家总人数的三分之二。满足以下条件的项目，经学校正式申请，可以参与省教育厅统一组织的项目验收：

- 1.项目已完成立项时设定的主要建设任务和目标；
- 2.项目已取得标志性建设成果，且该成果已在教学实践中得到检验和有效应用；
- 3.已按照要求完成项目校内结题；
- 4.符合当年度省统一验收规定的其他条件。

各校质量工程建设项目管理情况，将作为学校下一年度项目立项限额的参考依据。

(四) 项目实施过程中，其名称、建设内容（任务）、建设目标、建设周期、主要负责人、预期成果等发生重大变更的，需由时任项目负责人在发生变更后及时提出，经学校项目主管部门审核后由学校正式来函说明原因；擅自变更上述内容的，验收评定

时列为不通过。

三、其他事项

(一) 2023 年度各校向省教育厅推荐并获得立项的项目，学校须将项目校内评审、推荐及论证相关材料妥善保存，留底备查。

(二) 各校要统筹本校“冲补强”提升计划资金及自有资金对立项项目予以资助，项目获得学校资助情况将作为项目结题验收时重要考察因素之一。如项目建设中取得具有推广价值的优秀成果，请及时形成书面材料报省教育厅高教处。

联系人：李成军、窦月月，联系电话：020-37626882。

附件：2023 年度广东省本科高校教学质量与教学改革工程建设项目立项名单



4.教学研究项目：关于广东省在线开放课程的立项通知及有关佐证材料

广东省教育厅

粤教高函〔2018〕179号

广东省教育厅关于公布2018年广东省 本科高校教学质量与教学改革工程 建设项目立项名单的通知

各本科高校：

按照《广东省教育厅关于开展2018年度广东省本科高校教学质量与教学改革工程项目申报推荐工作的通知》（粤教高函〔2018〕120号）安排，省教育厅组织了2018年我省本科高校教学质量与教学改革工程（以下简称“质量工程”）项目推荐工作。经学校遴选、公示及推荐、省教育厅审核、公示，现将2018年省本科高校质量工程建设项目立项名单予以公布，并就有关事项通知如下：

一、立项情况

确定立项建设省级在线开放课程134门、实验教学示范中心24个、大学生实践教学基地65个、教师教学发展中心2个、教学团队75个、产业学院9个、重点专业22个、特色专业64个。

此外，评审认定省级虚拟仿真实验教学中心 16 个。详细名单见附件。

二、项目管理

(一) 除虚拟仿真实验教学中心外，本次公布的其他类别立项项目仅为省质量工程建设项目，经学校组织建设、校内结题并通过省教育厅统一组织项目验收后，正式认定为省级项目。

(二) 省虚拟仿真实验教学中心采取直接认定的方式，自本文发布之日起五年内有效，五年后可提请重新验收评定，届时通过评定的，有效期延长五年。

(三) 项目正式实施前，请确保已对项目建设目标、建设举措、预期成果、建设进度安排等进行科学论证，论证专家应不少于 5 人，且至少有三分之一来自外校。论证后的目标、任务等将作为项目结题验收时的重要依据。

(四) 项目日常管理委托学校主管部门负责，学校应根据项目建设周期和规律，按期统筹做好项目中期检查、校内结题验收等工作。各校质量工程建设项目管理情况，将作为学校下一年度项目立项额度的参考依据。

(五) 项目实施过程中，其名称、建设内容、建设周期、主要负责人、预期成果等发生重大变更的，需由项目负责人提出，经学校项目主管部门审核后由学校正式来函说明原因；擅自变更上述内容的，验收评定时列为不通过。

三、其他事项

(一) 2018 年度各校向省教育厅推荐并获得立项的项目，学校须将项目校内评审、推荐及论证相关材料妥善保存，留底备查。

(二) 项目由各校统筹本校“创新强校工程”资金及自有资金予以资助，项目获得学校资助情况将作为项目结题验收时重要考察因素之一。如项目建设中取得具有推广价值的优秀成果，请及时形成书面材料报省教育厅高教处。

联系人：刘雨濛、李成军，联系电话：020-37626882、37629463；传真：020-37627963。

附件：2018 年广东省本科高校教学质量与教学改革工程建设项目立项名单



在线开放课程立项建设项目汇总表

序号	单位名称	项目名称	项目负责人
1	中山大学	儿科学MOOC课	蒋小云
2	中山大学	组织学与胚胎学	滕丽娜
3	中山大学	环境土壤学	王诗忠
4	中山大学	慕课视域下网球教学与实战策略	张保华
5	华南理工大学	植物景观规划与设计	冯婉慧
6	华南理工大学	多媒体技术及应用	郭芬
7	华南理工大学	学术英语	韩金龙
8	华南理工大学	西班牙语初级	何文君
9	华南理工大学	酒店房务运营与管理	曲波
10	华南理工大学	计算机网络	袁华
11	暨南大学	日语——自游自在日本行	陈凤川
12	暨南大学	理论·制度·道路——中国特色社会主义理论与实践研究	程京武
13	暨南大学	自然保护与生态安全:拯救地球家园	黄柏炎
14	暨南大学	管理咨询	李从东、苏晓艳
15	暨南大学	晋商的法律智慧:商法	朱义坤
16	华南农业大学	朋辈辅导与人际交往	钟向阳
17	华南农业大学	压花艺术	陈国菊
18	华南农业大学	工程力学	陈海波
19	华南农业大学	公共管理学	史传林
20	华南农业大学	概率论	肖莉
21	华南农业大学	工程造价管理	李高扬
22	华南农业大学	现代生活的文化解读	廖杨
23	华南农业大学	林业经济学	李怡
24	华南农业大学	操作系统	王金凤
25	南方医科大学	公共卫生实践技能	李华
26	南方医科大学	生物药剂学与药物动力学	刘孟华
27	南方医科大学	系统解剖学	欧阳钧
28	南方医科大学	舌尖上的中药——五脏健康百病消	袁立霞
29	南方医科大学	组织学与胚胎学	张琳
30	广州中医药大学	中药鉴定学	黄海波
31	广州中医药大学	走进针灸	李敏
32	广州中医药大学	岭南罗氏妇科生精三部曲	罗颂平
33	广州中医药大学	精讲精学中医基础理论	严灿
34	广州中医药大学	中医药创新创业	叶清
35	华南师范大学	职业教育概论	陈伟、吴世勇
36	华南师范大学	篮球	郭永波、邓飞
37	华南师范大学	教育心理学	何先友
38	华南师范大学	物理化学	孙艳辉
39	华南师范大学	移动互联网创业模式与实践	王洪红
40	华南师范大学	组织行为学	周丽婷
41	广东工业大学	生态城市规划的理论与方法	蔡云楠
42	广东工业大学	学术英语阅读	陈晓茹

5.教学研究项目：关于华南农业大学质量工程项目的立项通知及有关佐证材料

华南农业大学文件

华南农教〔2017〕75号

关于公布华南农业大学2017年度校级质量工程 立项项目的通知

各学院、部处、各单位：

根据《广东省普通本科高校“十三五”教学质量与教学改革工程建设实施方案》（粤教高函〔2016〕128号）要求，学校组织开展了2017年度项目评选工作。经项目负责人申请、单位推荐、学校组织专家评审、校内公示无异议等程序，确定“华南农业大学华阳国际设计公司校外实践教学基地”等45个项目为2017年度校级质量工程项目，现予以公布（详见附件）。

请各项目负责人按照项目任务书及时开展各项工作，推进学校人才培养改革。各单位要切实加强对项目建设的督促、指导，

- 1 -

以确保项目如期完成。

附件：华南农业大学 2017 年度校级质量工程立项名单

华南农业大学
2017 年 11 月 22 日

公开方式：主动公开

华南农业大学校长办公室

2017 年 11 月 24 日印发

- 2 -

附件

华南农业大学 2017 年度校级质量工程立项名单

项目编号	项目类别	项目名称	项目负责人	单位名称
zlgc17001	大学生校外实践教学基地	华南农业大学华阳国际设计公司校外实践教学基地	陈乃华	水利与土木工程学院
zlgc17002		华南农业大学广州珠江机床有限公司实践教学基地	王红军	工程学院
zlgc17003		华南农业大学森明公司校外实践教育基地	郝丽	艺术学院
zlgc17004		华南农业大学广东蓝洋科技有限公司实践教学基地	袁腾	材料与能源学院
zlgc17005		华南农业大学广东粤建设计研究院有限公司校外实践教育基地	闫恩成	水利与土木工程学院
zlgc17006		华南农业大学广州德讯阳光青少年体育俱乐部实践教学基地	卢三妹	体育教学研究部
zlgc17007		华南农业大学广州乐庚信息科技有限公司实践教育基地	张昕	数学与信息学院
zlgc17008		华南农业大学中山穗葡农科教合作人才培养基地	江定心	农学院
zlgc17009		金融学课程教学团队	柳松	经济管理学院
zlgc17010		生态学实验教学团队	章家恩	资源环境学院
zlgc17011	环境生态学教学团队	黎华寿	资源环境学院	
zlgc17012	线性代数教学团队	魏福义	数学与信息学院	
zlgc17013	能源与环境系统工程专业主干课程群教学团队	蒋恩臣	材料与能源学院	
zlgc17014	刑事法实践教学团队	杜国明	人文与法学学院	
zlgc17015	园艺专业核心课程教学团队	胡桂兵	园艺学院	
zlgc17016	音乐表演专业教学团队	郑颖文	艺术学院	
zlgc17017	分析类数学教学团队	曹广福	数学与信息学院	
zlgc17018	国际经济与贸易教学团队	熊启泉	经济管理学院	
zlgc17019	精品视频公开课(选題)	生活中的数据分析	杨德贵	数学与信息学院
zlgc17020		概率论	肖莉	数学与信息学院
zlgc17021		操作系统	王金凤	数学与信息学院
zlgc17022	精品资源共享课	草坪学	陈曙	林学与风景园林学院
zlgc17023		嵌入式 Linux 系统	李震	电子工程学院
zlgc17024		数字电路与逻辑设计	黄瑛	数学与信息学院
zlgc17025		工程力学	陈海波	工程基础教学与训练中心
zlgc17026		计算机组成原理	黄润杰	数学与信息学院
zlgc17027		系统工程	俞守华	数学与信息学院
zlgc17028		林业经济学	李怡	经济管理学院
zlgc17029		基础工程	何春保	水利与土木工程学院
zlgc17030		水利计算	姜俊红	水利与土木工程学院
zlgc17031		水工建筑物	刘远	水利与土木工程学院
zlgc17032		公共管理学	史传林	公共管理学院

6.教学研究项目：关于华南农业大学一流课程的立项通知及有关佐证材料

华南农业大学文件

华南农教（2020）32号

关于公布华南农业大学 2020 年校级线上线下 混合式课程、线下课程建设 立项项目的通知

各学院、部处、各单位：

为贯彻落实《教育部关于加快建设高水平本科教育 全面提高人才培养能力的意见》（教高〔2018〕2号）精神，根据学校《关于开展校级线上线下混合式课程、线下课程培育项目申报工作的通知》要求，经课程教师申报、学院审核、教务处组织专家评审，校内公示无异议等程序，确定线上线下混合式课程 50 门、线下课程 29 门，共计 79 门课程为 2020 年度校级课程建设立项项目，现予以公布（详见附件）。

本次立项项目视同于学校质量工程项目，每门立项课程资助前期经费 1 万元，项目如期完成后另外资助 1 万元用于课程的后

续建设。请各项目负责人按照项目任务书及时开展各项工作，各单位要切实加强对项目建设的督促、指导，以确保项目如期完成。

- 附件：1. 线上线下混合式课程立项项目名单
2. 线下课程立项项目名单

华南农业大学
2020年7月9日

公开方式：主动公开

华南农业大学党政办公室

2020年7月10日印发

附件 1

线上线下混合式课程立项项目名单

序号	学院	姓名	立项课程
1	农学院	王少奎	分子生物学
2	农学院	王新荣	植物病理学
3	林学与风景园林学院	刘小蓓	世界遗产概论
4	林学与风景园林学院	翁殊斐	风景园林植物应用
5	林学与风景园林学院	陈崇贤	风景园林规划设计
6	林学与风景园林学院	高 伟	传统园林技艺
7	园艺学院	黄旭明	果树生理
8	兽医学院	曹伟胜	兽医流行病学原理
9	动物科学学院	张 哲	家畜育种学
10	资源环境学院	吕辉雄	低碳生活与绿色文明
11	资源环境学院	余光伟	水污染控制工程
12	海洋学院	杨慧荣	水生生物学
13	海洋学院	付京花	贝类学及增养殖
14	生命科学学院	朱国辉	生物化学
15	工程学院	魏德仙	单片机原理及接口技术
16	工程学院	张铁民	自动控制原理

序号	学院	姓名	立项课程
17	食品学院	周爱梅	食品化学
18	食品学院	王 丽	食品微生物检验学
19	水利与土木工程学院	郭焕宇	建筑美学与岭南文化
20	材料与能源学院	李春远	有机化学（公共课）
21	材料与能源学院	郭 琼	家具设计
22	材料与能源学院	林雅铃	药物分析
23	材料与能源学院	王 磊	（药物）毒理学
24	材料与能源学院	肖 勇	基础化学实验 II
25	数学与信息学院	肖 莉	概率论
26	数学与信息学院	王金凤	操作系统
27	电子工程学院	韩宇星	人工智能技术智慧农业应用导论
28	电子工程学院	杨初平	光电图像处理技术
29	经济管理学院	张 蓓	广告管理
30	经济管理学院	牛卫平	外贸函电
31	经济管理学院	贾 莉	市场营销学
32	经济管理学院	董 丽	会计学原理
33	经济管理学院	周小春	财务管理
34	经济管理学院	左 伟	管理学原理
35	经济管理学院	莫易娴	互联网金融

7.教学研究项目：关于华南农业大学计算机系统类课程教研室的立项通知及有关佐证材料

关于公示2021年省级质量工程拟推荐项目和校级质量工程拟立项项目的通知

审核发布：本科生院（招生办公室） 来源单位及负责人： 发布日期：2021-11-27 浏览次数：1410

各学院（部）、各相关单位：

根据学校《关于开展2021年度质量工程项目申报推荐工作的通知》要求，学校组织开展了本年度质量工程项目遴选工作。经项目负责人申请、所在单位推荐和学校组织专家评审等环节，本年度拟计划立项校级质量工程项目53项，并在本年度校级立项的基础上，推荐“经管类专业《数据处理及分析》平台课教研室”等17个项目为省级质量工程项目。现予以公示，具体项目见附件。

公示期自2021年11月27日至2021年12月1日。公示期间，如有异议请以书面方式提交（附必要的佐证材料），并署真实姓名，否则不予受理。

联系人：王欢

地址：行政楼329

联系电话：85280052

邮箱：gjy@scau.edu.cn

附件：2021年省级质量工程拟推荐项目和校级质量工程拟立项项目名单.pdf

华南农业大学本科生院

2021年11月27日

2021年省级质量工程拟推荐项目和校级质量工程拟立项项目名单

序号	拟项目类别	学院	项目名称	项目负责人	备注
1	在线开放课程	食品学院	食品化学（全英）在线开放课程	兰雅祺	拟校级立项
2	在线开放课程	农学院	种子生物学	周玉衣	拟校级立项
3	在线开放课程	农学院	作物栽培学	唐湘如	拟校级立项
4	在线开放课程	兽医学院	兽医微生物学	郭雪峰	拟校级立项
5	在线开放课程	资源环境学院	土壤学（双语）	任宗玲	拟校级立项
6	在线开放课程	资源环境学院	普通生态学	舒迎花	拟校级立项
7	在线开放课程	水利与土木工程学院	混凝土结构基本原理	吕艳梅	拟校级立项
8	“四新”示范课程	公共管理学院	土地经济学	李灿	拟校级立项
9	“四新”示范课程	艺术学院	中国传统器具设计研究	刘源	拟校级立项
10	“四新”示范课程	外国语学院	《（英语）学术论文写作》	肖好章	拟校级立项

23	全英课程	林学与风景园林学院	《树木学（全英）》	施诗	拟校级立项
24	全英课程	水利与土木工程学院	管理学原理（Principles of Management）	陈幸斌	拟校级立项
25	全英课程	兽医学院	细胞生物学	苏仁伟	拟校级立项
26	全英课程	食品学院	食品添加剂全英课（Food Additives）	吕慕雯	拟校级立项
27	全英课程	兽医学院	兽医传染病学	郑泽中	拟校级立项
28	全英课程	食品学院	畜产品综合利用/Comprehensive Utilization of Animal Products	吴绍宗	拟校级立项
29	全英课程	水利与土木工程学院	水利水电工程专业概论全英课（Introduction to hydraulic engineering）	王慧琳	拟校级立项
30	全英课程	工程学院	汽车拖拉机学	李庆	拟校级立项
31	全英课程	生命科学学院	细胞生物学	王浩	拟校级立项
32	全英课程	林学与风景园林学院	《森林生态学》全英课程建设	马玲	拟校级立项
33	课程教研室	数学与信息学院、软件学院	计算机系统类课程教研室	王金凤	拟校级立项
34	课程教研室	农学院	分子生物学思政课堂教研室	王少奎	拟校级立项

8. 教学研究项目：关于华南农业大学教学改革项目“翻转云课堂”在《操作系统》教学中的创新应用研究”的结题证书



9.教学研究项目：关于华南农业大学教学改革项目“基于创新能力培养的专业实践课中混合教学模式的探索与实践--以《操作系统》实践课为例”的立项通知及有关佐证材料

华南农业大学2020年校级教改项目拟立项名单

序号	项目名称	申报人	项目类别
1	基于创新能力培养的《养鱼学》教学模式改革研究与实践	孙宝霞	重点资助
2	以传承中华优秀传统文化为特色的应用型音乐人才培养模式改革研究	郑雁文	重点资助
3	风景园林专业美术课程群模块构建与教学创新	李静	重点资助
4	社会主义核心价值观教育融入“英美文学选读”教学的研究与实践	尹晓燕	重点资助
5	基于学生工程能力培养的机械基础实验课程改革与实践	甄文斌	重点资助
6	基于“新农科”建设的农林经济管理专业人才培养模式探索与实践	陈有华	重点资助
7	基于校企合作的风景园林协同育人模式探索与实践	黄东	重点资助
8	大思政视域下高校思政课程内实践教学基地建设的研究与实践——以华南农业大学为例	叶昕存	重点资助
9	课程思政教育在《禽病学》教学中的应用与实践	覃伟新	重点资助
10	中国民族音乐课程中的思政教学探索	朱虹	重点资助
11	“三全育人”背景下《无机功能材料》课程思政的建设与实践	倪春林	重点资助
12	普通大学高水平运动员培养模式改革与研究——以华南农业大学为例	陈基泰	重点资助
13	智慧植保导向下的精准农业人才培养体系构建	徐汉虹	重点资助
14	基于教育数据增强的高校学习评价方法研究	倪如燕	重点资助
15	基于大规模在线开放课程的线上“金课”建设探索——以“压花艺术”课程为例	陈国菊	重点资助
16	工程基础实践课在高校“冲一流”背景下的改革与实践	徐祖华	重点资助
17	新工科背景下工程类人才培养中课程思政育人模式研究	李润扬	重点资助
18	农林院校创新创业型人才培养视角下知识产权教育课程体系建设研究	张日新	重点资助
19	新农科背景下多学科交叉融合的园林专业人才培养模式研究与实践	詹继斐	重点资助
20	数学类课程线上线下混合教学的探究与实践	江雪萍	重点资助
21	新工科背景下基于能力培养的城乡规划专业“双线融合式”实践教学改革创新	赵雄华	重点资助
22	大规模远程在线考核关键技术研究	陈湘颖	重点资助
23	国际视野下生态环境类专业人才培养的研究与实践	蔡昆争	重点资助
24	高等农业院校工程类专业“普通化学”在线开放课程建设与应用	刘海峰	重点资助
25	大数据背景下基于文本挖掘的新农科人才需求分析与人才培养模式改革研究	张建雄	重点资助
26	基于移动平台的自然地理野外实习虚拟仿真系统建设与研究	钟晓兰	重点资助
27	以新工科人才培养为导向的发酵工程立体教学模式探索	徐学锋	重点资助
28	“植物学类专业课+古诗词+立德树人”的三全育人教学模式探索	解新明	重点资助
29	植物学虚拟协同融合实验教学体系创新路径与实践	方颖	重点资助
30	基于高开放性、创新性与挑战性“三位一体”的“金课”改革研究与实践——以《中国文学批评史》为例	孙宗美	重点资助
31	法医学虚拟仿真实验项目的研究与实践	杜国明	重点资助
32	新农科背景下应用型管理人才培养模式改革与实践——以土壤资源管理专业为例	李灿	重点资助
33	基于PBL模式的温氏塘《养牛学实训》课程改革与实践	邓铭	重点资助
34	校企协同育人下的应用型知识产权人才培养新模式研究	王红	重点资助
35	“三全育人”理念下“课程思政”前向度与路径探究——以非通识教育课的教学改革为例	薛晓梅	重点资助

36	基于混合式教学的课程思政改革探索——以《当代中国社会问题研究》课程为例	张运红	重点资助
37	DWap在大学英语的教学改革及研究	钟志英	重点资助
38	新农科背景下农林院校旅游管理专业建设的内涵与提升路径	林敏慧	重点资助
39	基于创新能力培养的专业实践课中混合教学模式的探索与实践——以《操作系统》实践课为例	王金涛	重点资助
40	“立德树人”视域下中外合作办学课程思政教学模式探索与实践——以日语专业中日合作办学项目为例	赵无运	重点资助
41	面向新农科建设的农林院校历史专业改革与发展研究	赵飞	重点资助
42	“MOOC+线下课”教学模式的探索与实践——以《生物化学》“MOOC+线上见面课”课程教学为蓝	初志成	重点资助
43	艺术类史论课程融入课程思政的探索与实践	胡辉	重点资助
44	基于创新能力培养的遗传学实验教学资源体系改革研究	李楠	重点资助
45	产教融合视域下复合应用型旅游管理人才培养体系创新与实践	陈丽丽	重点资助
46	农业高校广播电视编导专业涉农人才培养模式研究	郑文华	重点资助
47	动物专业产教融合协同育人模式的研究	陈瑰丽	重点资助
48	分级教学背景下大学英语B级线上线下混合式写作教学模式研究	张雅娜	重点资助
49	基于产教深度融合的农林经济管理实践教学体系内涵建设	李巧敏	重点资助
50	高等农业院校与农业产业园合作育人模式创新研究——以华南农业大学为例	潘军	重点资助
51	核心素养助推智慧教育导向的《水产动物组织胚胎学》教学探索与实践	范兰芬	重点资助
52	新农科建设视域下的哲学本科人才培养模式改革研究与实践	谢刚	重点资助
53	国际化新形势与新工科背景下林学专业人才培养模式改革与探索	周玮	一般资助
54	“后疫情时代”线上线下混合式教学模式在病毒学课程中的应用	马金成	一般资助
55	计算机基础课程线上实验平台与资源建设应用研究	李玉峰	一般资助
56	“线上+线下”混合式教学中协作学习体系的构建——基于经管专业数据分析类课程的探索与实践	李宗璋	一般资助
57	高校教师应对在线教学的教育技术应用能力提升路径探索	李玉玲	一般资助
58	《植物生理学实验》微课的建设	钱春梅	一般资助
59	后疫情时期BTE教学法在音乐通识课“线上”教学模式研究与实践	李伟	一般资助
60	十年来国内高等农林教学研究的可视化分析	杨永忠	一般资助
61	基于MOOC+SPOC或线上线下混合模式的林学实验教学改革与实践	苏怡	一般资助
62	高校思政“基础”课专题式教学研究	郑海	一般资助
63	互联网+背景下《电工电子技术》教学中的微课运用研究与实践	史艳玲	一般资助
64	新冠肺炎疫情后基于中国大学MOOC异步SPOC的混合式教学模式推进策略与效果评价——以“二·一·一”理科基础课《公共管理导论》的对比研究	吴法	一般资助
65	“新农科”背景下创新型媒体人才的教学探究——以《新媒体实务》课程为例	赵娜	一般资助
66	互联网背景下“产出导向法（POA）”在大学英语教学中的应用研究	牛莉	一般资助
67	基于MOOC的SPOC混合教学模式在《动物学》教学中应用探索	佟青春	一般资助
68	线上翻转课堂的教学模式在健美操体育课中的应用研究	何凤捷	一般资助
69	基于供给侧结构性改革理论的资料专业《土壤学》课程教学改革研究与实践	任京峰	一般资助
70	新工科理念下“食品添加剂”的“学-研-产”协同应用型课程设计与实践	肖杰	一般资助
71	模拟法庭课程建设质量评价体系研究	张艳琼	一般资助
72	以“传统文化”为切入点的课程思政融入金属工艺课程教学模式探索	潘子广	一般资助
73	基于教育大数据的高校学生学业预警机制构建研究——以华南农业大学为例	曾典	一般资助
74	基于网络教学平台的大学英语翻译模块课程混合式教学改革研究与构建	仇如慧	一般资助

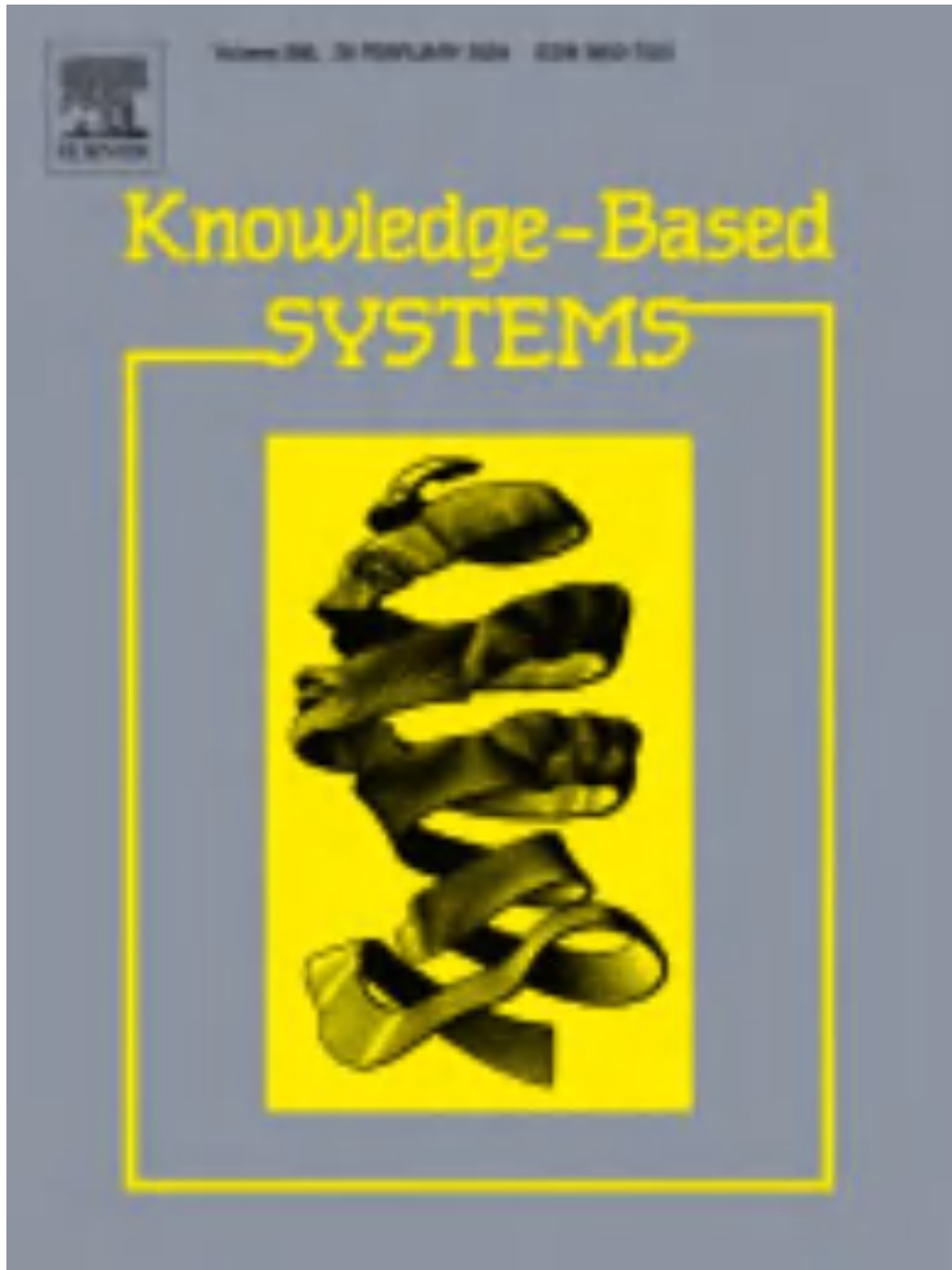
10.教学研究项目：关于华南农业大学教学改革重点项目“工程认证背景下基于知识图谱的新工科一流专业建设和提升的研究与实践”的立项通知及有关佐证材料

附件1

2023年拟推荐省级教学改革项目和校级教学改革拟立项项目名单

序号	项目名称	项目负责人	拟立项级别	备注
1	高阶思维视域下《种子生物学》“四维融合”混合式教学模式改革与实践	周玉亮	校重点项目	拟推荐省级
2	新林科视域下森林培育学知识图谱课程建设与教学模式创新	邱叔	校重点项目	拟推荐省级
3	一流专业建设背景下风景园林规划设计课程教学改革研究	陈崇贤	校重点项目	拟推荐省级
4	一流农科高校基础实验课程思政教学一体化改革与实践——以大学物理实验课程为例	劳婧媚	校重点项目	拟推荐省级
5	双一流高校“科-教-思”融合培养学生高阶思维的实践教学改革创新——以遗传学实验为例	李楠	校重点项目	拟推荐省级
6	工程结构设计软件课程的“融合+分层”教学	李文雄	校重点项目	拟推荐省级
7	基于竞赛和创新方法提升大学生创新能力的实证研究	易欣	校重点项目	拟推荐省级
8	基于创新创业能力培养的“赛教融合”《养羊学》课程改革与实践	柳广斌	校重点项目	拟推荐省级
9	人工智能赋能生态化大学英语混合教学改革研究	苏君	校重点项目	拟推荐省级
10	基于创新能力培养的《电路实验》教学改革研究	王建华	校重点项目	拟推荐省级
11	新时代高校思政课“问题链”教学模式探究——聚焦《习近平新时代中国特色社会主义思想概论》课程	何艳玲	校重点项目	拟推荐省级
12	三产融合理论在《茶叶生物化学》中的应用	张钰乾	校重点项目	拟推荐省级
13	工程认证背景下基于知识图谱的新工科一流专业建设和提升的研究与实践	王业斌	校重点项目	拟推荐省级
14	融合数据分析思维和学科交叉的《线性代数》课程教学创新与实践	张作峰	校重点项目	拟推荐省级
15	基于创新能力培养的统计学专业数据分析类实验课程改革创新与实践	周洪	校重点项目	拟推荐省级
16	基于知识图谱的农业院校课程思政建设探索——以公共数学基础课为例	张娜	校招标项目	拟推荐省级
17	基于新晋医人才培养的《分子生物学》课程改革与实践探索	沈永义	校重点项目	拟推荐省级
18	思政视域下《国际金融》课程教学改革与实践	周颀	校重点项目	拟推荐省级
19	新农科背景下基于“数字标本”的智慧植保实验体系研究与实践	李云峰	校重点项目	拟推荐省级
20	低碳农业背景下基于“一特两融”教学模式的环境土壤学课程体系构建	林庆祺	校重点项目	拟推荐省级
21	“双一流”建设视野下安全教育模式嵌入环境化学实验室课程体系的实践与优化	高峰	校重点项目	拟推荐省级
22	“三台协同，以美育人”——以课程教学为基础的高校舞蹈美育建设研究与实践	郑琳喆	校重点项目	拟推荐省级
23	劳动教育融入设计类专业课程路径探索——以《纤维艺术设计》课程为例	林汉聪	校重点项目	拟推荐省级
24	人工智能背景下的新文科艺术专业人才培养研究与创新：以动画专业为例	王柯	校招标项目	拟推荐省级

11. 教改论文： EduLGCL: Local-global contrastive learning model for education recommendation



Knowledge-Based Systems

Supports open access

Submit your article

Menu

Volume 286

28 February 2024

[Download full issue](#)

[< Previous vol/issue](#)

[Next vol/issue >](#)

Receive an update when the latest issues in this journal are published

[Sign in to set up alerts](#)

[Editorial board](#) [Full text access](#)

Editorial Board

Article 111486

[View PDF](#)

Research Article

[Research article](#) [Full text access](#)

Research on the attribute reduction method based on the best approximation set

Fachao Li, Kunyan Li, Chenxia Jin, Yuanjian Lin

Article 111362

[View PDF](#) [Article preview](#)

[Research article](#) [Full text access](#)

EduLCCL: Local-global contrastive learning model for education recommendation

Yijun Zhao, Fajian Jiang, Yin Pang, Yunxi Deng, ... Jinfeng Wang

Article 111357

[View PDF](#) [Article preview](#)

[Research article](#) [Full text access](#)

Knowledge-based Dual External Attention Network for peptide detectability prediction

Xiaocai Zhang, Hui Peng, Tao Tang, Yuansheng Liu, ... Jianjia Zhang

Article 111378

[View PDF](#) [Article preview](#)

[Research article](#) [Full text access](#)



EduLGL: Local-global contrastive learning model for education recommendation

Yijun Zhao^a, Fajian Jiang^a, Yin Pang^b, Yunxi Deng^a, Youyou Han^b, Jinfeng Wang^{a,b,*}

^a College of Mathematics and Informatics, South China Agricultural University, Guangzhou, 510642, China

^b Key Laboratory of Smart Agriculture of Guangzhou, South China Agricultural University, Guangzhou, 510642, China

ARTICLE INFO

Keywords:
Self-supervised learning
Recommendation
Graph convolutional network
Supervised learning
Collaborative filtering

ABSTRACT

Contrastive learning has been used in recommendation to learn user and item representations from sparse and long-tail user-item interaction histories. Recommendations applied in education have several key challenges. Firstly, the frequency of student-course interactions is lower, resulting in limited data. Secondly, existing models rely on multimodal data for multi-objective recommendations and they limit the model's generalizability. Lastly, it is important to consider the impact of static features on different students. Aiming at addressing these challenges, a novel course recommendation system based on Local-Global Contrastive Learning called as EduLGL is proposed, in which the mutual information between course features and student features is computed to select course features. After expanding the connections, the relationship between students and courses is described from different perspectives. Multiple graph convolutional operations are performed on the interaction history graph which can connect students and courses and represent student clusters. Through the feature extraction layer, the final student representation is obtained. Extensive experiments were conducted on three public datasets and a course dataset collected from a real-world environment to validate the performance of the EduLGL. The results demonstrate that EduLGL outperforms baseline methods, particularly in handling user cold-start problems. Our code is available at <https://github.com/mh7k/EduLGL>.

1. Introduction

Recommendation systems proactively suggest items that users may like or find interesting, alleviating user concerns caused by information overload. In the context of the rapid development of the Internet, online learning has become an essential part of the educational journey for students. Online education helps students break through the limitations of time and space, bringing them a higher degree of learning freedom. The phenomenon of information overload due to the huge amount of educational resources on the Internet prevents students from quickly finding courses and learning directions that interest in [1,2]. Learning in the wrong direction can easily lead to student dropout and a decline in learning interest. How to reduce students' selection costs, meet their personalized learning needs, and increase their interest in learning is the goal of recommendation systems in education [3].

Existing recommendation systems which are widely used in e-commerce scenarios [4,5], music scenarios [6], and video scenarios [7] mostly learn users' personalized preferences from their interaction history with items. However, in the educational scenario, student-course

interactions are cold behaviours, i.e. the act of choosing between students and courses exists only for a short period time in each semester. This means that it is difficult for recommender systems to accurately represent a student's learning interests from short interaction histories, and thus recommend courses to students that may be of interest to them. Good examples in other domains cannot be directly applied in the education scenario.

At the same time, there exists user cold-start problems in the education scenario, such as a freshman entering school without interaction with any courses, making it difficult to capture students' interests and preferences. No interaction or interactions less than 5 times are often removed in other recommender systems, which is not applicable to education. Therefore, learning a more accurate representation of students and courses based on the sparse history of student-course interactions becomes the key problem.

There is also a clear long-tail problem in the education scenario, where most students tend to opt for popular courses and there are some cold and excellent courses that tend to be overlooked. Because it is difficult to accurately express students' interests, it is difficult for

* Corresponding author at: College of Mathematics and Informatics, South China Agricultural University, Guangzhou, 510642, China.
E-mail address: wangjinfeng@scau.edu.cn (J. Wang).

<https://doi.org/10.1016/j.kbsys.2023.111357>

Received 12 October 2023; Received in revised form 11 December 2023; Accepted 28 December 2023

Available online 30 December 2023

0950-7051/© 2023 Elsevier B.V. All rights reserved.

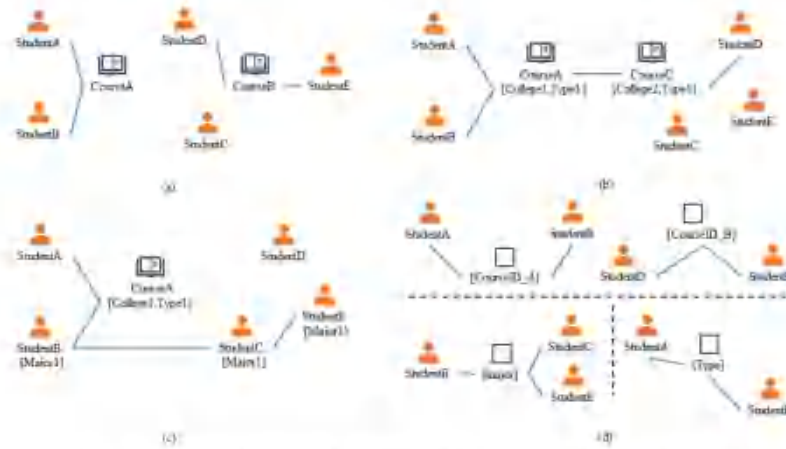


Fig. 1. Enhance connection to existing student-course interaction history using basic features of students and courses.

traditional recommender systems to help students explore these useful courses.

Recent researchers have also provided several solutions for the recommender systems applied in educational scenarios including leveraging social networks [6], constructing knowledge graphs with external knowledge sources [9], and learning student preference representations from multimodal data [10]. These methods have been adjusted to different recommendation objectives and have achieved promising results. However, this external knowledge is often not easily obtained and applied in educational scenarios. Therefore, the use of static profiles of students and courses can be considered to complement the interaction history of students and courses to learn more comprehensively the representation of students' interests. At the same time, there are student community relationships in the education scenario like social networks, as there are similarities in interests among students in the same majors. These community relationships are extracted from the static profile of students and do not need to be additionally collected like social network data. The use of static profiles of students and courses to supplement the interaction history is illustrated in Fig. 1.

Fig. 1(a) demonstrates a traditional recommendation approach that utilizes only the interaction history between students and courses for recommendations, limiting the representation of courses to mere ID numbers. Given the same courses taken by students A and B, and students D and E, interest similarity arises. Nonetheless, the system remains unaware of student C's interests. By the static attributes of courses, denoted as Type, it is possible to find the similarity between courses A and B. Consequently, this reveals that students D and E might have similar preferences and interests, as illustrated in Fig. 1(b). Utilizing the static profiles of students in the educational scenario, such as Major, to construct a student relationship network can reveal the connections between students B, C, and E. These three students might have similar preferences and interests, as illustrated in Fig. 1(c). Not only can this approach help identify the potential interests of student C, but it also reveals the interest relationships between students A, B, D, and E. Finally, to reduce the over-smoothing issue in the graph convolutional computation [11], the expanded interaction history is divided into multiple connected subgraphs based on different descriptive features. These subgraphs represent the connection relationships between students from different perspectives. As illustrated in

Fig. 1(d), a connection relationship among students is constructed from the perspectives of major, type, and ID.

Another key problem is how to choose the appropriate static features to better represent the student's preferences. Manual selection methods would require extensive experimentation. There are methods to measure feature importance based on average impurity [12] and gradient boosting decision tree selection based on gradient contribution [13,14]. These methods require more computer memory resources, and the effect of feature selection depends on the degree of dispersion of the data set. In this paper, we propose calculating the mutual information value between student ID features and course features as the basis for feature selection. By choosing the multiple features that have the most significant impact on student interests, we generate connected subgraphs from different perspectives.

To address the above problem of applying recommender systems in the education scenario, this paper proposes to complement the history of student-course interactions using static profiles of courses to describe student interest representations from different perspectives. At the same time, using static profiles of students to extract their community relationships and explore the learning interests of potential students. Constructing cross-graph contrastive learning to integrate information and mitigate long-tail problems in scenarios.

In this article, a recommendation model based on graph contrastive learning with mutual information is proposed to provide with students some courses in which may be interested. The main contributions of this work are listed as follows.

- A recommendation model using a local-global learning model with mutual information filtering features called EduLGCLE is designed to solve the data sparsity and user cold start in the task of course recommendation in educational scenarios.
- More accessible static characteristics of students and courses are proposed to extract group relationships among students and connection relationships under different perspectives and expand the connections between students and courses without the other data sources, which can improve the generalizability of the model.
- Mutual information is used to filter course statics and select appropriate feature combinations for extending existing student-course interaction history. This process generates a global relational view for contrastive learning and provides explanations for the recommended results.

- Extensive experiments on three real-world datasets demonstrate the improvement of the proposed model in recommendation accuracy. For application in real-world educational scenarios, two datasets of student control groups are collected for experiments. This further validates the feasibility and effectiveness of the proposed model in education.

The rest of the paper is organized as follows. Section 2 provides an overview of recent technologies. Section 3 presents the general form of the course recommendation problem in an educational scenario and the implementation details of EduLGCN. Section 4 reports the findings of experiments conducted to validate the enhancement of the model. Section 5 concludes the paper.

2. Related work

2.1. Content-aware recommendation

Employing embedding vectors to represent users and items, this study optimizes the learning of user-item interaction data. Similar to MF [35], it maps user and item IDs to low-dimensional embedding vectors and calculates their inner products to predict interactions. However, this method solely learns from user-item interaction data, while contextual features (e.g., location, time) also impact user interaction behavior. Recently, deep learning methods have emerged to leverage contextual feature interactions for improved recommendation performance [36–38]. For instance, NFM [36] proposes click operations and pooling on vectors, while AFM [37] introduces an attention mechanism to highlight the importance of feature interactions. GCM [38] encodes the context information through the encoder and adds it to the graph convolutional filter for learning. Despite their effectiveness, most of these works focus only on second-order feature interactions. The proposed EduLGCN extends the original graph-structured interaction data with static features of students and courses. By reducing the number of nodes at a fixed distance, it brings students closer together while better capturing higher-order collaborative filtering signals.

2.2. Graph convolutional network for recommendation

In early recommendation systems based on matrix factorization methods [35], sparse matrices were used to record the interaction relationships between users and items. The graph convolutional neural network (GCN) treats the interaction matrix as a bipartite graph that records the connectivity between user nodes and item nodes. By performing multiple graph convolution operations on this interaction graph and propagating knowledge between neighboring nodes, higher-order relationships between users and items can be captured. It can also handle data with asymmetric length of user interactions. NGCF [19] extends the traditional collaborative filtering approach by applying traditional graph convolutional networks to recommendation systems by capturing higher-order neighbor relationships in the interaction graph. LightGCN [20] considers the specificity of the recommendation problem to remove the unnecessary nonlinear activation and feature transformation matrices in traditional graph convolutional networks, which simplifies the computational process and uses the weighted sum to simulate the residual connections. A random-walk method was used in some models, such as PinSage [21] and GraphSage [22], to generate sub-graphs that sample and aggregate graph structure for efficient information transfer in large-scale graphs. However, this method is similar to node dropout and may not consider some cold nodes when walking. DiffNet [23] uses a differential attention mechanism to distinguish the degree of contribution of edges between nodes to improve recommendation performance. These approaches share the commonality of using graph structures to capture relationships between nodes, thereby improving the performance of recommendation algorithms.

2.3. Contrastive learning for recommendation

In recommender systems, low frequency data will lead in sparse supervised signals that the model can learn. By introducing contrastive learning, contrasting views with different neighbor structures are generated based on the main view by learning the representation of the same node on different views. Nodes in the main view can extract useful information from the contrastive view, which can effectively mitigate the data sparsity problem in the scenario [24]. The contrastive learning methods currently applied in recommender systems can be divided into two types based on generating contrastive views: methods with structural dropout and methods with local-global relationships. Contrastive learning based on structural dropouts, such as SGL [25], and SimGCL [26], can generate contrastive views in the main view or introduce random noise in the embedding space. This approach tends to introduce more natural noise and further exacerbates the sparsity of connections to cold nodes which are already sparse. Contrastive learning generating local-global relationships, such as NGL [27], HCCN [28], and LightGCL [20], may use external knowledge sources like clustering relationships or knowledge graphs to generate global representations of nodes as contrastive views. It can also perform matrix factorization on the main view to generate contrastive views including global information. While this approach effectively addresses the cold-start problem, its effectiveness relies on the quality of external knowledge. It may lose model generalization. The proposed EduLGCN constructs a global relational view using easily accessible static features of students and courses to describe students' preferred interests from different perspectives.

2.4. Application of recommendation system in education

In recent years, researchers have applied recommendation technologies in education scenarios to improve students' motivation and engagement in learning [3]. These methods often describe students' learning preferences from different perspectives based on knowledge graphs [20] constructed from external knowledge sources and multi-modal data such as video and audio [30]. However, these external knowledge sources typically require specific scenarios for information collection, which may result in poor generalization in different educational contexts. At the same time, these models make recommendations for objectives such as learning motivation in offline classes, and lack a quick filtering method for online classes. The proposed EduLGCN uses easily accessible static profile information to complement the connections of cold users by using potential community relationships between groups of students to complement. Freshman have not yet selected any university courses, which may explore in the wrong learning direction easily leading to an increase in absence rates and a decrease in motivation [31]. EduLGCN can use collective recommended preferences to help students explore new learning directions more accurately.

3. Methodology

3.1. The definition of course recommendation problem

The recommendation system [32] determines the user's preference for an item by analyzing the user's interaction history. This level of preference can be reflected by an item's rating made by the user, such as a course's rating. Using this paradigm, the historical interaction data between students and courses can be represented in the form of a rating matrix. However, in actual recommendation scenarios, students are only explicitly interested in a small portion of the courses by rating and reviewing and implicitly interested in a portion of the courses by browsing or clicking, but there is still a large portion of the courses without interaction with students, which is a long-tail phenomenon. The rating matrix is usually sparse and asymmetric. Using a bipartite graph to represent the interaction data can save storage space.

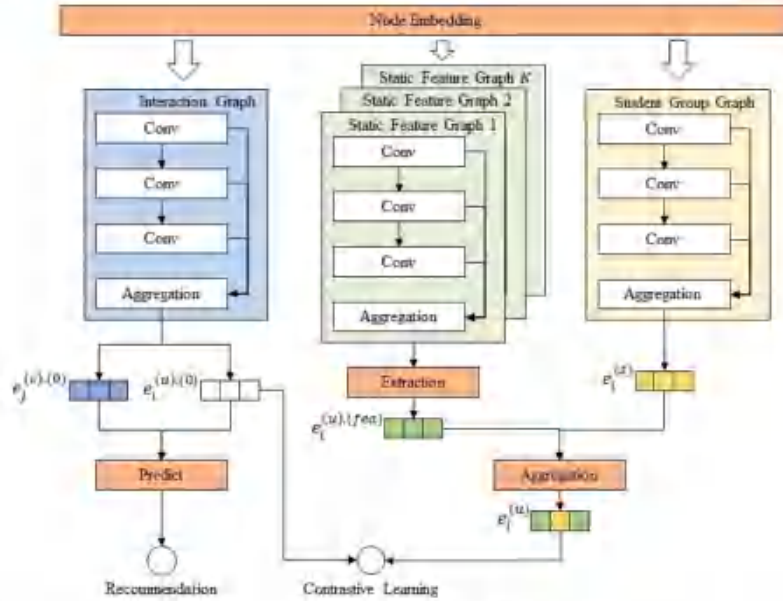


Fig. 2. The overall structure of EduAGCL.

The graph-based course recommendation problem can be defined in the following form. Let U and C denote the sets of N students and M courses, where $|U| = N$ and $|C| = M$. The interaction history of a student and a course can be represented as a student-course interaction graph G_{ij} , which is defined as $\{(u, v_{ui}) | u \in U, i \in C\}$, with edge $v_{ui} > 0$ indicating an observed interaction between student u and course i . The interaction behavior includes both explicit and implicit interactions. Conversely, $v_{ui} = 0$ indicates no observed interaction between student u and course i . The set of course features is denoted as D using K static features of the course to complement the course connection, where $|D| = k$. The course graph G_c records the relationships between course and their feature information, defined as $\{(i, link_{i,fem}) | i \in C, fem \in D\}$. When $link_{i,fem} = 1$, it indicates that course i can be described by static feature fem . $link_{i,fem} = 0$ means no relationship between them. However, using static features of different aspects to describe a course at the same time may lead to feature blurring [33]. The set of course features is denoted as $\{D_{k_1}, \dots, D_{k_T} \in D\}$ by classifying the static features into T groups with different indirect nodes according to the node types. The number of features contained in each indirect node is denoted as $\{k_1, \dots, k_T \in K\}$. Different feature-course sub-graphs can be obtained for each indirect node, which describes the connections between students and courses from different perspectives. These sub-graphs can be represented as $\{G_{c_1}, \dots, G_{c_T} \in G_c\}$.

Additionally, within the scenarios, there exist implicit connections among students beyond social relationships, such as shared interests between two students studying in the same major. The student graph, denoted as $G_s = \{(u, link_{uv}, friend) | u \in U\}$, captures the connectivity among students. Here, the edge $link_{uv} = 1$ signifies an observed connection between student u and their friend, while the edge $link_{uv} = 0$ indicates an unobserved connection between student u and their friend.

The inputs and outputs of the course recommendation model can be expressed as follows:

Input: a student-course bipartite interaction graph G_{ij} , a course graph G_c containing static features of the course, and a student group graph G_s .

Outputs: The edge weight v_{ui} between student u and course i will be predicted and sorted. The first several courses ranked in descending order by course prediction scores are used as a recommendation list.

3.2. The overall structure of EduAGCL

The overall structure of the EduAGCL proposed in this paper is shown in Fig. 2. The embedding layer provides trainable representations of student embeddings and embeddings of individual course features. All views share the same set of student embedding representations. The graph convolution process can be divided into three channels, including a main view channel with interaction history graphs, a static feature channel with all student-course connection graphs, and a student group channel containing student community information graphs. Three graph convolution operations are performed in each channel for propagating and aggregating low-order and high-order neighbour information for each node. In the main view channel, the learned embedding representations of student and course are fed into the prediction layer. An inner product operation is performed in the prediction layer to generate a similarity score indicating the student's preference for the course. The student embeddings learned in the static feature channel go through a fully connected feature extraction layer to produce the final embedding representation of student with global information according to each student's interests. A contrastive learning task with the global representation of the student is performed. The embedding representation of the student in the main view channel is considered as a local representation to extract useful information from the global representation of the student.

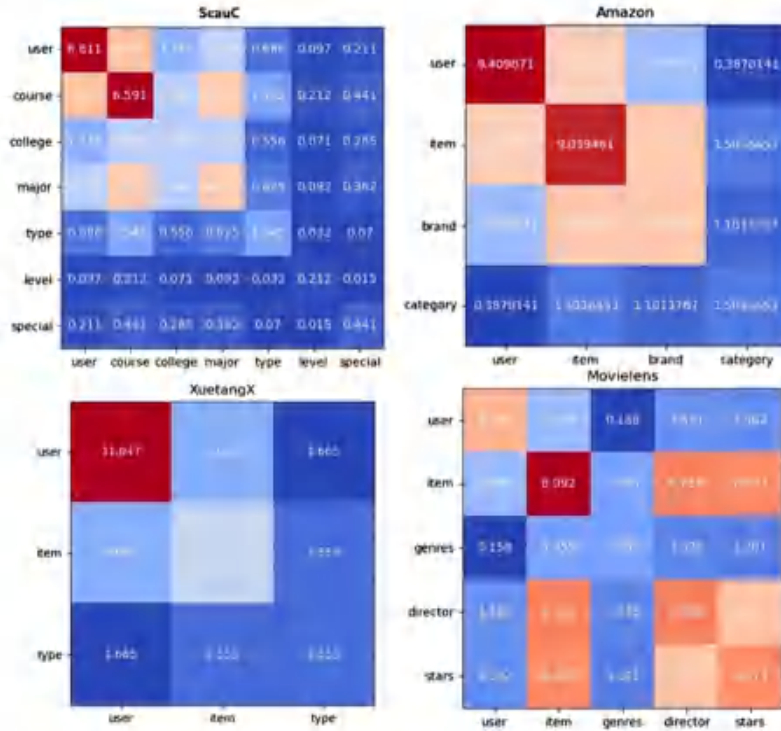


Fig. 3. Heat map of mutual information values for each dataset.

3.2.1. Embedding layer

Using a unique identifier maps the student, course, and feature labels into a vector, an embedding finding table is initialized as the corresponding embedding representations $e_i^{(s)}$, $e_j^{(c)}$, and $e_l^{(f)}$, for the student node u_i , the course ID node c_j and course static feature node $l_l^{(f)}$. All student nodes and course nodes can be represented as $E^{(s)} \in \mathbb{R}^{N \times d}$ and $E^{(c)} \in \mathbb{R}^{M \times d}$, where N and M is the number of students and courses; d denotes the dimension of the embedded representation.

3.2.2. Local relation learning

Neighbour information propagation is performed three times on the student-course interaction graph G_0 which represents the interaction history of the student and the course. The first-order neighbour and high-order neighbour information of the nodes are learned. Local representations of student nodes and course nodes are obtained. Referring to the LightGCN and MFGCL [34], EduLGCCL is designed using a graph convolutional encoder with the feature transform matrix, non-linear activation method, and initial layer embedding removed. At layer $(l + 1)^{th}$, the information aggregation process can be represented in the following form.

$$e_{i,j+1}^{(s)} = \sum_{c \in N_c^{(l)}} \frac{1}{\sqrt{|\sum_j R_{ij}^{(l)}|} \sqrt{|\sum_j R_{ij}^{(l)}|}} e_{i,j}^{(s)} \quad (1)$$

$$e_{j,l+1}^{(c)} = \sum_{s \in N_s^{(l)}} \frac{1}{\sqrt{|\sum_i R_{ij}^{(l)}|} \sqrt{|\sum_i R_{ij}^{(l)}|}} e_{j,l}^{(c)} \quad (2)$$

where $e_{i,j+1}^{(s)}$ and $e_{j,l+1}^{(c)}$ denote the results of student node u_i and course node c_j after aggregating and propagating the neighbour information in G_0 in the $(l + 1)^{th}$ layer; $\sum_j R_{ij}^{(l)}$ and $\sum_i R_{ij}^{(l)}$ denote the degree of student node u_i and course node c_j in G_0 ; R is defined as a degree matrix; $N_s^{(l)}$ and $N_c^{(l)}$ denote the aggregation of neighbour nodes of student node u_i and course node c_j in G_0 . The final representation of student node u_i and course node c_j in G_0 is obtained by aggregating the node representations of all layers in G_0 . The calculations can be expressed as the following form.

$$e_i^{(s)} = \alpha_1 e_{i,1}^{(s)} + \alpha_2 e_{i,2}^{(s)} + \alpha_3 e_{i,3}^{(s)} \quad (3)$$

$$e_j^{(c)} = \alpha_1 e_{j,1}^{(c)} + \alpha_2 e_{j,2}^{(c)} + \alpha_3 e_{j,3}^{(c)} \quad (4)$$

To simplify the calculation, the weights in EduLGCCL are set as $\alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{3}$. It is also possible to use the attention mechanism to generate the weights [35,36].

The final representations of student node u_i and course node c_j are subjected to an inner product to generate a similarity score for predicting the weights y_{ij} of the edges between student node u_i and course node c_j in G_0 . After sorting the values y_{ij} in similarity, the top 10 and top 20 courses are selected and output as a recommendation list. The calculations are shown below.

$$y_{ij} = e_i^{(s)} \cdot e_j^{(c)} \quad (5)$$

3.2.3. Global relation learning

In order to reduce feature blurring and over-smoothing in graph convolution computation, the student-course connectivity graph is divided into multiple student-course sub-graphs in different perspectives, denoted as $\{G_{c_1}, \dots, G_{c_T}\} \in \mathcal{G}_T$ in which T denotes the total number of selected course feature nodes and G_{c_i} is denoted as G_{ij} . The number of feature nodes in each graph can be noted as $\{K_0, \dots, K_T\} \in \mathcal{K}$, where \mathcal{K} represents the set of static course characteristics, such as course, college, and major. By calculating the mutual information value between course characteristics and student identification features, T distinct features are selected. The heat maps of mutual information values for each dataset are illustrated in Fig. 5. It was found that selecting the feature group with the largest mutual information value and the smallest worked best. Because this kind of combination is least affected by the feature blurring problem.

A three-layer graph encoder is applied to each student feature graph to learn and aggregate higher-order neighbor relationships for each node. Referring to SocialGN [33], an information extraction mechanism is designed. It utilizes a fully connected layer to aggregate the final representations of student nodes and feature nodes. The aggregation is performed on T student-course graphs. This process generates a personal global representation of student nodes. The calculations can be expressed as the following form.

$$z_c^{global} = w_{conv} \left\{ \sigma \left(w_{gcn}^{global} \right) \left[\sigma \left(w_{c_i} z_c^{local} \right) \right] - \left[\sigma \left(w_{c_{i+1}} z_c^{local} \right) \right] \right\} \quad (6)$$

$$z_c^{local} = \frac{z_c^{global}}{\left\| \frac{z_c^{global}}{\|z_c^{global}\|_1} \right\|_1} \quad (7)$$

where σ denotes the nonlinear activation function \tanh ; $w_{gcn}, w_{c_i}, \dots, w_{c_{i+1}}, w_{conv}$ is the trainable matrix used for adaptive extraction of personal preferences; $w_{gcn} \in \mathbb{R}^{(T+1) \times K}$ is the trainable matrix used for dimension reduction of the concatenated node representation. z_c^{local} denotes the final representation of student node c_i in the student-course graph G_{c_i} , and particular z_c^{global} is $z_c^{global} = z_c^{local} \otimes z_c^{local}$ denotes the final representation of student node c_i . The normalization operation prevents z_c^{global} from being too large.

Similarly, the connections between groups of students can effectively broaden their interests, especially the interactions among freshmen. The model performed three rounds of graph convolution on the student group graph to uncover both direct and second-order connections among students. After that, the model combined this information with the ultimate global representation of student nodes, creating node representation for comparative learning. The specific calculation process can be presented in the following.

$$z_i^c = \text{aggregation} \left(z_i^{global} \otimes z_i^c \right) \quad (8)$$

where z_i^c represents the representation of student node c_i obtained from the student graph G_{c_i} , encompassing the student node c_i as well as information from the higher-order neighbors, which is computed as in Eq. (7). To streamline computations, the aggregation here is performed by taking the average.

3.2.4. Local-global contrastive learning

In the joint learning approach constructed in this paper, the main view performs the recommendation task, using BPR Loss [17] as the loss function. The contrast view performs the contrast learning task using InfoNCE loss [28] shown below.

$$L_{InfoNCE} = - \sum_{(i,j) \in N_p \cup N_n} \ln \left(\frac{\exp(\gamma_{ij})}{\sum_{(k,l) \in N} \exp(\gamma_{kl})} \right) \quad (9)$$

During training in a mini-batch, the unobserved student-course connections are treated as negative samples, while the observed connections are treated as positive samples. N_p denotes the set of all courses with

Table 1
The comparison of time complexity.

	LightGCN	LGCAF	SGL-ED	EduLGCN
Adjacency matrix	$2 A $	$4 A + 2K A $	$4 A + 4 A $	$4 A $
Graph traversal	$O(A)$	$O(L + 2K \cdot A)$	$O(L + 4) \cdot A $	$O(2K \cdot A)$
Prediction	$O(2 A)$	$O(2 A)$	$O(2 A)$	$O(2 A)$
Constant	-	-	$O(BN)$	$O(BN)$

which student a interacts in the current mini-batch; γ_{ij} denotes the similarity scores of negative sample pairs (i, j) , σ denotes the sigmoid activation function.

By employing contrastive learning loss, the distances between the local and global representations of the same node in the embedding space are reduced, while the distances between different nodes are amplified. This method allows the extraction of useful information from global signal representations to enhance the local signal representations, thereby mitigating the issue of data sparsity. The contrastive learning loss function is shown below.

$$L_{CL} = - \sum_{i \in \mathcal{N}} \log \frac{\exp(\langle z_i, z_i^{local} \rangle / \tau)}{\sum_{j \in \mathcal{N}} \exp(\langle z_i, z_j \rangle / \tau)} \quad (10)$$

where z_i and z_i^{local} denote the global and local representations of the same node i respectively; z_j denotes a negative sample node, referring to other nodes of the same type within the current mini-batch, excluding the node itself; τ represents the temperature parameter of the softmax function, which controls the degree of the contribution of the negative sample pairs during the training process.

The joint learning method for the recommendation task and the contrastive learning task is shown below.

$$L = \lambda_1 L_{BPR} + \lambda_2 L_{CL} + \lambda_3 \|\Theta\|_2^2 \quad (11)$$

where λ_1 controls the strength of contrastive learning; λ_2 controls the strength of L2 regularization; Θ is the set of model parameters, including the initialised embedding of all nodes and the matrix of trainable parameters for the fully connected layer.

3.3. Time complexity analysis

In this subsection, the theoretical time complexity of LightGCN [20], LGCAF [22], SGL-ED [25], and EduLGCN will be analyzed in a mini-batch shown in Table 1. Let $|A|$ denotes the number of edges observed in an interaction graph; K denotes the number of selected features; r denotes the probability of edge discard; L denotes the number of layers of convolution; d denotes the dimension of the embedding space; B denotes the mini-batch size; and N denotes the number of nodes in the current mini-batch.

- Both EduLGCN and LGCAF utilize item description information to expand the interaction graph, and the regularized adjacency matrix constructed has $2|A| + 2K|A|$ non-zero elements. Due to the feature ambiguity problem, not all description information can bring positive improvements. LGCAF attempts various information combinations through manual methods, while EduLGCN uses mutual information for feature selection. For SGL-ED, because it uses edge dropout to construct two contrastive learning channel views, each contrast channel's adjacency matrix has $2r|A|$ non-zero elements. LightGCN does not use graph augmentation operations, so the regularized adjacency matrix has only $2|A|$ non-zero elements.
- In the graph traversal, since SGL-ED is a three-channel graph comparison learning framework, the time consumption is almost three times higher compared to the single-channel LightGCN.

LGCACF and EduLGCCL, on the other hand, experience increased time consumption because they use decomposition of multiple sub-graphs.

- The prediction layers of the above methods are trained in each mini-batch using BPR Loss, resulting in similar time consumption for this part.
- Only SGL-ED and EduLGCCL utilized contrastive learning frameworks, which results in time consumption due to the acquisition of positive and negative samples. Treating the different representations of nodes themselves as positive samples requires $O(MN)$ time, while considering other nodes within the same mini-batch as negative samples takes $O(M^2N)$ time, where M is the ratio of negative samples in positive samples.

4. Experiments and analysis

To verify the effectiveness of the proposed EduLGCCL, we perform massive experiments to explore the following research questions. EduLGCCL performs on different datasets compared with Baseline. EduLGCCL performs with different numbers of features. EduLGCCL performs for user cold-start problems on different datasets. Different parameter settings affect the performance of EduLGCCL.

4.1. Experimental setting

4.1.1. Datasets

The experiments are conducted on three real-world datasets: MovieLens [40], Amazon [41] and XuetangX [42]. EduLGCCL is proposed for educational applications, so the real data is collected from different grades of one university. Offline questionnaires are adopted to collect information from students and their intention to interact with the courses. The used public datasets are sourced from various application scenarios, exhibiting variations in both dataset size and feature dimensions. Divide the data into training set, validation set, and testing set in the ratio of 7:1:2. The detailed information of each dataset is shown in Table 2.

- **MovieLens** [40]: It is a film dataset widely used to evaluate the performance of recommender systems. Film genre, film director, and film rating were chosen as the static profiles of the movies. After removing invalid films with an interaction count of less than 5, the user's rating behavior is considered as implicit interaction data.
- **Amazon** [41]: A large and widely used e-commerce dataset, of which Amazon Electronics was selected to evaluate the performance of EduLGCCL. All interactions between users and products are treated as implicit data, except for products with less than 5 interactions and users with less than 10 interactions.
- **Taobao** [43]: A dataset from the e-commerce domain, collected from Taobao.com, was utilized in this study. The interaction data encompassed "purchase" and "add to cart" behaviors. Users with less than 10 interactions were excluded.
- **XuetangX** [42]: Data was collected in 2018 at XuetangX, which was one of China's largest MOOC platforms, including 1,302 courses, 82,535 users, and 458,454 user-course interactions. All interactions were recorded with the same weight as invisible interests.
- **ScauC** [Ours]: This dataset was collected from our university including 2002 core courses taken as the course set. Students' interests in the courses were gathered through off-line questionnaires. The students were divided into two groups: the freshman group consisting of 445 students and the senior group consisting of 350 students. The active interest preferences and existing course enrollment information were adopted from both groups of students. After filtering out courses without interaction at all, all expressed interests between students and courses were considered as explicit interests and recorded with equal weight.

Table 2
The details of each dataset.

Dataset	Users	Items	Rating	Density	Features	Primary count
MovieLens [40]	610	9606	10000	0.0171	genre director actor	21 2942 1034
Amazon [41]	1220	14004	24060	0.002	brand category	2771 15
XuetangX [42]	82535	1302	458454	0.0002	type	19
Taobao [43]	89,098	89,406	918,881	0.00008	category	167
ScauC [Ours]	897	2074	7227	0.0029	course type level quest	21 31 3 4 3

4.1.2. Baselines

The following models are used as baselines for comparison, including Matrix Factorization (MF) [16], LightGCN [20], LGCACF [26], SGL-ED [23], BUB [44], SimGCL [36], and MFGCL [34].

- **MF** [16]: Matrix Factorization is a classic and widely used technique in recommendation systems to decompose a user-item interaction matrix into latent features. It enables the identification of hidden patterns and preferences, allowing personalized recommendations based on these latent factors.
- **LightGCN** [20]: The application of graph convolutional networks was proposed in recommendation models by extending traditional collaborative filtering methods and addressing the inability to effectively capture latent high-order relationships. Unnecessary nonlinear activations and feature transformation matrices have been removed for the recommendation task.
- **LGCACF** [26]: Using LightGCN as the foundational encoder, distinct interaction graphs are constructed for various static item features and user identifier features to represent user embeddings from different perspectives. Global user representations are generated through averaging, allowing exploitation of item connections from different viewpoints.
- **SGL-ED** [23]: This is a joint recommendation model applying graph contrastive learning. To address the long-tail and sparse supervision issues in recommendation scenario, contrastive learning was introduced by SGL. Different contrastive learning views were generated using edge dropout, establishing a three-channel contrastive learning mode comprising a main view and contrastive learning views. A unified learning mode was constructed to simultaneously optimize the recommendation loss and contrastive loss. ED was the best-performing variant to represent the method using edge dropout in the SGL model.
- **BUB** [44]: The recommendation model, using only first-order collaborative filtering signals, adopts a two-tower architecture for the target network and online network. Self-supervised learning is applied to positive sample pairs to extract augmented views from neighborhood information, mitigating the issue of data sparsity. However, negative sample pairs are not considered in the optimization learning process.
- **SimGCL** [36]: Unlike SGL-ED, this model proposes the addition of uniform noise in the embedding space to generate different representations of the same node, enabling contrastive learning without changing the original graph structure. As this method introduces perturbations in the embedding space, it can be considered a form of local-global contrastive learning model.
- **MFGCL** [34]: A Top-K recommendation model based on contrastive awareness and contrastive learning is proposed. LightGCN serves as the base encoder, dividing the binary user-feature interaction

graphs into multiple parts based on node types. Distinct representations are obtained from different perspectives on these interaction graphs, which are then aggregated to form the final representation. A three-channel contrastive learning structure is constructed, wherein two contrastive learning channels generate disturbed representations by adding random noise to the embedding space.

4.1.3. Hyper-parameter setting

For the embedding layer setup, the Xavier initializer is used to create a 512-dimensional initialized user and message embedding. All models are optimized using Adam optimizer to search for appropriate learning rates in $\{0.01, 0.005, 0.001\}$. The coefficient λ_2 of the L2 regularization is fixed to 10^{-4} . The batch size is set as 2048. The appropriate temperature coefficients in the contrastive learning models, SGL-ED and EduLGC, are determined in $\{0.2, 0.175, 0.15, 0.125, 0.1\}$, and the contrastive learning strengths λ_1 are searched for the appropriate scale in $\{0.2, 0.1, 0.05, 0.01\}$. All experiments were conducted on a computing platform with an Intel-5 12600K processor and an RTX-3060 12G.

4.1.4. Evaluation metrics

Two evaluation matrices are used to measure recommendation results $\{1, 0, 0.5\}$: $Recall@K$ and $NDCG@K$. K is set as 10 and 20, which means recommending the top 10 or 20 courses that are most likely to be interested. $Recall@K$ measures how many relevant courses interesting to students are included in the recommendation list. It is used to assess whether the recommendation model successfully recalls courses interesting to students. Indicators are calculated as shown in Eq. (12). $NDCG@K$ refers to Normalized Discounted Cumulative Gain at K , which takes into account the relationship between the first K recommendation results and the correlation used to measure the quality of the recommendation. The higher the indicator is, the closer the output to the real preference of students is, indicators are calculated as shown in Eq. (13). All reported experimental results are averages after 5 runs.

$$Recall@K = \frac{TP@K}{TP@K + FN} \quad (12)$$

$$NDCG@K = \frac{DCG@K}{IDCG@K} = \frac{\sum_{i=1}^K \frac{rel_i}{2^{(i-1)}}}{\sum_{i=1}^K \frac{rel_i^*}{2^{(i-1)}}} \quad (13)$$

where $TP@K$ denotes the number of courses with real interactions in the top K results of the recommendation list; FN denotes the number of courses with real interactions that are not in the recommendations list; $DCG@K$ refers to the discounted cumulative gain in the top K recommendation results; $IDCG@K$ indicates the discounted cumulative gain in the top K recommendation results in the ideal cases; rel_i refers to the relevance of the i th recommendation result; rel_i^* refers to the relevance of the i th recommendation result in descending order of relevance.

4.2. Performance comparison

The experimental results are presented in Table 3, Tables 4 and 5, in which $R@K$ represents $Recall@K$ and $N@K$ represents $NDCG@K$. In order to simulate the recommendation effect in the real environment, the XuetangX and ScouC datasets were simply cleaned, and the cold-start user items with less than 10 interactions were retained. In all experiments, the contrastive learning methods yielded better results. Compared to traditional models, SGL-ED and EduLGC consistently demonstrated great advantages owing to their contrastive learning can obtain embeddings with a more uniform distribution.

As shown in Table 3, the performance of EduLGC surpasses all baseline methods. The results indicate that expanding the observed interaction data by utilizing selected item static features can represent the

Table 3
Performance comparison with baselines on multi-feature datasets.

Dataset	Moyilems				Amazon			
	R@10	N@10	R@20	N@20	R@10	N@10	R@20	N@20
MF	0.1072	0.2222	0.1762	0.2389	0.0459	0.1625	0.0309	0.0945
LightGCN	0.1360	0.2582	0.2093	0.3392	0.0568	0.0792	0.0810	0.0995
IGCAGF	0.1169	0.3172	0.1825	0.3447	0.0636	0.0794	0.0851	0.0844
R3R	0.1438	0.4201	0.2039	0.3161	0.0355	0.0282	0.0572	0.0661
SGL-ED	0.1479	0.3415	0.2859	0.3498	0.0709	0.0825	0.0932	0.0917
SimGCN	0.1491	0.3429	0.2718	0.3480	0.0585	0.0794	0.0811	0.0886
MRGCN	0.1466	0.3468	0.2776	0.3531	0.0607	0.0728	0.0815	0.0832
EduLGC	0.1572	0.3601	0.2392	0.3501	0.0768	0.0845	0.0886	0.0943
Improvs	+5.03%	+5.01%	+2.27%	+2.34%	+3.37%	+2.42%	+5.74%	+3.54%

Table 4
Performance comparison with baselines on single-feature datasets.

Dataset	TaoBao				XuetangX			
	R@10	N@10	R@20	N@20	R@10	N@10	R@20	N@20
MF	0.4761	0.2876	0.2797	0.4225	0.4781	0.3383	0.2797	0.4215
LightGCN	0.6479	0.6343	0.6710	0.6419	0.5210	0.4076	0.6158	0.4796
R3R	0.0210	0.0201	0.0408	0.0557	0.4948	0.2945	0.3246	0.4911
SGL-ED	0.0290	0.0403	0.0865	0.0515	0.5418	0.4719	0.6461	0.5930
SimGCN	0.0283	0.0403	0.0850	0.0506	0.5282	0.4730	0.6374	0.5070
MRGCN	0.0277	0.0412	0.0868	0.0510	0.5312	0.4527	0.6267	0.4669
EduLGC	0.0289	0.0423	0.0865	0.0514	0.5197	0.4701	0.6421	0.2038
Improvs	-0.64%	+0.33%	-0.08%	-0.10%	+0.39%	-0.61%	+0.77%	-0.61%

Table 5
Performance comparison with baselines on real educational scenario dataset.

Dataset	ScouC			
	R@10	N@10	R@20	N@20
MF	0.1904	0.1175	0.2381	0.1275
LightGCN	0.1757	0.1192	0.2558	0.1402
IGCAGF	0.1839	0.1190	0.2645	0.1430
R3R	0.1705	0.1170	0.1448	0.1460
SGL-ED	0.1829	0.1196	0.2528	0.1414
SimGCN	0.1810	0.1191	0.2497	0.1406
MRGCN	0.1793	0.1209	0.2534	0.1451
EduLGC	0.1895	0.1238	0.2724	0.1490
Improvs	+1.61%	+2.99%	+2.99%	+3.23%

distance between two target nodes from different perspectives. This effectively discovers users with similar interests, thus better representing user preferences and enabling personalized recommendations.

As illustrated in Table 4, employing a single static feature to expand the interaction paths alone renders no better method to explore and minimize the paths between two user nodes. In this case, utilizing the expanded interaction graph for global information extraction is significantly hampered by the ambiguity issue. Furthermore, challenges such as data noise and overly short interaction sequences persist. Consequently, the overall performance of EduLGC on this dataset is somewhat weaker.

As presented in Table 5, EduLGC enhances the performance of recommendation metrics in the real-world interactive dataset ScouC. The method of combining existing student-course interaction history with static features to generate a comprehensive relationship view strengthens the interpretability of the recommendation outcomes, enabling the provision of socially responsive and personalized course recommendations for students.

4.3. Ablation experiment

To explore the importance of contrastive learning and static feature enhancement in EduLGC, ablation experiments are performed to validate and analyse the performances of the following variants: EduLGC without static feature enhancement (w/o info) and EduLGC without

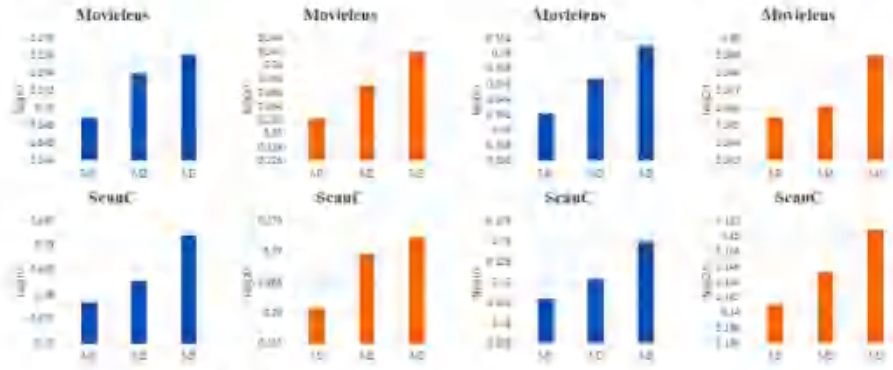


Fig. 4. The results of EduLGCN on MovieLens and ScauC datasets with various feature combinations.

Table 6

The ablation results of EduLGCN and its variants on the XuetangX and Amazon

Dataset	Metric	w/o CL	β -value	w/o info	β -value	EduLGCN
Amazon	AUC	0.5010	0.0043	0.0944	0.0033	0.0746
	F1	0.0587	0.0026	0.0592	0.0045	0.0545
	F2	0.0833	0.0047	0.0810	0.0012	0.0780
XuetangX	AUC	0.5208	0.0019	0.5222	0.0001	0.5407
	F1	0.4543	1.62E-06	0.4479	2.87E-05	0.4760
	F2	0.6286	0.0003	0.6177	1.52E-07	0.6451
ScauC	AUC	0.4903	1.40E-02	0.4865	1.13E-07	0.5306

contrastive learning (w/o CL) for ScauC, XuetangX, and Amazon. Especially, the situation without the supplementation of student graph information (w/o student) is tested for the ScauC dataset. The results of the ablation experiments are shown in Table 6 and 7, where a β -value less than 0.05 indicates that there is a statistically significant difference between the EduLGCN and the recommended metrics of each variant in terms of their significance in the context of statistical testing.

The variant without the contrastive learning (w/o CL) was employed, retaining the information expansion module, making the EduLGCN structure closely resemble LGCAF. Further removal of the connection expansion module from EduLGCN (w/o info) resulted in a structure similar to LightGCN. It can be observed that the removal of the contrastive learning module led to the direct utilization of global node representations for similarity computation across the four metrics on the Amazon and XuetangX datasets, resulting in a decrease in performance due to the presence of global noise. Moreover, the removal of the connection expansion module resulted in a scenario where only the interaction history was available for model learning. A further decline in performance is caused by missing comprehensive representations.

The variant without student graph (w/o student) for the ScauC dataset includes only the contrastive learning module and the information expansion module. Due to the presence of a substantial number of Freshman courses with interaction history lengths less than five, the model could not effectively validate whether the node representations are learned, thus affecting the optimization performance in subsequent iterations. Consequently, a notable decrease in performance was observed under this circumstance.

4.4. Performance of EduLGCN with different features

In order to explore the importance of mutual information static feature selection in EduLGCN, we used different numbers of feature

combinations on different datasets in order to conduct the experiments. In the MovieLens dataset, there are four available item features: movie, genre, director, and star. Different combinations of features were selected for experimentation under the condition of the fixed selection of "Movie". There are five available features in the ScauC dataset and many combinations of features cannot be manually filtered, even with a fixed selection of "course" features. The experimental results on MovieLens and ScauC are illustrated in Fig. 4 for the three cases of the highest theoretical mutual information value, the lowest theoretical mutual information, and the best performing feature combination.

As depicted in Fig. 4, M1 denotes the situation with the lowest theoretical mutual information values, specifically indicating feature combinations below one or two-feature. M2 represents the situation with the highest theoretical mutual information values, encompassing full feature combinations or those greater than one. M3 represents the optimal performing feature combination.

Overall, a higher number of classifying features in a given type indicates a greater mutual information value between the feature and user characteristics. However, selecting the feature combination with the highest mutual information value does not necessarily yield the best results. This is because feature blurring may occur, causing difficulties in distinguishing the differences between two closely related challenging negative samples in the embedding space. For instance, the combination (college1, major1) and (college1, major2) might result in being lack of sufficient training samples for the model to differentiate between major1 and major2 effectively within the embedding space, which led to blurring.

Furthermore, inherent noise may be introduced from various scenarios. So, the combination of a few features with the highest mutual information value and minority features with the lowest mutual information value can often yield favourable results.

4.5. Validation on user cold start problem

In the context of recommendation systems, users who have fewer interactions with items are referred as "cold users". Recommending items to cold users who have limited interaction history is known as a "cold start problem". In educational scenarios, cold users specifically embody a freshman entering school without interaction with any courses, making it difficult to capture students' interests and preferences. The users in the XuetangX and ScauC datasets are grouped according to the length of their interaction history. The experimental results are shown in Fig. 5.

Different user groups are defined based on varying levels of user interaction. Group 0 represents users whose interaction frequencies

Table 7
The ablation results of EduLGCN and its variants on the ScauC.

Dataset	Metric	w/o CL	ρ -value	w/o info	ρ -value	w/o student	ρ -value	EduLGCN
ScauC	RMSE	0.1712	$1.10E-06$	0.1757	$3.26E-07$	0.1767	$3.14E-06$	0.1695
	NDCG@10	0.1094	$5.32E-07$	0.1162	$1.89E-07$	0.1066	$5.12E-07$	0.1277
	AUC@10	0.2648	$3.79E-07$	0.2613	$8.12E-07$	0.2220	$1.87E-05$	0.2924
ScauC	NDCG@20	0.1403	$2.22E-07$	0.1445	$2.36E-07$	0.1270	$1.22E-06$	0.1474

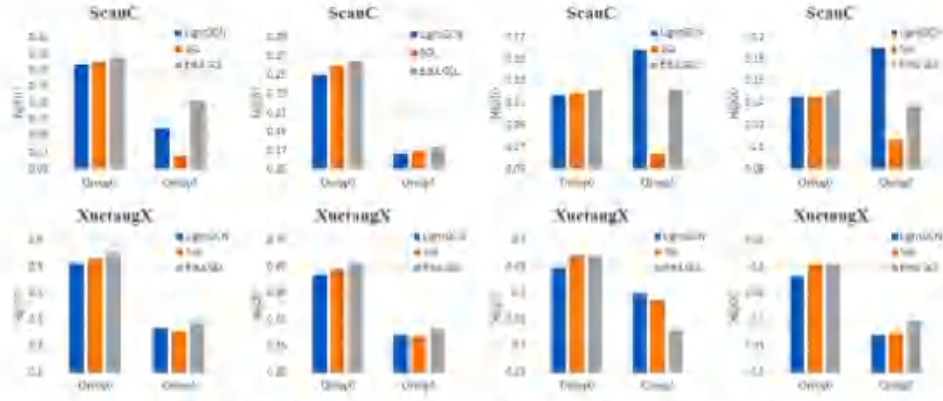


Fig. 5. The comparison results of various models on different user groups in the ScauC and XuetangX datasets.

range from 1 to 5 times, which corresponds to freshman in the educational scenario. Group 1 comprises users whose interaction frequencies range from 5 to 10 times, which represents a mix of seniors and some freshmen.

It is obvious in Fig. 3 that EduLGCN exhibits excellent performance in the freshman group's metrics for the ScauC dataset, which extends the relationship between student nodes using student clustering relationships and course static features. However, its performance in recommendation quality is weaker than that of the conventional LightGCN model for the senior group. Although EduLGCN recommends courses of interest to students, these courses might not accurately represent the students' true preferences. It could reflect differences between the inferred similar preferences of student clustering and the actual preferences learned by the model from existing data.

Even though EduLGCN is weaker than SGL in overall performance for XuetangX dataset, it can assist in solving the cold-start problem. Despite the relatively limited number of available course features for expansion, these features are helpful for connecting with cold-start users. Moreover, since the number of cold-start students in XuetangX is small, SGL's edge dropout has a limited impact on this user category. However, an interesting phenomenon arises in the senior group, that is, traditional models outperform contrastive learning models in recommendation quality. It could be attributed to the number of students in this range, as the limited additional information introduced by contrastive learning might result in redundancy for the recommendation task within this range.

4.6. Hyperparameter analysis for educational scenario

In order to investigate the effect of adding contrastive learning intensity on the performance of EduLGCN, the proportion of adding contrast learning λ_c is set as the set of $\{0.1, 0.1, 0.05, 0.01\}$ and the strength

of adding negative sample r is set as the set of $\{0.2, 0.175, 0.15, 0.1\}$. The experimental results with different proportions and strengths are demonstrated in Fig. 5. The comparison is made by fixing the same negative sampling strength while varying the proportions, and by fixing the same proportions while varying the strengths.

As shown in Fig. 6, when incorporating the same proportion of negative samples in the ScauC and XuetangX datasets, the model's performance tends to be improved as the proportion of contrastive learning decreases. It validates that adding a small amount of the global interest representation of students can promote the performance of recommendations. With a fixed proportion of contrastive learning, a higher strength of negative sample incorporation leads to more difficult negative samples being added. In this case, as the number of negative samples decreases, the model's performance also degrades. This phenomenon may be related to the sparsity of the educational dataset. Additionally, the issue of feature ambiguity caused by multiple feature combinations may affect the model's ability to distinguish difficult negative samples.

5. Conclusion

This study addresses the data sparsity and user cold-start issues faced by recommendation systems in educational scenarios, presenting a graph-based contrastive learning recommendation model named EduLGCN. The model employs the static features of students and courses to expand the sparse interaction data, adding more neighbor nodes to student nodes with fewer interactions, and supplementing their learning interest representations. It describes students' interest preferences from diverse perspectives based on the static features of courses, mitigating the user cold-start problem in recommendation systems. By introducing a contrastive learning method, the model integrates local and global student interest representations, alleviating the data sparsity

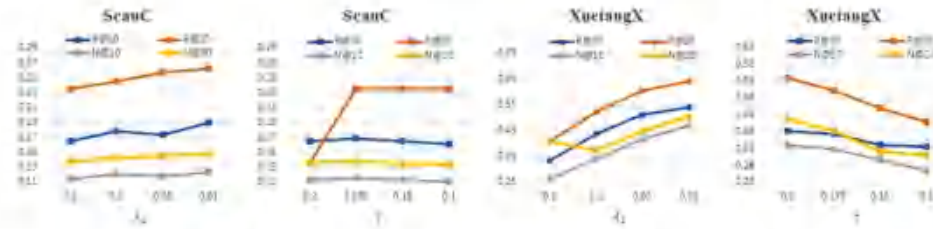


Fig. 6. The performance of EdatGCL with different parameters on the SeauC and XuefangX datasets.

issue in educational scenarios. This model explores more courses that students might find interesting, enhancing their learning motivation and enthusiasm. Unlike existing recommendation systems applied in educational scenarios, it utilizes static profile features that are easier to obtain and often overlooked, allowing the model to describe students' interest preferences from various angles. Experimental results on multiple public datasets from diverse domains and densities demonstrate the effectiveness of the proposed method. Furthermore, data collection and experimentation in real-world educational scenarios validate that the method can alleviate cold-start issues and assist students in better expanding their learning interests.

However, the experimental results of this study indicate that utilizing graph convolution and static features to expand interaction history does not yield improvements in datasets with only a single feature. Employing random walk or edge perturbation methods for generating global representations in heterogeneous graphs is a superior approach for such single-feature datasets. Future research will conduct specific experiments addressing the situation of single-feature datasets.

Furthermore, the proposed method in this study focuses solely on the expansion of students' interests, which belongs to the matching stage task in recommendation systems. Future research will explore the exploration of ranking and re-ranking stage tasks for addressing short-term interest changes of students.

CRedit authorship contribution statement

Yijun Zhao: Writing – original draft, Software, Methodology, Conceptualization. **Fajian Jiang:** Software, **Yin Pang:** Investigation, Formal analysis. **Yunxi Deng:** Investigation, Formal analysis. **Youyou Han:** Data curation. **Jinfeng Wang:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

This work is supported in part by Scientific Research Platforms and Projects of Guangdong Provincial Education Department (2023ZDZX 4002, 2021ZDZX1079), Guangzhou Key Laboratory of Intelligent Agriculture (2019 02010091). The authors thank all editors and reviewers for the suggestions and comments.

References

- [1] Fuzhen Wu, Yinyan Wu, *Graph-based graph representation learning*, *Arxiv Preprint arXiv:1808.08438*, 2018.
- [2] Anu Shrivastava, Sreyas Shridhar, Junyong Park, Arjun Gu, *Graph-based graph representation learning*, *Arxiv Preprint arXiv:1808.08438*, 2018.
- [3] Alexandros Kipf, Wendelin Wainwright, *Graph convolutional networks*, *Artificial Intelligence Review*, vol. 30, pp. 409–447, 2016.
- [4] Jiepei Chen, Xiang Chen, Wei Li, Peng Liang, Wenjie Liu, *Graph neural networks for recommendation systems: A survey*, *ACM Computing Surveys*, vol. 55, no. 1, pp. 1–36, 2022.
- [5] Sijun Luo, Jiarui Gao, *Graph neural networks for recommendation systems: A survey*, *Proceedings of the 2022 International ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 25–34.
- [6] Arvids Bagdasarian, Arseniy Zhukov, Arseniy Zhukov, *Graph neural networks for recommendation systems: A survey*, *Proceedings of the 2022 International Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3–5.
- [7] Li Wu, Xiangbo Hu, Wang Wang, Jun Wang, Xiang Wang, *A survey of recommendation system personalization from collaborative filtering to federated learning*, *International Journal of Intelligent Systems*, vol. 35, no. 12, pp. 1155–1172, 2020.
- [8] Li Wu, Junjie Li, Pengfei Bai, Jun Wang, Wang Wang, *Graph neural networks for recommendation systems: A survey*, *Proceedings of the 2022 International Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3–5.
- [9] Danyu Shi, Peng Wang, Jun Wang, *Graph neural networks for recommendation systems: A survey*, *Proceedings of the 2022 International Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3–5.
- [10] Wei Gu, Yuhao Zhou, *Graph neural networks for recommendation systems: A survey*, *Proceedings of the 2022 International Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3–5.
- [11] Li Guo, Li Wu, Pengfei Bai, Jun Wang, *Graph neural networks for recommendation systems: A survey*, *Proceedings of the 2022 International Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3–5.
- [12] Gaurav Srivastava, Anshu Kumar, *Graph neural networks for recommendation systems: A survey*, *Proceedings of the 2022 International Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3–5.
- [13] Delia Wang, Yang Zhang, Yi Zhu, *Graph neural networks for recommendation systems: A survey*, *Proceedings of the 2022 International Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3–5.
- [14] Hongyi Luo, Yang He, Michael Bressan, *Graph neural networks for recommendation systems: A survey*, *Proceedings of the 2022 International Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3–5.
- [15] Yuhao Zhou, Junjie Li, *Graph neural networks for recommendation systems: A survey*, *Proceedings of the 2022 International Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3–5.
- [16] Xiangbo Hu, Jun Wang, *Graph neural networks for recommendation systems: A survey*, *Proceedings of the 2022 International Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3–5.
- [17] Jun Xian, Han Yu, Xiangbo Hu, *Graph neural networks for recommendation systems: A survey*, *Proceedings of the 2022 International Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3–5.
- [18] Junjie Li, Xiangbo Hu, *Graph neural networks for recommendation systems: A survey*, *Proceedings of the 2022 International Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3–5.

- [19] Xiang Wang, Xingran He, Meng Wang, Fuli Feng, Tai-Seng Chua, Neural graph collaborative filtering, in: Proceedings of the 42nd international ACM SIGIR conference on Research and development in information retrieval, 2019, pp. 165–174.
- [20] Xiangnan He, Kuan Deng, Yang Wang, Yan Li, Yongsheng Zhang, Meng Wang, LightGC: Simplifying and powering graph convolution network for recommendation, in: Proceedings of the 43rd International ACM SIGIR conference on research and development in information retrieval, 2022, pp. 639–648.
- [21] Rex Ying, Ruining He, Kaifeng Chen, Peng Eksombatchai, William L. Hamilton, Jure Leskovec, Graph convolutional neural networks for web-scale recommender systems, in: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, 2018, pp. 974–983.
- [22] Will Hamilton, Zhifu Yang, Jure Leskovec, Inductive representation learning on large graphs, in: Advances in neural information processing systems, vol.30, 2017.
- [23] Le Wu, Xieyi Sun, Yanjie Fu, Wotang Hong, Xing Wang, Meng Wang, A neural influence diffusion model for social recommendation, in: Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, 2019, pp. 235–244.
- [24] Tianhang Yin, Xinyang Yu, Derek Zhiyuan Cheng, Yishi Yu, Ting Chen, Aditya Menon, Lichun Huang, Ed H Chi, Steve Joo, Jinyi Kang, et al., Self-supervised learning for large-scale item recommendations, in: Proceedings of the 20th ACM International Conference on Information & Knowledge Management, 2021, pp. 4321–4330.
- [25] Jianan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jiaxian Lian, Xing Xie, Self-supervised graph learning for recommendation, in: Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval, 2021, pp. 726–735.
- [26] Jianliang Yu, Huihui Yin, Xian Xun, Tong Chen, Lichen Cui, Quoc Van Hung Nguyen, Are graph augmentations necessary? simple graph contrastive learning for recommendation, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 1294–1303.
- [27] Zhan Lin, Guanglin Tian, Yupeng Hou, Wayne Xin Zhao, Improving graph collaborative filtering with neighborhood-enriched contrastive learning, in: Proceedings of the ACM Web Conference 2022, 2022, pp. 2320–2329.
- [28] Lingshao Xia, Chao Huang, Yang Xu, Jianshi Zhao, Lowsi Yu, Jianyi Huang, Hypergraph contrastive collaborative filtering, in: Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval, 2022, pp. 70–79.
- [29] Xuheng Chi, Chao Huang, Lingshao Xia, Xuhui Ren, LightGC: Simple yet effective graph contrastive learning for recommendation, 2023, arXiv preprint arXiv:2302.08191.
- [30] Xiaohui Zou, A survey on application of knowledge graph, J. Phys. Conf. Ser. 1487 (2020) 012016.
- [31] Hao Zhang, Heng Yang, Tao Huang, Guoqing Zhao, DBNCR: Personalized course recommendation system based on DBN in MOOC environment, in: 2017 International symposium on educational technology, ISET, 2017, pp. 106–108.
- [32] Jie Lian, Wei Zhou, Fengli Luo, Junhao Wen, Min Gao, Xuhua Li, Jun Zeng, SocialIGN: Light graph convolution network for social recommendation, Inform. Sci. 589 (2022) 595–607.
- [33] Zhihui Zhou, Lili Zhang, Ning Yang, Contrastive collaborative filtering for cold-start item recommendation, 2023, arXiv preprint arXiv:2302.02131.
- [34] Yiqun Zhao, Fajian Jiang, Jiafeng Wang, MPGL: Light graph contrast learning with multi-format information for recommendation, in: 2023 International Joint Conference on Neural Networks, IJCNN, 2023, pp. 01–08.
- [35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, in: Advances in neural information processing systems, vol.30, 2017.
- [36] Minghao Tang, Peng Zhou, Shaobo Li, Yuanmeng Zhang, Jianyi He, Aoxi Zhang, Multi-head multimodal deep matrix recommendation network, Knowl.-Based Syst. 130(5) 110989.
- [37] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, Lars Schmidt-Thieme, BPR: Bayesian personalized ranking from implicit feedback, 2012, arXiv preprint arXiv:1205.2618.
- [38] Michael Gormann, Aapo Hyvärinen, Noise-contrastive estimation: A new estimation principle for unnormalized statistical models, in: Proceedings of the thirtieth international conference on artificial intelligence and statistics, 2016, pp. 297–304.
- [39] Dinghui Mei, Niu Huang, Xin Li, Light graph convolutional collaborative filtering with multi-object information, IEEE Access 9 (2021) 34433–34441.
- [40] F. Maxwell Harper, Joseph A. Konstan, The movielens dataset: history and context, ACM Trans. Interact. Intell. Syst. (TIIS) 5 (4) (2015) 1–19.
- [41] Raining He, Julian McAuley, Up and down: Modelling the visual evolution of fashion trends with one-class collaborative filtering, in: Proceedings of the 25th international conference on world wide web, 2016, pp. 507–517.
- [42] Jang Zhang, Bowen Hao, Be Chen, Cuiqing Li, Heng Chen, Jianing Sun, Hierarchical reinforcement learning for course recommendation in MOOCs, in: Proceedings of the AAAI conference on artificial intelligence, 2019, pp. 426–442.
- [43] Han Zhu, Xiang Li, Pengye Zhang, Guozhong Li, Ji He, Han Li, Kun Gu, Learning tree-based deep model for recommender systems, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 1079–1088.
- [44] Dongha Lee, Seongju Kang, Hyeonjun Ju, Chanyoung Park, Hwanbyu Yu, Bootstrapping user and item representations for one-class collaborative filtering, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 317–328.
- [45] Jieyang Tang, Chib-Ming Chen, Chuan-Hsi Wang, Ming-Feng Tsai, HDP-IR: High-order proximity for implicit recommendation, in: Proceedings of the 12th ACM conference on recommender systems, 2018, pp. 140–144.

12. 教改论文：计算机类专业课程思政探索与实践——以操作系统课程为例



目次

刊名: **软件导刊**

卷期: 2022年第21卷第7期

创刊日期: 2002年10月20日

主管: 湖北省科学技术厅

主办: 湖北省科技信息研究院

编辑: 《软件导刊》编辑部

出版: 湖北省科技传媒有限责任公司

主编: 何美祥

执行主编: 何丽

责任编辑: 孙娟

黄健

尹晨茹

刘嘉文

国际标准连续出版物号: ISSN1672-7800

国内统一连续出版物号: CN42-1671/TP

地址: 武汉市洪山路2号湖北科教大厦D座5楼

邮政编码: 430071

电话: (027)87891823 87821070

网址: www.rjtk.org(在线投稿)

电子邮箱: softwareguide@163.com

微信公众号: 软件导刊杂志



2021中国高校计算机教育大会(CCEC2021)专题

面向代码质量提升的软件工程能力素质培养	陈立言,董 焱,尹良洋,等(1)
面向产业需求的高校人工智能人才培养研究	齐 帅,关东海,闫建秋,等(16)
工程认证背景下的大数据创新实验室建设	何金凤,马天宏,杜明珠(12)
物联网工程专业实践教学体系构建研究	陈 辉,刘向举,蒋社志,等(16)
新工科背景下NoSQL数据库技术课程建设探索与实践	袁燕妮,吴 斌,王 楠(21)
新时代下计算机网络课程“三新”探索及实施	孙伟峰,张丽君,姚 琳(27)
翻转课堂教学改革当议	张 策,初相辉,王 峰,等(33)
数字图像处理课程的混合式教学方案	姜秀艳,徐迎晚,张 睿(41)
混合教学模式下Python程序设计教学改革探索	李向群,许新征,高 颖(45)
混合式教学在算法分析与设计课程中的应用与实践	袁梁华,高万林,孙瑞志,等(49)
多种教学模式融合的计算机课程实践研究	王圣伟,李桂林,魏江源,等(53)
基于云课堂的“互联网+”混合式教学改革研究	毛宇亮,朱 丽,杜 波(58)
“三点一贯穿”培育线上线下混合式金课	王 妍,齐苏航(63)
基于Excel的模糊筛选系统设计与实现	王乾乾(68)
OBE理念下的串口实验改革探索	李 海,张 欣(75)
基于“产教研思”的计算思维培养教学模式探讨	郭万江,杨金翠,孙鹏飞,等(79)
操作系统课程思政探索与实践	李文生,叶 文,刘德鸿(83)
面向一流课程目标的图形学课程思政设计	马 天,张杰慧,符玉梅,等(89)
融合思政元素的数据结构实验教学改革与实践	刘伯红,吴思远,夏 英(93)
面向新工科的Python程序设计课程思政教学探索	李锐英,何首武,陈 佳,等(98)
计算机类专业课程思政探索与实践	王金凤,张 静,张丽霞,等(103)
“CDO+思政”的Java程序设计改革与实践	谢艳娇,陈俊杰,左东石(109)
疫情之下留学生编程类课程线上教学方法探究	柏 涛,王书婷,洪海波(114)
基于PBL教学法的统计学习课程教学改革与实践	陈 娟,陈文字,文 泉(118)
OBE理念下的操作系统课程教学改革与实践	李 辉,李 征,王 鹏,等(122)

地方高校本科一流课程建设探索	张 健, 赵娟兰, 毕 卉, 等(126)
计算机组成原理教新论文主题比较分析	刘建春, 郑 婷, 廖国琼, 等(130)
面向应用型人才培养的计算机通识教育实践	张根南, 马杨辉, 马伟特(137)
基于系统能力培养的计算机课程体系改革与实践	张 旭, 胡东华, 朱付保(142)
基于资源整合的案例驱动式计算机网络教学改革	王 霞, 谢雄奇, 谷瑞军, 等(147)
高职计算机应用基础课程思政教学设计与实践	徐草娟, 陆 璐, 梅绍斌(151)
基于BOPPPS教学模型的程序设计课程线上教学实践	许智宏, 吕 华, 马建红, 等(157)
现代产业学院工程实践课程改革与探索	张向群, 张 柯, 杜根远(161)
基于问题导向—互动建构的课堂教学模式研究与探索	谷 斌, 宁 彬, 胡春阳, 等(165)
基于SpringBoot的在线教育系统研发及应用实践	陶 磊, 谢仁平(170)
2021中国计算机研究生教育大会(CCGEC2021)专题	
网络空间安全学科人才培养机制探索	李翰青, 庄晓强, 廖一中, 等(175)
SAR遥感技术专业创新型人才研修培养研究	李 宁, 曹 琳, 郭磊亮, 等(180)
新工科背景下电子信息人才培养模式探索	毕 纯, 李春燕, 张于龙(184)
硕士研究生培养质量影响因素分析	陈平华, 吴亚迪, 罗汉元, 等(188)
计算机研究生教育现状及趋势研究	袁 科, 程自伟, 王树霖, 等(195)
面向科研能力培养的研究生专业课程改革与实践	罗 红, 孙 岩(203)
物联网课程教学创优探索与实践	张 策, 吕为工(208)
基于知识图谱的课程思政素材库构建	汤宇轩, 齐 健, 申彦明, 等(214)
融合思政教育的“新工科”研究生创新思维培养模式	谭 云, 辜悦华, 向旭宇(220)
基于产教融合的实践教学基地建设探索	雷 茹, 田 琦, 周 珂, 等(224)
计算机类研究生课程建设改革与探索	钱文华, 姜 磊, 李敏华, 等(228)
问答式教学方法设计	朱季平, 韩且亮, 吴旭光(231)
面向研究生的凸优化课程教学设计与实践	魏 倩, 尚 可, 金 勇, 等(236)
面向研究生思辨能力培养的数据挖掘课程教学实践	叶志伟, 高 松, 林 旭, 等(240)
控制工程专业研究生教育综合改革试点建设	胡振涛, 杨琳琳, 周 琳, 等(244)
科普激励师范院校计算机研究生实践能力培养	李 燕, 郝培勇, 孙艳歌, 等(248)

中国计算机学会会刊
 计算领域高质量科技期刊
 中国核心期刊(遴选)数据库收录期刊
 科学引文数据库(SCI)源期刊
 中国学术期刊(光盘版)全文收录期刊
 中文科技期刊数据库(全文版)收录期刊
 中国学术期刊综合评价数据库来源期刊
 《中国学术期刊影响因子年报》统计源期刊
 湖北省优秀科技期刊

印刷:武汉市洪林印务有限公司
 发行:武汉报刊发行局
 邮发代号:38-431
 订购、零售:全国各地邮局
 中科期刊淘宝店、微店



广告经营许可证:4201003400008
 出版日期:2022年7月15日

值班编辑:孙 凯
 翻译:尹晨茹

计算机类专业课程思政探索与实践

——以操作系统课程为例

王金凤, 张 猜, 张丽霞, 房三虎, 孙微微

(华南农业大学 数学与信息学院, 广东 广州 510642)

摘 要:在全国上下课程思政总体布局的大背景下,如何在专业类课程中融入思政元素成为一线教师需要思考并付诸行动的课题。以计算机专业核心课程操作系统为例,从省级一流本科线上线下混合课程建设出发,对课程思政环节的改革与探索进行总结和介绍,从教学理念、教学模式、教学内容等方面说明了计算机类专业课程与思政不可分割的必然联系,以及实施专业思政教育的必要性。课程线上运营5学期,选课人数已达1 821人。通过实践教学,使学生能够通过操作系统的工作原理推及自身,深入思考如何树立社会主义核心价值观,强化软件强国的思想,从而培养专业基础扎实、动手能力强、有情怀、有温度的计算机类专业创新型人才。

关键词: 计算机类专业;课程思政;操作系统;软件强国

DOI: 10.11907/rjdk.212070

开放科学(资源服务)标识码(OSID):

中图分类号:G434

文献标识码:A

文章编号:1672-7800(2022)007-0103-06



Exploration and Practice of Ideological and Political Course Education in Computer Major

——Taking Operating System Course as an Example

WANG Jin-feng, ZHANG Cai, ZHANG Li-xia, FANG San-hu, SUN Wei-wei

(College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China)

Abstract: Under the background of the overall layout of ideological and political education in the whole country, how to integrate the ideological and political elements into the professional courses has become a topic that front-line teachers need to think about and put into action. As a core course of computer major, operating system was taken as an example, which is the provincial first-class undergraduate course with online-offline hybrid way. Summarize and introduce the reform and exploration of the ideological and political aspects of the course. It illustrates the inextricable connection between the courses of computer major and the ideological and political aspects from the aspects of teaching philosophy, teaching mode and teaching content, and the necessity of implementing professional ideological and political education. At the same time, it improved the effect of online teaching, which has been operated for 5 terms and enroll up to 1 821 students. Students can infer OS principles to them and think deeply about how to establish socialist values, strengthen the idea of a software power and foster innovative talents of computer major with strong professional skill, active manipulative ability, feelings and temperature.

Key Words: computer major; ideological and political education of curriculum; operating system; software power

收稿日期:2021-08-17

基金项目:教育部协同育人项目(201901134024);广东省一流本科课程建设项目(粤教高函[2020]16号);广东省高等学校质量工程——《操作系统》开放课程建设项目(粤教高函[2018]179号-24);广东省教育科学规划课题(2021GXJK257);华南农业大学百个号任教师党支部课程思政精品示范课建设项目(2021)

作者简介:王金凤(1978-),女,华南农业大学数学与信息学院副教授,硕士生导师,研究方向为机器学习、教育信息化;张猜(1986-),男,华南农业大学数学与信息学院副教授,研究方向为量子计算;张丽霞(1978-),女,华南农业大学数学与信息学院副教授,研究方向为图像识别;房三虎(1977-),男,华南农业大学数学与信息学院副研究员,研究方向为高等教育管理;孙微微(1971-),女,华南农业大学数学与信息学院副教授,研究方向为信息安全。本文通讯作者:孙微微。

0 引言

2016年12月,在全国高校思想政治工作会议上,习近平总书记强调将思想政治工作中的传统优势与现代信息技术相结合对于思想政治工作的重要性,要求灵活运用新媒体技术,以增强思想政治教育的吸引力和时代感^[1]。在党中央的号召与推动下,全国高校掀起了课程思政的高潮。2017年,中共中央、国务院《关于加强和改进新形势下高校思想政治工作的意见》提出坚持全员全过程全方位育人(简称“三全育人”)的要求。高校实施“三全育人”综合改革,是改进新形势下高校思想政治工作,全面落实立德树人根本任务的战略举措。为了使高校教师有据可依,不偏离轨道,2020年6月,教育部又印发了《高等学校课程思政建设指导纲要》(以下简称《纲要》),强调要把思想政治教育贯穿人才培养体系,全面推进高校课程思政建设,发挥好每门课程的育人作用,提高人才培养质量。

1 相关工作

立德树人是高校的根本任务,做好德育工作是每位高校教师的义务和使命。课程思政是高校立德树人的突破口和抓手,是培养社会主义核心价值观的有效途径,有助于提高学生综合素质,促进学生德智体美劳全面发展^[2]。在当前疫情防控的关键时期,专业课教师如何守好“思政政治”的建设岗,种好“立德树人”责任田,成为新时代高校工作的新挑战。立德树人,在思政课程中引领,在通识类课程中浸润,在社科类课程中深化^[3]。在理工类课程中更应体现科技兴国,实干兴邦的精髓,而理工类课程的思政融入是课程思政实施过程中的难点。有学者基于理工科课程特点,从教学目标入手,提出革新教学理念,调整教学方式,重塑教学内容的方法,为理工科课程思政建设指明了方向^[4]。即便如此,计算机类专业课程在教师和学生印象中,与思政教育还是有一定距离。

近两年关于计算机类专业课程的思政研究开始初见成效,包括数据结构^[5]、程序设计^[6]、软件需求分析^[7]等课程,给予了教育工作者以改革启示。操作系统是计算机类专业课程的核心基础课程,对于计算机类各专业的基础课和实践课起到了承上启下的作用。一些学者从不同角度出发,针对操作系统的思政改革开展了研究。如文献[9]中强化了操作系统与关联课程之间的融合,调整了课程内容,以提升学生的系统应用能力;文献[10]将微课教学实践与课程思政相融合,基于BOPPPS教学模式,实现显性专业知识教育与隐性思政教育的有机结合。以往多从某一点着手进行思政融入研究,而文献[11]从课程目标、组织方式和评价体系优化等方面探索了以混合式教学改革驱动课程思政建设的可能路径。

本文借鉴相关专业课程的思政路径,基于华南农业大学课程团队自建的慕课平台,依托省级一流本科课程建设项目,以线上线下混合授课的方式开展实践了社会主义核心价值观、大国工匠精神、自主可控国产软件强国等思政思想,并全面介绍了思政融入的关键点,以期能不断探索提升课程思政建设水平,力争让思政从细处处渗透,从大局上把控,全程、全方位地融入课程内容,潜移默化润物无声,从而培养出具有新时代中国特色社会主义价值观的优秀人才。本文分别从教学理念、教学模式、教学内容几个方面展开课程思政工作,并取得了良好的教学效果。

2 思政融入专业措施

在传统的专业课程上,教师们对思政教育的重要性认识不够,通常认为思政只是思政课程的任务,与专业课程无关,从而错失了全程、全方位开展思政教育,培养高素质人才的良机。对于计算机类专业学生而言,近80%的课程都是专业相关课程,因此在专业课程中融入思政环节是必要的课程改革举措。同时,在专业课程讲解过程中,穿插一些传统文化、励志人物与时政要闻等内容,也会使得原本枯燥的专业课堂变得生动、活泼,有温度,大大提高了学生的学习兴趣。本文以操作系统课程为例,开展了专业课程思政融入的实践与探索。

2.1 教学理念

课程创新从“以本为本”出发,转换课程主体角色,从传统的教师、教材为主体向以学生为主体转换,进行个性化教学。教师掌控方向,遵循创新性、高阶性原则,创新教学模式。相应环节的思政要点由学生主动体会和感悟,自始至终体现了实践出真知的道理,并令学生自主完成部分知识点的学习,从而提升了学习积极性和学习效果。

在充分利用线上资源的前提下,不脱离教师的指导主线,发挥学生的主观能动性,打破原有模式的依赖因素,形成“一主线两主体三点评”模式,贯穿整个课程授课过程。一主线是课—系列线上线下结合、穿插交替的线上主线;线下指导+线上示范+线下解答+线上答疑+线下点评+线上展示;两主体是教师从传统的教师为主转变为教师和学生同为主体,相互影响,相互反馈;三点评是指课程最后对学生的考核包括教师对学生创新性的评价、学生对自己的点评、学生之间的互评,形成不同层次、不同角度、多维度的学生水平考核机制。同时,还要遵循课程思政润物无声的理念,时刻将思政教育贯穿整个课程教学。

教学理念实施方案如图1所示。首先,经过课程组讨论,确定课程改革沿着“一主线两主体三点评”的方向展开深入调研,广泛征集参与主体意见,制定课程的主要内容和考核范围,形成完整的课程改革体系;然后,将理论体系应用于具体课程,采用主线循环的技术路线,完成一个周期的课程实践;最后,通过学生成绩数据和主体反馈数据

对学生的创新能力进行评估,发现问题,并对理论体系进行完善,以备在学科范围内推广。



Fig. 1 Implementation plan of teaching concept
图1 教学理念实施方案

2.2 教学方法创新

本文选取的操作系统课程对融入课程思政的线上线下一混合授课方法进行了实践与验证,从传统的课堂教学模式向线上线下混合模式转变,依托智慧树平台上的慕课视频创建翻转课堂,实现线上预习、复习、作业与小测等,并在线下课堂完成知识点串讲、习题答疑、小组讨论等环节,线上线下互为补充、相辅相成。教学方法路线图如图2所示。

(1) 对下示例课程操作系统而言,教师需列出能够覆盖各个知识点的实验题目,同时对每个题目的设计要能充分调动学生的学习兴趣和创新动力。教师在线下完成题目布置和环境配置指导工作,这一环节充分显示了教师的重要作用,线上线下混合式教学不等于放任不管。

(2) 学生在自行理解教师讲解内容的基础上,可在线上观看实验视频,学习其中的实验技巧。学生可自由选择熟悉的编程语言和环境,在此过程中如有疑问可在下一环节进行咨询,以调动学生积极探索知识、勇于提出并解决问题的能力。

(3) 经过第二阶段的线上学习,部分学生可能会产生一些问题,教师可通过小班进行线下答疑,也可借助多媒体工具进行答疑,为学生提供必要的建议和支持。这一环节可锻炼学生的沟通能力,因此鼓励学生积极与教师进行互动。

(4) 学生进入本文构建的基于云平台的个人实践习题库,发挥自己的创新能力,针对具体题目进行设计与实现。在每位学生完成的题目之后可配备自我介绍和自我评价选项,供学生自己打分,完成考核第一部分内容,以培养学生勇敢展现自我、实事求是自我评价的精神。

(5) 带有自我评价标签的学生项目可推送给任课教师进行复查,同时召开现场答辩会,让学生有机会充分展现自己的成果,由教师点评打分。该环节可实现教师与学生对项目认知对接与评价信息交换。

(6) 将各个学生的作品在系统平台展示区进行公开展示,由同学观看学习,同时互相点评,打分,最后综合三方评定结果输出总评成绩。该环节体现了批评与自我批评的重要性。

在课程实施的各个环节中都需要现代信息手段和工具辅助。因此,在课程建设过程中,本文充分利用现代信息技术手段辅助线上线下教学的开展,主要包含3种形式:①QQ公共社群,开展线上答疑等活动,方便快捷;②线上慕课平台,满足学生自主学习的需求;③腾讯课堂:教师与学生面对面直播授课的工具,以提高学习效率。将现代信息技术融入教学的具体方法如图3所示。

2.3 教学内容

2019年8月,中共中央办公厅、国务院办公厅印发《关于深化新时代学校思想政治理论课改革创新若干意见》,要求“深度挖掘高校各学科门类专业课程和中小学语文、历史、地理、体育、艺术等所有课程蕴含的思想政治教育资源,解决好各类课程与思政课相互配合的问题,发挥



Fig. 2 Roadmap of teaching method
图2 教学方法路线图



Fig. 3 Integrating modern information technology into teaching methods
图3 现代信息技术融入教学方法

所有课程的育人功能,构建全面覆盖、类型丰富、层次递进、相互支撑的课程体系”。传统专业课程中注重专业知识的直接传递,而本课程的创新在于把思政思政融入各个章节,深度挖掘该专业课程体系中所蕴含的思想价值和精髓内涵,激励学生锤炼本领,掌握核心技术,培养爱国情怀,从而拓展专业课程广度、深度和温度。

在本课程教学过程中,首先从各章节内容出发,抓住重点,引导学生展开思考,用哲学的思想和眼光学习专业知识,提高学生的辩证思维能力,树立正确的社会主义价值观,培养爱国主义情怀,反过来也会促进对专业课的学习。根据不同章节内容,在各个教学环节融入思政观点。操作系统包括五大部分,分别是操作系统概论、进程、内存管理、调度、输入输出设备及文件系统。课程思政渗透点如图4所示。

以下案例进行说明:

例1:在首次启蒙课上,本文从操作系统实际应用角度切入。国家当前倡导开展自主可控软件的研究与开发,以摆脱国外在软件行业的垄断和控制,因此首先需要从计算机操作系统入手,引导学生思考如何研发我国自主可控的操作系统软件。在本课程的慕课中加入了现代操作系统介绍,其中重点介绍了华为公司推出的鸿蒙操作系统,这是我国软件企业研发出的一款极具发展前景的国产操作系统,因此在课堂上引起了学生们的浓厚兴趣,激发了学生们的爱国情怀,立志为打造中国自主可控软件而奋发学习。

例2:操作系统课程的第一部分是进程概念,该概念对于计算机专业学生来说是全新的,在其知识背景中只知道什么是程序,而进程就是静态的程序代码进入计算机系统后得以动态执行的身份,才有可能输出执行结果。此时,可将思政观点“做事必须付诸行动才能卓有成效”融入其中,不但告知了学生日常做事时需要牢记的法理,而且反过来让学生深刻体会到了进程与程序的本质区别,一语两得。

例3:在讲到操作系统中最重要的一个特殊变量“信号量”时,可以首先介绍信号量的提出者,即荷兰科学家迪杰克斯(Dijkstra)。学生们对于Dijkstra都是熟知的,但

对于这位传奇的结构化程序之父却知之甚少。首先简要介绍Dijkstra的生平和求学经历,引发了学生们阵阵感叹,起到了励志效果,从而激发了学生对于本节知识点“信号量”的求知欲望。同时还可以不失良机地告诉他们,科学家离我们并不遥远,就像本校的卢永根院士几十年如一日地为科学事业作出贡献,母校的钟海山院士八十岁高龄仍奋战在抗疫一线,无私无畏。因此我们更应该奋发图强,刻苦准备,提升思想和专业素质。

诸如此类思政内容可随时加入到专业课程教学中,自然科学与社会哲理是密不可分的,既可以讲授经久传世的哲理,也可以见缝插针地融入身边发生的要闻,做到全方位渗透,全程跟进,提升所有人的思政水平。

2.4 实验教学

计算机类专业是必须将理论与实践相结合的专业。在每门课程教学中,坚持技能培养导向,让理论指导实践,将学与行有机结合,将情的现实、情的理论融入教学中,让课堂变得有温度、有质量。通过在理论课堂讲授知识体系,实验课堂进行素质拓展,令学生将两项内容兼容并内化成适合社会需求的核心竞争力。

操作系统课程的实验环节依托教育部协同育人项目——基于鲲鹏的操作系统实践教学改革,采用容器技术打造实验云平台。山华为提供国产鲲鹏服务器,并安装国产服务器操作系统CentOS,可让学生在平台上切身感受到自主可控系统软件的优势和发展空间,为其将来促进国家在软件领域的发展奠定良好的思政基础。同时,在实验云平台上展开实验,让学生践行理论联系实际的思政观点,从动手操作中体会基础知识的重要性,从中发现问题并提出相应解决方案,为培养创新型人才奠定基础。

总之,整个实验教学过程本身是一项重要的思政教育活动。同理,可将此模式推广到其他计算机专业课程,让学生们首先对我们的大国制造有信心、有期盼,才能促使学生积极动手进行实践操作。在这种积极向上的心态下展开的学习,也才是自主、高效、有质量的。

3 课程思政实施成效

通过在专业课中融入思政环节,不但培养了学生的思政情怀,而且提升了授课教师自身的政治素质。尤其对于计算机类专业的教师和学生,需时刻牢记科技兴国的使命,因此需从细微处做起,达到全方位的思政教育渗透,专业技术与课程思政共荣共生的局面。

3.1 师生同步思政

课程思政的实施目的是培育与践行社会主义核心价值观,提高个人的爱国、敬业、诚信、友善修养,弘扬中华优秀传统文化,实现国家富强、人民富足、社会和谐的局面。因此,教育者需先接受教育,通过课程思政的培训和规范要求,更好地理解课程思政实施的必要性,将自身树立的

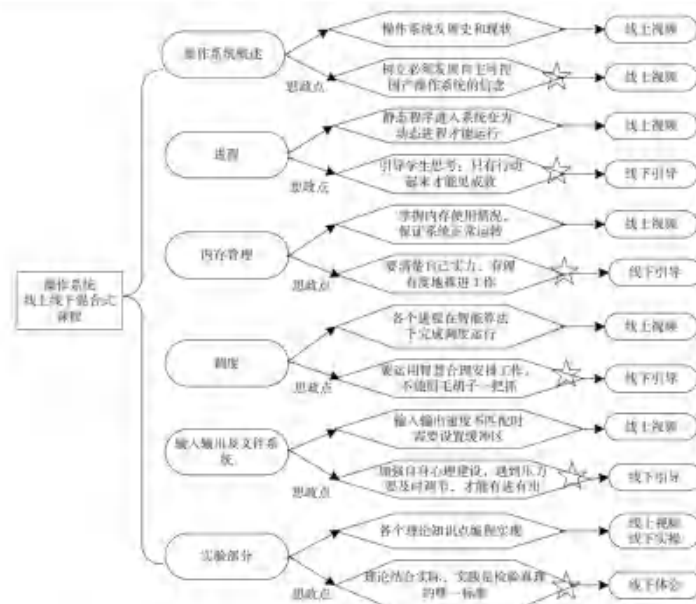


Fig. 4 Ideological and political points of each chapter

图4 各个章节思政点

正确价值观伴随着专业知识传递给学生,从而更好地落实立德树人根本任务。操作系统课程教师已掌握了扎实的专业知识,必然深刻理解操作系统各个模块的工作原理,而这些原理与人们的日常生活紧密相关,例如先来先服务(FIFS)的调度策略,因此可以在授课过程中恰当举例,帮助学生理解专业知识。

3.2 辩证思维悄悄渗透

自然辩证法与自然科学相互作用,相互促进,共同发展。当一个计算机专家研究分析一个计算机问题时,不可避免地要思考其哲学层次的问题,并依据哲学的思维方式去解决问题。哲学层次问题的解决,也可以从自然科学的研究中去总结、归纳。很多自然辩证法的问题,在自然科学中得到了解决。在此方面,图灵可称为计算机学界的哲学大师^[2]。因此,在计算机专业知识讲授过程中,所隐含的哲学道理也会悄然渗透其中。教师在不断揣摩理解相关知识时,也会深刻体会其中的辩证哲理,提高思政能力。

3.3 科技兴国之己任

习近平总书记指出,科学技术是第一生产力,创新是引领发展的第一动力。教师作为在授业解惑第一线的科技工作者,更应该弘扬科学精神,传播科学思想,倡导科学方法,为增强公众科学素质,促进科学成果共享、推动构建人类命运共同体作出贡献。通过专业知识与思政教育的

结合,让学生们时刻牢记以科技兴国之己任,提升个人的专业素养和思政素养,做一名能够担当社会主义现代化强国建设重任的合格接班人。

4 结语

本文主要依托省级一流课程的教学改革实践经验,从操作系统课程的教学理念、教学内容、实施方法等层面展开,介绍如何融入思政课程,同时对于其他相关计算机专业课程建设具有一定指导与借鉴意义。在未来的教学中,可以不断结合时政,加强思政点的渗透,继续加强对学生专业技能和爱国情怀的培养。

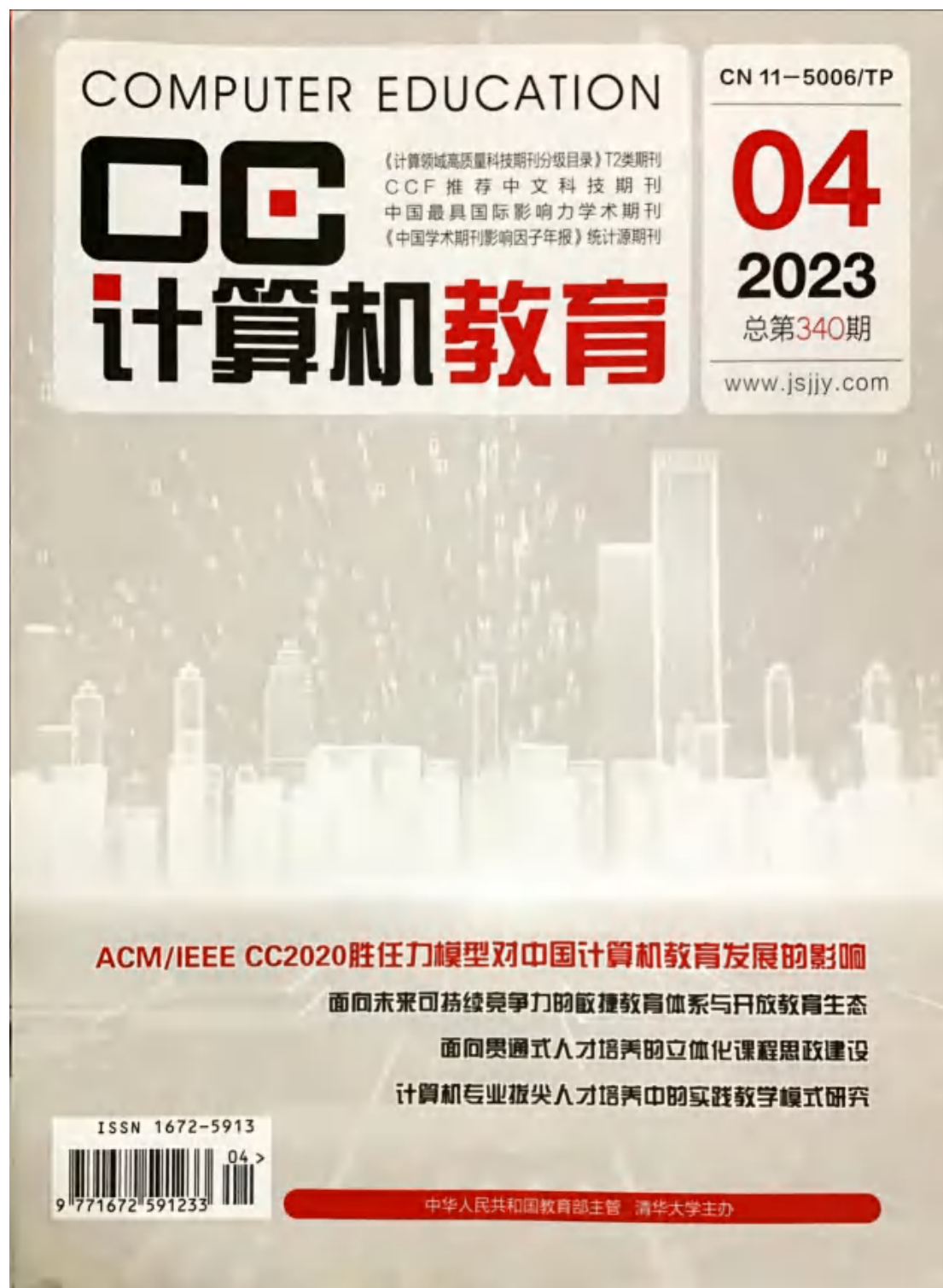
参考文献:

- [1] People's Daily, Xi Jinping stressed in the national conference on ideological and political work in colleges and universities: through ideological and political work throughout the whole process of education and teaching, create a new prospect for the development of higher education in China. [N]. People's Daily, 2016-12-09(1).
人民日报:习近平在全国高校思想政治工作会议上强调:把思政课程贯穿所有专业课程教学全过程,开创我国高等教育事业发展新局面[N]. 人民日报,2016-12-09(1).
- [2] E-AS D Y., ZHANG A D. Curriculum thought and politics: an inevitable choice to effectively play the role of the main channel of classroom education.

- tion[J]. Leading Journal of Ideological and Theoretical Education, 2017(1):31-34.
高德毅, 宗爱东. 课程思政: 有规定解课堂育人主渠道作用的必然选择[J]. 思想理论教育导刊, 2017(1): 31-34.
- [3] GAO D Y, ZONG A D. From ideological and political curriculum to curriculum ideological and political education: constructing the curriculum system of Ideological and political education in colleges and universities from a strategic perspective [J]. China Higher Education, 2017(1): 43-46.
高德毅, 宗爱东. 从思政课程到课程思政: 从战略高度构建高校思想政治课程思政体系[J]. 中国高等教育, 2017(1): 43-46.
- [4] LI J Y. Ideological and politics and emotion in science and engineering courses[J]. China University Teaching, 2019(12):20-23.
李俊琦. 理工科课程中的思政与情感[J]. 中国大学教学, 2019(12): 20-23.
- [5] WANG X Y, PAN Y Q. Design and practice of ideological and political teaching of data structure course [J]. Computer Education, 2021(1): 97-101.
王新宇, 潘雨青. 数据结构课程思政教学设计与实践[J]. 计算机教育, 2021(1): 97-101.
- [6] GUO Y Y. Exploration and practice of ideological and political implementation path of computer professional courses [J]. Computer Education, 2021(1): 80-85.
郭艳燕. 计算机类专业课程思政实施路径探索与实践[J]. 计算机教育, 2021(1): 80-85.
- [7] DAI D D, TANG W M. Method and practice of the ideological and political course of Python programming [J]. Software Guide, 2021, 20(7): 202-206.
代丹丹, 唐万梅. Python 程序设计课程思政方法与实践[J]. 软件导刊, 2021, 20(7): 202-206.
- [8] SHAO K, HUO X, GONG L. Practice of ideological and political teaching reform in software requirement analysis [J]. Software Guide, 2020, 19(12): 101-103.
邵望, 霍星, 巩琳. 软件需求分析课程思政教学改革实践[J]. 软件导刊, 2020, 19(12): 101-103.
- [9] HE J Y. The Exploration and practice of ideological politics——taking the computer professional course operating system as an example [J]. The Guide of Science & Education: Electronic Edition, 2021(1): 137-138.
何静敏. 新时代下“课程思政”的探索与实践——以计算机专业课程《操作系统》为例[J]. 科技导刊: 电子版, 2021(1): 137-138.
- [10] DANG X Y, LI Z L, SUN J P. Application of the curriculum ideology and politics integrated into operating system micro-class teaching [J]. Office Informatization, 2021(26): 57-59.
党向薇, 李子龙, 孙金萍. 思政教育融入《操作系统》微课教学的应用[J]. 办公自动化, 2021(26): 57-59.
- [11] KONG X, WU D. The path of the ideological and political construction of the mixed teaching reform service curriculum [J]. China University Teaching, 2021(1): 59-62.
孔翔, 吴德. 以混合式教学改革服务课程思政建设的路径初探[J]. 中国大学教学, 2021(1): 59-62.
- [12] WU W, CHENG B X, LI Z. Introduction to dialectics of nature [M]. Guangzhou: Sun Yat-sen University Press, 2019.
吴玮, 程本学, 李珍. 自然辩证法概论 [M]. 广州: 中山大学出版社, 2019.

(责任编辑: 黄 健)

13. 教改论文：工程认证背景下软件工程专业实训模式探索



计算机教育

2023年度协办单位

清华大学计算机科学与技术系	书记 汤晓鸥
启迪数字学院	执行院长 汤晓鸥
北京航空航天大学软件学院	院长 吕卫
北京航空航天大学计算机学院	副院长 姜晓
国防科技大学计算机学院	副院长 毛德
天津市大学软件学院	院长 姜晓
北京林业大学信息学院	院长 姜晓
哈尔滨工业大学计算学部	主任 姜晓
复旦大学软件学院	副院长 姜晓
北京信息科技大学信息管理学院	院长 姜晓
青海东软信息技术职业学院	院长 姜晓
西安邮电大学计算机学院	院长 姜晓
吉林大学计算机科学与技术学院	副院长 姜晓
吉林大学软件学院	院长 姜晓
北京邮电大学软件学院	执行院长 姜晓
北京交通大学软件学院	院长 姜晓
苏州大学计算机科学与技术学院	院长 姜晓
上海交通大学软件学院	常务副院长 姜晓
武汉科技大学计算机科学与技术学院	副院长 姜晓
哈尔滨理工大学计算机科学与技术学院	院长 姜晓
深圳大学计算机与软件学院	院长 姜晓
山东工商学院计算机科学与技术学院	副院长 姜晓
湖北工业大学计算机学院	院长 姜晓
温州商学院信息工程学院	院长 姜晓
嘉兴技师学院	副院长 姜晓

特别支持单位

清华大学出版社计算机与信息分社	社长 姜晓
百度在线网络技术(北京)有限公司	姜晓

- 140 基于多方位、递进式的数据库实验教学研究 姜 敏, 谢 佳, 汪花梅
- 146 新工科背景下的高校双创教育实践 王瑞峰, 吴祖峰, 李勇昊, 张菊洁, 周世杰
- 152 工程认证背景下软件工程专业实训模式探索 王金凤, 黄 琼, 林玉萍, 肖 磊, 陈晓伟, 邓维豪
- 158 基于知识图谱的综合实践教学设计 陈 飞, 张春阳, 于元隆

课程思政

- 163 通识教育课程中的思政教育探索与实践——以C语言程序设计课程为例 杨 鑫, 崔志波, 王宇新, 刘 旭, 魏小明
- 168 应用型大学软件工程专业课程思政教改探索 王海峰, 刘 旭, 张 明, 王九如, 何永庆, 赵德玉

核心课程建设

- 173 认知负荷视角下的数据结构实验题库设计 司国东, 彭利民, 张义青

教改纵横

- 177 以产教融合为导向的软件工程课实践教学改革 赵 伟, 竺 琳
- 181 基于翻转课堂的数据库系统课程教学改革 张夏雷, 庞桂芬
- 186 面向赋能教育的传感器原理及应用教学改革 范文婷, 王 晓

教育与教学研究

- 192 面向专业学位研究生教育的模式识别课程教学改革 王家宝, 张旭旭, 潘志松
- 198 基于成果导向的C++程序设计课程培养目标达成度评价方法探索 邢艳丽, 申 华, 高文娜
- 204 对分课堂联合TBL教学新模式设计与实践 郭建华, 刘以印
- 209 基于设计思维和项目驱动的STEM课程框架构建与应用 李玉敏
- 216 双维度分层实训方法在软件项目开发实训教学中的应用 丁 俊
- 220 “导师牵引+竞赛驱动”的人工智能专业人才培养 姜 琳, 黄 华, 刘会会, 杨俊平, 刘夏宝
- 225 新工科背景下“岗证思训”融合教学模式探索 徐安宁, 孙 彬, 周然群, 徐 琳, 占 艳, 袁慧灵



扫一扫
或搜索sjjyzz
微信也精彩

本刊为

中国知网数据库(CNKI)全文收录期刊
中国期刊全文数据库(CJED)全文收录期刊
中国学术期刊综合评价数据库(CAJCED)统计源期刊
中国重要会议论文全文数据库(CPCI)收录期刊
万方数据—数字化期刊群全文收录期刊

工程认证背景下软件工程专业实训模式探索

王金凤¹, 黄琼¹, 林丕源¹, 肖磊¹, 陈晓伟², 邓振豪³

¹ 华南农业大学数学与信息学院, 广东 广州 510642;

² 四川华迪信息技术有限公司, 四川 成都 610500;

³ 广州东软轨道交通教育技术有限公司, 广东 广州 510000

摘要: 分析双一流建设背景下软件工程专业存在的问题, 以工程认证理念为指引, 以实践能力产出为导向, 提出校企深度融合、沉浸式指导, 全方位构建实训实践课程体系结构, 优化软件工程专业人才培养模式, 结合华南农业大学软件工程的实训课程开展经验, 从建设思路、设计理念、实施过程、效果评价几个方面阐述该模式的建设过程。

关键词: 工程认证; 软件工程; 实训模式; 双一流建设

1 研究背景

工程教育是我国高等教育的重要组成部分, 而工程教育专业认证是国际通行的工程教育质量保障制度, 要求专业课程体系设置、师资队伍配备、办学条件配置等都围绕学生毕业能力达成这一核心任务展开, 并强调建立专业持续改进机制, 以保证专业教育质量和专业教育活力^[1]。未来工程人才的培养目标注重知识能力、能力和素质全方位发展^[2-3]。2017年教育部启动了“新工科研究与实践”项目, 强调培养的学生应具有工程实践能力强的特征^[4]。

软件工程是普通高等学校本科专业, 该专业目标是培养学生适应软件产业的发展, 使其具备计算机软件的基础理论、基本知识和基本技能, 具有用软件工程的思想、方法和技术来分析、设计和实现计算机软件系统的能力^[5-6]。软件工程专业课程体系包括通识类知识、学科基础知识、专业知识和实践性教学等, 必须支持各项毕业要求的有效达成, 进而保证专业培养目标的有效实现。

2 专业实践现存问题

传统的专业教学体系中, 多是以校内教师授课为主, 通过课堂教学、作业训练、期末测验等固定环节, 完成学生对基础理论知识的掌握, 其结果是理论与实践脱节。学校在专业建设过程中逐渐引入行业内企业, 让学生在真实的实验场景内得到锻炼, 但仍存在一些普遍问题。

(1) 广度不足, 实训课程体系不完善。

(2) 深度不够, 传统的实训题目脱离实际, 浮于表面。

(3) 高度不达, 实训企业基础薄弱, 无法支撑工程教育的毕业要求。

3 实训建设思路

鉴于实际问题, 为了满足国家工程人才培养需求, 以国家级一流本科专业软件工程专业为例, 我们开展了基于成果导向教育(OBE)理念的实习实训环节改革和创新, 提出了基于CDIO教育模式的工程认证能力培养模式^[7]。首先, 拓宽课程体系广度, 将多门工程类课程打造成为

基金项目: 广东省一流本科课程建设项目“操作系线上线下混合式一流课程”(202012039); 国家级一流本科专业建设项目“软件工程”(教高厅函[2022]14号); 教育部产学研协同育人项目“基于鲲鹏的操作系统课程实践教学建设”(201901134024)。

作者简介: 王金凤, 女, 副教授, 研究方向为机器学习、生物信息学、教育技术学, wangjinfeng@scau.edu.cn; 黄琼(通信作者), 男, 教授, 研究方向为信息安全, qhuang@scau.edu.cn。

...课程基础上,构成实训课程体系,以便于学生能...理论课程和实践课程的平滑过渡和完美结...课程,加大实训课程深度,与实训企业一起...课程教学大纲和人才培养方案,根据学校定...学院发展方向,重新制订实训项目。根据...实际项目,经过脱敏处理,进行模块化编...学生小组分工完成子模块,更加深入地...和学习实训题目,同时培养学生的团队协...力。最后,加大实训企业的合作范围和力...实训基地的专业素质和水平,淘汰一批

小型,无创新理念的合作单位,增加与华为、百度等五百强企业之间的合作机会,使得学生在校期间就能够有机会拓宽眼界,树立软件兴国的信念。

实训课程体系的整体设计采取反向设计法,以毕业要求为引导方向,设计制订课程群的教学大纲和实施方案,在确定了课程对毕业要求能力预期达成度的基础上,建立反向映射关系,以构成整个课程体系的良性循环。具体课程体系架构如图1所示。

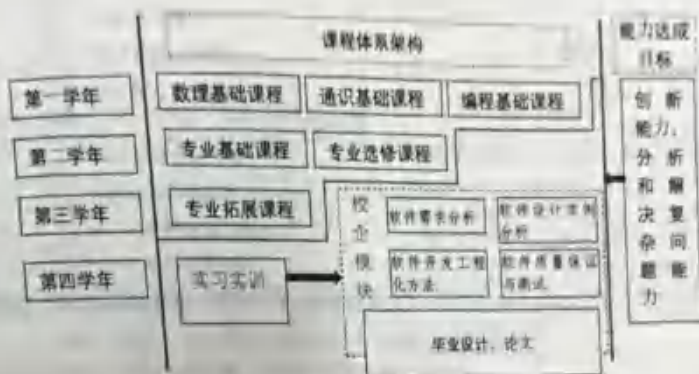


图1 课程体系架构

校企培养模块跨越贯穿第三和第四学年,软件工程相关技术课程融入整体实训过程,以实体项目为牵引,将各科课程分解渗透进入项目子模块,让学生在项目完成的过程中同步完成各项能力的训练与加强,而替代了拘泥于课堂讲授的形式,大大提升了学习兴趣和积极性。最后,依托实训过程积累的经验和打下的技术基础,学生在企业导师指导下确定毕业设计题目,完成开题和系统的初步设计和搭建,之后在校内导师指导下完成毕业论文的撰写和毕业答辩环节。

软件工程专业的基础课程学习和专业技术课程在第1—6学期已经基本完成,为了进一步深化技术理解,提升解决复杂工程问题的能力,将实训实践课程安排在第三学年和第四学年的跨越阶段,共计13周。其先后分为4个阶段:包括专业技术补强阶段、案例剖析、全真项目实战阶段、企业岗位实习。校企联合培养模块的5门课程分别为软件需求分析、软件开发工程化方法、软件设计实例分析、软件质量保证与测试、团队

激励与沟通,其中团队激励与沟通课程教学贯穿于项目实训整个过程,主要集中于第3阶段:团队协同职业岗位综合应用,理论授课结合实际操作。考评过程结合项目的实训,遵从工程教育专业认证的评价过程进行,最后考评时再独立给成绩。整个项目推进过程见表1。

4 具体实施开展

在学院的实际教学过程中,实践基地随时掌握了解各在线实训项目的执行情况和执行效果。按照教学实施方案,项目管理与质量监控部学生管理专员每周末了解本周学生对所学内容的掌握情况,形成项目例会纪要进行记录和存档。

1) 代码管控。

授课工程师每天收集学生完成的日常练习代码,并对学生提交的成果进行点评。通过对学生成果提交管控,督促学生每天保质保量完成规定的学习任务和实操训练。学生每天的练

表 1 实训小组项目推进过程设计表

项目阶段	分类	交付工件
计划	管理文档	项目开发计划
	项目管理	各个计划的同行评审记录、同行评审检查单
	配置管理	配置管理计划
先启阶段	技术文档	客户需求原始记录（由指导讲师提供）
		涉众请求
	需求开发	软件需求规约（含用例模型）
		词汇表
需求开发	项目管理	软件需求规约同行评审记录、同行评审检查单 项目组成员工作日志、项目周报 里程碑会议纪要、里程碑报告（需求）
	配置管理	配置状态报告 需求基线发布报告
	PPQA	由 PPQA 工程师提供 需求过程评审报告、需求产品审计报告、PPQA 周报
精化阶段	技术文档	系统架构设计说明书
		数据库设计说明书
		概念数据模型（CDM） 物理数据模型（PDM）
	分析设计	分析模型（Rose 模型：用例实现、时序图、状态图） 设计模型（Rose 模型：类图）
		各个技术文档的同行评审记录、集成测试用例与系统测试用例的同行评审记录、同行评审检查单
	项目管理	项目组成员工作日志、项目周报 里程碑会议纪要、里程碑报告（分析设计）
	配置管理	配置状态报告 设计基线发布报告
测试	集成测试计划 集成测试用例、系统测试用例	
软件实施	技术文档	源代码和可执行代码
	项目管理	源代码的同行评审记录、同行评审检查单 项目组成员工作日志、项目周报
	配置管理	配置状态报告
	测试	单元测试用例、集成测试用例、系统测试用例 单元测试日志、集成测试日志、系统测试日志
	构建阶段	技术文档
项目组成员工作日志、项目周报 里程碑会议纪要、里程碑报告（测试）		
配置管理		配置状态报告 测试基线发布报告
项目管理		项目问题跟踪表 项目工作量统计及偏差记录
整个过程	配置管理	基于 CMM13 的配置库 变更申请单

对成果提交情况作为学生成绩,构成职业素养成绩的重要组成部分。企业导师收集学生完成的每日进度报告文档,并对学生提交的进度报告进行

点评。学生每天的进度报告提交情况作为学生成绩构成职业素养成绩的重要组成部分,如图2所示。

周次	姓名	班级	项目名称	项目进展
1	王鑫	计算机科学与技术	基于Web的电商系统配置管理环境搭建	需求分析、需求规格、业务需求分析、建立系统需求、系统需求分析
2	曹亦凡	计算机科学与技术	基于Web的电商系统配置管理环境搭建	完成需求分析以及项目初步需求,进行需求分析,业务需求转化,完成系统需求
3	李兆典	计算机科学与技术	基于Web的电商系统配置管理环境搭建	需求分析、需求规格、业务需求分析、系统需求分析、需求规格、建立系统需求
4	王洪伟	计算机科学与技术	基于Web的电商系统配置管理环境搭建	讨论项目分工,进行需求分析,讨论确定项目的业务需求,建立系统需求,需求规格
5	刘震宇	计算机科学与技术	基于Web的电商系统配置管理环境搭建	讨论项目分工,进行需求分析,讨论确定项目的业务需求,建立系统需求,需求规格
6	张国强	计算机科学与技术	基于Web的电商系统配置管理环境搭建	讨论项目分工,建立业务需求,建立系统需求,需求规格,需求规格及需求

图2 学生每日项目进度报告样例截图

2) 课程安排。

实训具体实施安排见表2,期间穿插企业文化、职业讲座、求职技巧等系列讲座,提升学生的中华文化素养,贯穿爱国主人教育,力求业务和思政的全方位发展。

3) 实施管控。

授课工程师指导学生完成配置管理环境搭建、完成实训项目分组及实训成员角色分工。学生在项目开发过程中,授课工程师每节课随机抽查学生完成的实践成果,根据学生实践情况以及

表2 实训课程具体安排

周次	课程内容
1	开营仪式; 互联网 Web 前端开发技术; JavaScript 核心、ES6 高级编程技术
2	智能终端应用开发(应用端); jQuery 前端框架技术、Node.js 平台应用技术
3	智能终端应用开发(应用端); MongoDB 分布式数据库应用、ExpressWeb 框架技术
4	智能终端应用开发(服务器); ElementUI 框架技术、Vue.js/React 框架技术、微信小程序开发
5	互联网服务器端开发技术; WebPack 平台工具应用
6	软件设计案例分析: 跨境电商系统分析与设计实践; 规范/配置管理
7	软件需求分析: 跨境电商需求分析; 跨境电商交易平台研发实战; 专题/职业讲座
8	软件开发工程化方法; 农副产品交易平台研发实战; 功能评审考核
9	农副产品交易平台研发实战; 软件质量保证与测试; 系统评审验收
10	企业岗位实习; 实践中提升需求分析、软件设计、软件开发、软件质量等; 阶段性总结与交流

抽查过程中发现的问题,及时线上单独沟通,组织召开小组会议沟通。

授课工程师采用1+3+6教学模式,即在整个教学过程中,核心关键知识讲解占比10%,项目案例应用分析讲解占比30%,学生项目开发实践

操作训练占比60%。在实训过程中,授课工程师在核心关键知识点讲解和项目案例应用分析环节采用在线直播讲解方式实施。学生项目开发实践操作训练过程中的过程管理尤为关键,重点在以下过程做好实施管控,见表3。

表3 进度管控节点

工作项	主体	时间点	内容
进度抽查	授课工程师	每课	随机抽查学生实践成果
线上指导	授课工程师	每课	根据学生实践情况以及抽查过程中发现的问题,线上组织召开小组会议,如果是普遍性问题则召开班级会议
进度报告	学生	每日	提交项目实践进度报告,授课工程师查阅,存储留档
项目例会	授课工程师 学生管理专员	每周	组织实训小组组长带领组员线上会议报告项目进度和实践情况
阶段评测	授课工程师	每阶段	各阶段结束时,组织学生线上测试方式完成阶段评测
总结评审	授课工程师	总结阶段	项目结束时,采用线上班级会议方式由各实训小组组长带领组员分别展示各自所完成的实训作品并详细讲解,进行总结评审

4) 成果交付管控。

按照实训进度要求,各项目开发团队根据进度安排分别提交阶段性成果工件,由企业导师审核评定。各个项目小组的阶段性成果都是按照

企业实际场景的商业项目验收评审进行,多个小组之间形成交叉循环模式,企业完成评定,指引和给出改进意见,最终执行到完整的系统交付流程,直至实训过程终结,如图3所示。

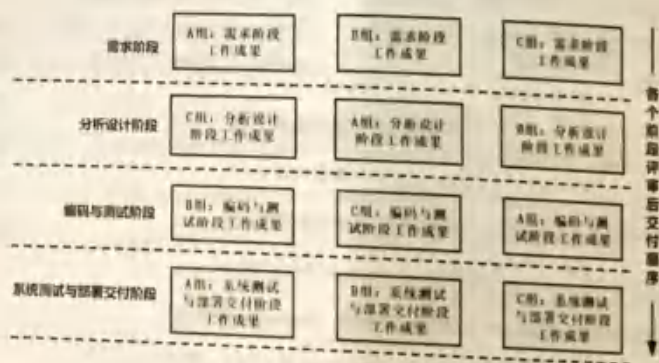


图3 项目小组交叉循环交付流程

5 特色分析和效果评价

1) 特色分析。

支撑工程教育专业认证的实训目标,能够满足10条专业认证毕业要求的12条子要求。针对不同能力的团队小组同学分类培养,规范与标准

一致,但是考核难度要求不同。大量采用最新真实商业案例贯穿始终,使学生尽快熟悉。覆盖大型和外资企业对IT工程师需要的所有岗位技能要求。工程师全过程跟踪记录管理,记录并掌握每一个学生的详细学习情况,确保实训质量。全方位的团队训练活动,包括日常例会、周例会、

同行评审、阶段评审、答辩演讲等。项目制牵引结合公司化运作的实训实践特色主要体现在：实训实践计划中把企业文化、新技术讲座、职业规划、面试技巧、简历制作、模拟面试等职业素养和双创课程贯穿在整个过程中。

2) 机制评价。

在机制改革基础之上，建立国家级“华南农业大学—四川华迪信息技术有限公司工程实践教学中心”，由四川华迪代表双方申报的“‘本科教学工程’大学生校外实践教育基地建设项目”，获中国高等教育学会认定为“校企合作双百计划”典型案例。根据麦可思公司的跟踪评价，本专业就业质量高。2011—2019届毕业生中有7届学生月收入位列我校第一。根据2017、2018年每年发布的《华南农业大学社会需求与培养质量跟踪评价报告》，我校2016届和2017届各专业毕业后半年月收入，软件工程专业位列第三和第二；根据《华南农业大学2019届毕业生培养质

量评价报告》，2018届和2019届毕业生月收入，软件工程专业两届都位列第一。此外，近两年对14家用人单位回访，他们评价本专业毕业生“综合素质好、专业水平高”“软件工程能力强”“工作积极主动”。

6 结语

在工程认证背景下，面向双一流建设需求，对国家级一流专业——软件工程的实训环节实施改革和创新举措，旨在培养能够分析和解决复杂工程问题的创新型、复合型人才，同时具有爱国主义情怀和团队协作精神，为软件行业输送高质量人才储备。本模式主要创新点是跨学年分配实训课程、双师制指导学生实践、交叉循环完成阶段考核，使得实训实践环节完成广度宽、深度足、高度达的实践课程体系水平，为同类工程专业提供模式借鉴。

参考文献：

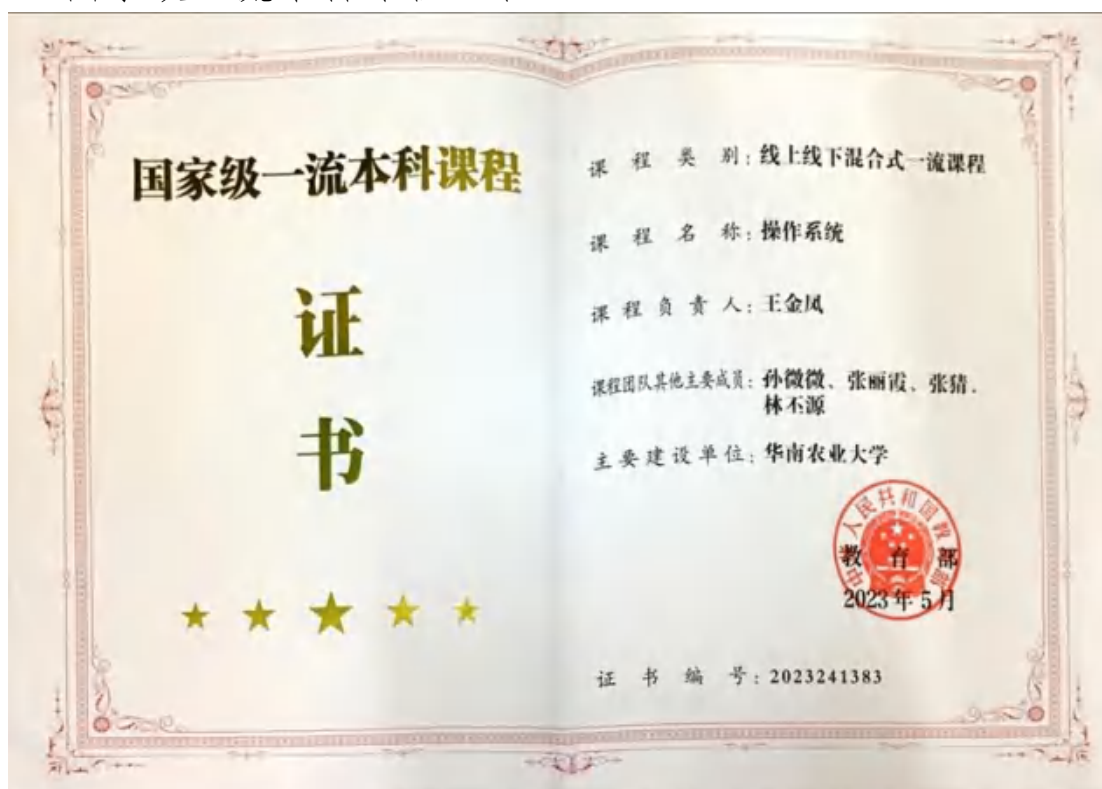
- [1] 高海燕. 高等工程教育人才培养模式改革若干问题探析[D]. 西安: 西安电子科技大学, 2020.
- [2] 杨永斌, 李佩琳, 刘曼君. 世界工程教育认证的发展趋势[J]. 高等工程教育研究, 2019(5): 5-10.
- [3] 葛耀群. 工程教育的未来已来[J]. 桥梁, 2021(6): 104.
- [4] 吴建华, 侯永峰, 杨秋波, 等. 加快发展和建设新工科 主动适应和引领新经济[J]. 高等工程教育研究, 2017(1): 1-9.
- [5] 教育部高等学校教学指导委员会. 普通高等学校本科专业类教学质量国家标准(上)[M]. 北京: 高等教育出版社, 2018(3): 321-329.
- [6] 罗兴宇, 刘玲, 闫会峰. 计算机科学导论[M]. 上海: 上海交通大学出版社, 2017: 109-114.
- [7] 李攀, 崔家瑞, 阎群, 等. 工程教育认证下自动化专业实践类课程改革[J]. 实验技术与管理, 2016, 33(12): 4.

(编辑: 孙怡铭)

14. 教学成果奖证书



15. 国家级一流本科课程证书



16.广东省一流本科课程证书



17.教学比赛证书



18.主编教材：《操作系统原理-以 OpenEuler 为例》



内容简介

本书为示范性软件学院联盟建设的首批软件工程系列教材之一。本书以操作系统工作原理为主线，深入讲解操作系统如何实现系统资源的调用、管理，以及如何协助用户程序的调度与执行。

本书首先从计算机系统结构入手，由支撑操作系统运行的硬件层面，包含操作系统在内的软件层面，以及保障整个系统正常运转的软硬协调机制三方面展开介绍；并对当前国产操作系统的发展现状，尤其是 openEuler 系统的概况进行了介绍。而后按照操作系统功能模块划分，分别通过进程管理、内存管理、处理器调度、设备管理和磁盘调度、文件管理五方面，深入剖析操作系统的管理机制和工作原理；同时，结合 openEuler 系统实例，展示操作系统原理在现代操作系统中的应用实例。每章后面会根据内容的重要程度编配相应的习题和计算题，部分题目选自历年全国研究生入学考试试题。本书为新型教材，结合国家统一本科课程慕课视频，将重要知识点的视频二维码穿插于教材中的相关文字部分，读者可以根据兴趣和需要扫码完成线上学习。

本书主要面向高等院校计算机类专业本科生，既包含了本课程专业知识的内容，又可以满足具有考研意向学生的需求，同时还拓展了读者对 openEuler 系统的了解。

图书在版编目(CIP)数据

操作系统原理：以 openEuler 为例 / 王金凤主编；
孙微微，张丽霞，张倩副主编. —北京：高等教育出版社，2024.7

ISBN 978-7-04-062179-2

I. ①操… II. ①王… ②孙… ③张… ④张… III.
①操作系统-高等学校-教材 IV. ①TP316

中国国家版本馆 CIP 数据核字 (2024) 第 095733 号

Osasuo Xitong Yuanli, yi openEuler Wei

策划编辑 张 曦 责任编辑 张 曦 封面设计 李小微 版式设计 杨 树
责任绘图 马天迪 责任校对 高 歌 责任印制 赵义民

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100120
印 刷 北京中科印刷有限公司
开 本 787mm × 1092mm 1/16
印 张 16.5
字 数 340 千字
购书热线 010-58581118
咨询电话 400-810-0598

网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.hepmall.com.cn>
<http://www.hepmall.com>
<http://www.hepmall.cn>

版 次 2024 年 7 月第 1 版
印 次 2024 年 7 月第 1 次印刷
定 价 37.00 元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。
版权所有 侵权必究
物料号 62179-00



第1章 计算机系统概述1	1.7.2 创新性..... 40
1.1 计算机系统..... 2	1.7.3 openEuler 小结..... 41
1.1.1 硬件基本构成.....3	本章小结..... 41
1.1.2 现代微处理器.....5	习题..... 42
1.1.3 指令系统与指令执行.....6	第2章 进程管理 43
1.2 中断技术..... 8	2.1 进程..... 44
1.2.1 中断的概念.....8	2.1.1 进程概念..... 44
1.2.2 引入中断后的指令周期..... 10	2.1.2 进程状态..... 45
1.2.3 中断系统的软硬协同实现..... 11	2.1.3 进程描述..... 48
1.2.4 中断处理过程..... 13	2.1.4 进程切换与模式切换..... 52
1.2.5 多个中断的处理..... 15	2.1.5 操作系统的执行方式..... 54
1.3 存储器层次结构..... 17	2.1.6 openEuler 的进程..... 55
1.3.1 高速缓存..... 18	2.2 线程..... 60
1.3.2 内存缓存..... 20	2.2.1 多线程概述..... 60
1.3.3 外存储器..... 22	2.2.2 线程分类..... 62
1.4 操作系统目标..... 22	2.2.3 openEuler 的线程..... 64
1.4.1 交互方便性..... 23	2.3 并发原理..... 64
1.4.2 资源管理有效性..... 24	2.3.1 开发的基本概念..... 65
1.4.3 系统可扩展性..... 25	2.3.2 并发进程的制约关系..... 65
1.5 操作系统的发展与类别..... 26	2.3.3 操作系统设计问题..... 67
1.5.1 串行处理时代..... 26	2.4 同步与互斥..... 68
1.5.2 简单批处理系统..... 27	2.4.1 互斥..... 68
1.5.3 多道批处理系统..... 29	2.4.2 信号量..... 71
1.5.4 分时操作系统..... 30	2.4.3 管程..... 79
1.5.5 实时操作系统..... 31	2.4.4 消息传递..... 82
1.5.6 嵌入式操作系统..... 32	2.4.5 openEuler 实现案例..... 85
1.6 国产操作系统概述..... 34	2.5 死锁..... 90
1.7 openEuler 系统简介..... 37	2.5.1 死锁原理..... 90
1.7.1 平台优势..... 39	2.5.2 死锁预防..... 92

2.5.3 死锁避免	95	4.1.3 进程调度策略	151
2.5.4 死锁检测与恢复	99	4.2 单处理器调度算法	152
2.5.5 一种综合的死锁策略	101	4.2.1 先来先服务算法	152
2.5.6 哲学家就餐问题	101	4.2.2 最短进程优先算法	153
本章小结	104	4.2.3 最短剩余时间算法	155
习题	105	4.2.4 轮转调度算法	155
3.1 多级反馈队列算法	157	4.2.5 多级反馈队列算法	157
4.2.6 最高响应比优先算法	159	4.2.7 优先级调度算法	159
4.3 多处理器与多核调度	161	4.3 多处理器与多核调度	161
4.3.1 多处理器调度的设计	161	4.3.1 多处理器调度的设计	161
4.3.2 多处理器进程调度	162	4.3.2 多处理器进程调度	162
4.3.3 多处理器线程调度	163	4.3.3 多处理器线程调度	163
4.4 实时调度算法	165	4.4 实时调度算法	165
4.4.1 实时系统和实时任务	165	4.4.1 实时系统和实时任务	165
4.4.2 实时调度的类型	166	4.4.2 实时调度的类型	166
4.4.3 最后期限调度	167	4.4.3 最后期限调度	167
4.4.4 速率单调调度	170	4.4.4 速率单调调度	170
4.4.5 优先级反转	170	4.4.5 优先级反转	170
4.5 openFiler 的调度	172	4.5 openFiler 的调度	172
4.5.1 进程类别与调度策略	173	4.5.1 进程类别与调度策略	173
4.5.2 调度类与调度队列	173	4.5.2 调度类与调度队列	173
4.5.3 CPS 调度算法	174	4.5.3 CPS 调度算法	174
本章小结	176	本章小结	176
习题	177	习题	177
第 3 章 内存管理	108	第 5 章 设备管理与磁盘调度	180
3.1 内存管理需求	109	5.1 I/O 系统	181
3.1.1 重定位	109	5.2 I/O 硬件原理	182
3.1.2 保护	110	5.2.1 设备、设备控制器与通道	182
3.1.3 共享	110	5.2.2 I/O 控制方式	184
3.1.4 逻辑组织	110	5.3 I/O 软件原理	189
3.1.5 物理组织	111	5.3.1 中断处理程序	190
3.2 内存管理技术	111	5.3.2 设备驱动程序	191
3.2.1 固定分区	112	5.3.3 与设备无关的 I/O 软件	191
3.2.2 动态分区	114	5.3.4 用户层 I/O 软件	192
3.2.3 伙伴系统	116	5.4 缓冲技术	195
3.2.4 简单分页	118	5.4.1 缓冲的作用	195
3.2.5 简单分段	121		
3.3 虚拟内存	123		
3.3.1 虚存机制的内存分配	123		
3.3.2 虚存机制的可行性	124		
3.3.3 虚拟内存管理技术	124		
3.3.4 虚拟分页的管理策略	132		
3.4 openFiler 的虚拟内存管理	138		
3.4.1 openFiler 的虚拟内存地址空间	138		
3.4.2 openFiler 中的多级页表	139		
3.4.3 openFiler 中页交换的实现	140		
本章小结	142		
习题	142		
第 4 章 处理器调度管理	147		
4.1 处理器调度	148		
4.1.1 调度类型	148		
4.1.2 调度算法的性能指标	150		

5.4.2 缓冲池	197	6.1.1 文件的相关概念	220
5.4.3 UNIX SVR4的缓冲区 高速缓存	198	6.1.2 文件管理系统	221
5.4.4 减少磁盘I/O开销的措施	199	6.1.3 文件组织	223
5.5 磁盘调度	200	6.1.4 文件目录	227
5.5.1 磁盘结构与性能参数	200	6.2 文件物理存储	230
5.5.2 磁盘调度算法	202	6.2.1 文件物理组织	230
5.5.3 Linux的磁盘调度	205	6.2.2 磁盘管理	231
5.6 独立磁盘冗余阵列RAID	208	6.3 文件系统安全	238
5.6.1 RAID的技术基础	208	6.3.1 崩溃一致性	239
5.6.2 RAID的分级	209	6.3.2 文件保护	240
5.6.3 用软件实现RAID	213	6.4 文件系统实例	241
本章小结	214	6.4.1 MS-DOS文件系统	241
习题	215	6.4.2 UNIX文件系统	243
第6章 文件管理系统	219	6.4.3 Linux文件系统	244
6.1 概论	220	6.4.4 openSolaris文件系统	245
		本章小结	249
		习题	249



二、科研项目

1.主持：关于风湿云平台（FSC）关键技术开发

科目收文明细账								
部门：(0600)数学与信息学院、软件学院 项目：(0240675)风湿云平台（FSC）关键技术开发与示范								
科目信息：(610102010302-)自然科学类- 起始年月：2024年01月								
截止年月：2025年07月								
单位：元								
凭证日期	凭证编号	科目编号	科目名称	摘要	项目收入	项目支出	项目借款	项目还款
2024-09-03	(Z)01051	610102010302	自然科学类	阳江市民泰养生保健有限公司未款入账H240675	1,500,000.00	0.00	0.00	0.00
负责人:王金凤				H240675/风湿云平台（FSC）关键技术开发与示范合计	1,500,000.00	0.00	0.00	0.00
总计				总计	1,500,000.00	0.00	0.00	0.00

打印单位：华南农业大学
打印时间：2025-07-24 15:48:21

11x67m 720241749

合同编号:

技术开发合同

项目名称: 风湿云平台(FSC)关键技术开发与示范

委托方(甲方): 阳江市民泰养生保健有限公司

受托方(乙方): 华南农业大学

签订时间: 2024.06.12

签订地点: 华南农业大学

有效期限: 2024.6.30-2025.12.31

中华人民共和国科学技术部印制



填写说明

一、本合同为中华人民共和国科学技术部印制的技术开发合同示范文本，各技术合同认定登记机构可推介技术合同当事人参照使用。

二、本合同书适用于一方当事人委托另一方当事人进行新技术、新产品、新工艺或者新材料及其系统的研究开发所订立的技术开发合同。

三、签约一方为多个当事人的，可按各自在合同关系中的作用等，在“委托方”、“受托方”项下（增页）分别排列为共同委托人或共同受托人。

四、本合同书未尽事项，可由当事人附页另行约定，并可作为本合同的组成部分。

五、当事人使用本合同书时约定无需填写的条款，应在该条款处注明“无”等字样。

技术开发（委托）合同

委托方（甲方）：阳江市民泰养生保健有限公司

住 所 地：广东省阳江市阳西县沙扒镇渔港新村 D 座 63 号

法定代表人：黄拯民

项目联系人：黄拯民

联系方式：13824354313

通讯地址：广东省阳江市阳西县沙扒镇渔港新村 D 座 63 号

电话： 传真：

电子信箱：

受托方（乙方）：华南农业大学

住 所 地：广东省广州市天河区五山路 483 号

法定代表人：薛红卫

项目联系人：王金凤

联系方式：15815860686

通讯地址：广东省广州市天河区五山路 483 号华南农业大学

数学与信息学院

电话：15815860686 传真：020-85285893

电子信箱：wangjinfeng@scau.edu.cn

本合同甲方委托乙方研究开发“风湿云平台（FSC）关键技术开发与示范”项目，并支付研究开发经费和报酬，乙方接受委托并进行此项研究开发工作。双方经过平等协商，在真实、充分地表达各自意愿的基础上，根据《中华人民共和国合同法》的规定，达成如下协议，并由双方共同恪守。

第一条 本合同研究开发项目的要求如下：

1. 技术目标：打造国内首个风湿病垂直领域的云服务平台。实现多元、异构、多模态、流式风湿病数据的采集、存储、调用与可视化；实现医患、中医、中药用户的在线互动交流；实现药剂、临床数据的存储。

2. 技术内容：(1) 开发一个能够支持风湿病患者、方剂、药剂以及临床数据的数据中台；

(2) 开发一个风湿病患者愈疗管理系统；

(3) 开发一个风湿病医生管理系统

(4) 开发一个风湿病药剂管理系统

(5) 开发一个风湿病相关活动主体交易系统

3. 技术方法和路线：项目将打造一个支持风湿病愈疗的智能化交互平台。一是融合数据库与人工智能技术，支持风湿病多模态、流式、异构数据的采集、汇集、存储、融合、治理及创新应用。二是聚焦风湿病中药/复方/活性成分及功能，收集风湿病古籍、现代文献、英文等多语种风湿病药/复方/活性成分及功能信息，实现风湿病大数据统一入库、统一建模，形成数据资产。三是打通风湿病愈疗病患、医生、原材料、方剂、临床等主体的数据断点，实现不同场景数据交换、安全迁移、访问共享和业务协同。让风湿病相关病患、原材料供应商、医生的管理和业务、服务、应用等人员快速获取数据服务内容。

第二条 乙方应在本合同生效后10日内向甲方提交研究开发计划。研究开发计划应包括以下主要内容：

1. 项目的研究内容、研究目标，以及拟解决的关键科学问题；

2. 拟采取的研究方案；

3. 研究计划。

第三条 乙方应按下列进度完成研究开发工作：

1. 2024.06.20-2024.06.30: 调研、资料搜集、整理, 细分需求;
2. 2024.07.01-2024.08.31: 完成各系统开发;
3. 2024.09.01-2024. 9.30: 沟通完善产品_____;
4. 2024.10.1-2025.12.31: 协助甲方推向市场、回答甲方技术问

询。

第四条 甲方应向乙方提供的技术资料及协作事项如下:

1. 技术资料清单: 风湿云平台(FSC)应用方向; 具体需求; 推广预期、预算、预期进度等。

2. 提供时间和方式:

(1) 提供项目调研必要的支持_____;

(2) 提供线下示范必要的场所_____;

3. 其他协作事项: 友好协商确定_____。

本合同履行完毕后, 上述技术资料按以下方式处理: 乙方自行销毁_____。

第五条 甲方应按以下方式支付研究开发经费和报酬:

1. 研究开发经费和报酬总额为 150 万人民币_____。

其中: (1) 设备费 10 万人民币_____;

(2) 研究开发费 140 万人民币_____。

2. 研究开发经费由甲方_____一次_____ (一次、分期或提成) 支付乙方。具体支付方式和时间如下:

(1) 支付方式: 电汇(银行汇款)_____

(2) 支付时间: 合同生效 10 日内_____

乙方开户银行名称、地址和帐号为:

开户银行: 广州工行五山支行_____

地址: 广州市天河区五山路_____

帐号：3602002609000310520

第六条 本合同的研究开发经费由乙方以华南农业大学横向课题经费管理办法的方式使用。甲方有权以委托第三方审计的方式检查乙方进行研究开发工作和使用研究开发经费的情况，但不得妨碍乙方的正常工作。

第七条 本合同的变更必须由双方协商一致，并以书面形式确定。但有下列情形之一的，一方可以向另一方提出变更合同权利与义务的请求，另一方应当在30日内予以答复；逾期未予答复的，视为同意：

1. 甲方合同执行环境发生重大变更；
2. 乙方合同执行环境发生重大变更。

第八条 未经甲方同意，乙方不得将本合同项目部分或全部研究工作转让第三人承担。但有下列情形之一的，乙方可以不经甲方同意，将本合同项目部分或全部研究工作转让第三人承担：

1. 受空间限制，节约成本，保障质量及后期维护；
2. 受乙方单位软硬件条件限制，无法完成的。

乙方可以转让研究开发工作的具体内容包括：

1. 线下疗愈中心示范应用必须的设计、硬件环境改造
2. 风湿病相关信息采集、录入及加工
3. 系统检测
4. 其他经双方友好协商确定的

第九条 在本合同履行中，因出现在现有技术水平和条件下难以克服的技术困难，导致研究开发失败或部分失败，并造成一方或双方损失的，双方按如下约定承担风险损失：

1. 参照项目方案中的项目进度，按比例退还甲方研发经费；
2. 乙方研发后续投入自己负责。

双方确定，本合同项目的技术风险按第三方评估的方式认

定。认定技术风险的基本内容应当包括技术风险的存在、范围、程度及损失大小等。认定技术风险的基本条件是：

1. 本合同项目在现有技术水平条件下具有足够的难度；
2. 乙方在主观上无过错且经认定研究开发失败为合理的失败。

一方发现技术风险存在并有可能致使研究开发失败或部分失败的情形时，应当在30日内通知另一方并采取适当措施减少损失。逾期未通知并未采取适当措施而致使损失扩大的，应当就扩大的损失承担赔偿责任。

第十条 在本合同履行中，因作为研究开发标的的技术已经由他人公开（包括以专利权方式公开），一方应在10日内通知另一方解除合同。逾期未通知并致使另一方产生损失的，另一方有权要求予以赔偿。

第十一条 双方确定因履行本合同应遵守的保密义务如下：

甲方：

1. 保密内容（包括技术信息和经营信息）：（1）乙方提交的技术方案；乙方提供的原代码。
2. 涉密人员范围：甲方项目组参与成员。
3. 保密期限：3年。
4. 泄密责任：支付乙方合同金额5%赔偿金。

乙方：

1. 保密内容（包括技术信息和经营信息）：甲方提供的风湿云平台（FSC）应用方向；具体需求；推广预期、预算、预期进度等。
2. 涉密人员范围：乙方项目组参与成员。
3. 保密期限：3年。
4. 泄密责任：支付甲方合同金额5%赔偿金。

第十二条 乙方应当按以下方式向甲方交付研究开发成果：

1. 研究开发成果交付的形式及数量：1. 软件著作权申请回执；
2. 风湿云平台（FSC）系统软件及原代码一套；3. 相关使用说明书一套。
2. 研究开发成果交付的时间及地点：线上传输或线下拷贝。

第十三条 双方确定，按以下标准及方法对乙方完成的研究开发成果进行验收：专家评审通过或甲方认可。

第十四条 乙方应当保证其交付给甲方的研究开发成果不侵犯任何第三人的合法权益。如发生第三人指控甲方实施的技术侵权，乙方应当针对侵权部分进行修改。

第十五条 双方确定，因履行本合同所产生的研究开发成果及其相关知识产权权利归属，按下列第1种方式处理：

1. 甲（甲、乙、双）方享有申请专利的权利。

专利权取得后的使用和有关利益分配方式如下：甲方获得100%收益权。

2. 按技术秘密方式处理。有关使用和转让的权利归属及由此产生的利益按以下约定处理：

(1) 技术秘密的使用权：全部归甲方

(2) 技术秘密的转让权：全部归甲方

(3) 相关利益的分配办法：全部归甲方

双方对本合同有关的知识产权权利归属特别约定如下：全部归甲方。

第十六条 乙方不得在向甲方交付研究开发成果之前，自行将研究开发成果转让给第三人。

第十七条 乙方完成本合同项目的研究开发人员享有在有关技术成果文件上写明技术成果完成者的权利和取得有关荣誉证书、奖励的权利。

第十八条 乙方利用研究开发经费所购置与研究开发工作有关的设备、器材、资料等财产，归乙（甲、乙、双）方所有。

第十九条 双方确定，乙方应在向甲方交付研究开发成果后，根据甲方的请求，为甲方指定的人员提供技术指导和培训，或提供与使用该研究开发成果相关的技术服务。

1. 技术服务和指导内容：（1）提供使用说明书；（2）现场演示指导

2. 地点和方式：双方协商确定。

3. 费用及支付方式：另行协商签署补充合同确定。

第二十条 双方确定，甲方有权利用乙方按照本合同约定提供的研究开发成果，进行后续改进。由此产生的具有实质性或创造性技术进步特征的新的技术成果及其权属，由甲（甲、乙、双）方享有。具体相关利益的分配办法如下：甲方享有 100%收益和处分权力。

乙方有权在完成本合同约定的研究开发工作后，利用该项研究开发成果进行后续改进。由此产生的具有实质性或创造性技术进步特征的新的技术成果，归乙（甲、乙、双）方所有。具体相关利益的分配办法如下：乙方享有 100%收益和处分权力。

第二十一条 双方确定，在本合同有效期内，甲方指定黄拯民为甲方项目联系人，乙方指定王金凤为乙方项目联系人。项目联系人承担以下责任：

1. 黄拯民、王金凤负责项目整体沟通，协调推进
2. 王金凤负责项目研发方案的设计、团队组织、开发及财务审批、验收等
3. 项目实施进度的监督管理

一方变更项目联系人的，应当及时以书面形式通知另一方。未及时通知并影响本合同履行或造成损失的，应承担相应的责任。未经甲方同意，乙方不能更换联系人。

第二十二条 双方确定，出现下列情形，致使本合同的履行成为不必要或不可能的，一方可以通知另一方解除本合同：

1. 因发生不可抗力或技术风险；

第二十三条：双方因履行本合同而发生的争议，应协商、调解解

甲方： 阳江市民泰养生保健有限公司 (盖章)

法定代表人/委托代理人： 易松民 (签名)

年 月 日

乙方： 华南农业大学 (盖章)

法定代表人/委托代理人： 薛红妮 (签名)

年 月 日

印花税票粘贴处：

(以下由技术合同登记机构填写)

合同登记编号：

2.主持：关于基于时空大数据的活鸡智能配送与调度系统关键技术与示范应用项目的合同

合同编号：

技术开发（合作）合同

项目名称：基于时空大数据的活鸡智能配送与调度系统关键技术与示范应用

项目类型：华农温氏科创中心揭榜项目

甲方：温氏食品集团股份有限公司

乙方：华南农业大学

丙方：温氏食品集团股份有限公司研究院

签订时间：2024年5月20日

签订地点：广东省云浮市新兴县

有效期限：2024年1月1日至2026年12月31日

填写说明

- 一、本合同为技术开发（合作）合同示范文本。
- 二、本合同书适用于当事人各方就共同进行新技术、新产品、新工艺、新材料或者新品种及其系统的研究开发所订立的技术开发合同。
- 三、本合同书未尽事项，可由当事人附页另行约定，并可作为本合同的组成部分。
- 四、当事人使用本合同书时约定无需填写的条款，应在该条款处注明“无”等字样。

技术开发（合作）合同

甲方： 温氏食品集团股份有限公司

住 所 地： 广东省云浮市新兴县新城镇东堤北路9号

法定代表人： 温志芬

项目联系人： 王凯

联系方式： 18819492878

通讯地址： 广东省云浮市新兴县

电 话： 0766-2986008 传 真： 0766-2986008

电子信箱： 373490238@qq.com

乙方： 华南农业大学

住 所 地： 广东省广州市天河区五山路483号

法定代表人： 薛红卫

项目联系人： 王金凤

联系方式： 15815860686

通讯地址： 广东省广州市天河区五山路483号

电 话： 020-85280320 传 真： 020-85285393

电子信箱： wangjinfeng@scau.edu.cn

丙方： 温氏食品集团股份有限公司研究院

住 所 地： 广东省云浮市新兴县新城镇惠能北路温氏研究院

法定代表人： 陈丽

项目联系人： 黄明亮（IT）、曾伟谦（业务）

联系方式： 13580544143、15016599626

通讯地址： 广东省云浮市新兴县新城镇惠能北路温氏研究院

电 话： 0766-2986008 传 真： 0766-2986003

电子信箱： wskygl@163.com

本合同合作各方就共同参与研究开发基于时空大数据的活鸡智能配送与调度系统关键技术与示范应用项目事项，经过平等协商，在真实、充分地表达各自意愿的基础上，根据《中华人民共和国民法典》的规定，达成如下协议，并由合作各方共同恪守。

第一条 本合同合作研究开发项目的要求如下：

1. 技术目标：

(1) 预期研究成果及达到的水平

1) 集成开发“基于时空大数据的温氏智能协同配送、调度执行系统”一套；系统实现与集团苍穹平台接口对接，支撑数据实时交互。

2) 实现在广东养鸡公司 90%以上棚面应用。

3) 联合温氏集团申请发明专利 2 项以上，至少授权 1 项；

4) 发表高水平论文 3 篇以上。

(2) 技术考核指标

1) 系统运行效率指标：结果在 30 分钟以内

2) 智能车辆调度算法时间误差：配送车辆的行驶和等待时间用时与算法得出的预计用时合计偏差小于 30 分钟；

3) 智能配送算法匹配率：配送双方契合度，匹配度为 90%或以上；

4) 智能配送算法时间误差：时间匹配度，配送偏差小于 20 分钟。

(3) 经济考核指标

1) 实现在广东养鸡公司 90%以上棚面应用。

2) 节省不少于 10 个人力成本。

3) 调度环节工作效率提高 30%以上

4) 节约人员、车辆及开栏导致的耗损等各种成本不少于 3000 万。

(4) 在公司的应用价值

1) 节约人工成本-将节约不少于 10 个人工成本/天。

①调度环节：经前期调研了解到目前广东养鸡公司有服务站 10 个以

上，以 10 个计，每个服务站每天从下午结单开始，排班时长 3 个工时，调度员用语音通知客户，工作总时长 5 个工时，合计 8 个工时。本系统运行后实行自动化排班和通知，预计每个服务站节约 5 个工时，示范点共节约 50 个以上工时。

②分拣环节：广东养鸡公司日均出鸡量约 60 万只，磅位工作总时长 200 小时以上，因匹配度高、棚面压鸡减少等，整体分拣、结算效率提高，按分拣环节效率提高 5% 计算，按每个磅位均需 5 人计算，共节约工时 50 个人工时。

2) 提升经济效益-事业部推广应用后，节约成本合计不少于 3000 万
在温氏指定区域的生产端、销售端以及运营端等关键环节开展示范应用，节约人员、车辆、物料等各种，整体工作效率提高 5% 以上。

①节约调度、排班人员人力成本 100 万元；

②降低开栏数，降低开栏因鸡的应激反应而导致的各种损耗，预期降低开栏数 10% 左右，有效降低开栏销售天数，0.8 天，节约成本 2000 万；

③降低运输成本，通过整体规划，就近运输，降低车辆空置率，防止寻租等，节约运费 1000 万；

④降低分拣耗损导致的损失。降低供货在棚面等待时长，降低耗损 50 万。

3) 提高工作效率--不低于 5%

①本系统可降低客户的等待时长，减少跨服务站取货，降低客户等待总时长 10%；

②可降低养户在棚面等待时长、运输时长（就近运输）和抓鸡时长（集中抓鸡装车），降低总时长 10%

2. 技术内容：面向肉鸡配送场景和每天的订单大数据，开展一个肉鸡智能配送与调度模型研究与系统实现的研究。（1）分布式增量数据库系统研究。（2）开展多元全时空联动的智能配送与调度模型研究。（3）

智能配送与调度系统集成与应用。

3. 技术方法和路线：(1) 需求提取转换。深入基层调研，对业务内容进行梳理，重点针对用户排单、养户排单、棚面调度、养户抓鸡送货等业务进行深入调研挖掘，提取面临的卡脖子、突出问题。如图所示，经前期调研，目前可将问题提取为时间、空间、商品、工位匹配问题。(2) 问题因素分析及建模。系统梳理每个问题的影响因素，并将业务问题提取转换为数学科学问题，归类动态特性下为多目标优化问题。重点基于建立时空网络流模型，构建订单（客户）分配调度模型、养户活鸡配送调度模型、活鸡配送反应式调度模型、服务站及棚面调度模型。(3) 形成基本的调度逻辑。在时空网络模型分析基础上，结合业务实际提出客户优先分配规则、棚面订单分配优先规则、反应式调度规则、养户订单分配优先规则。(4) 算法研究及优化。针对模型和规则及业务特征，设计算法。在此基础上具体分析生产流程与车辆调度，并提出一个包括生产序列确定、车辆调度和棚面调度的八组优先组合规则的前摄启发式算法来解决该调度问题。为评价应用不同组合规则的优劣，项目以总费用、惩罚费用和时间等因素建立相关目标函数，并根据活鸡配送案例中养户离散程度和订单数量将案例分成四类。(5) 生成调度方案。经过比较优化，自动生成各业务环节调度方案。

第二条 本合同合作各方在研究开发项目中，分工承担如下工作：

甲方：

1. 研究开发内容：_____ 无 _____。
2. 工作进度：_____ 同乙方 _____。
3. 研究开发期限：_____ 同乙方 _____。
4. 研究开发地点：_____ 同乙方 _____。

乙方：

1. 研究开发内容：基于时空大数据的温氏智能协同配送调度执行系

统。

2. 工作进度：（第一阶段）2024年度末，开展技术布局，完成第一版算法，并于2025年第一季度在一个试点单位上线。目标完成技术路线、需求分析、数据收集整理、模型构建工作，并在一个棚面试点上线；

（第二阶段）2025年度末，完成第二版算法优化、完成与苍穹平台的数据对接。目标完成配送智能调度系统的开发与数据集成工作；

（第三阶段）2026年度末，达成项目技术目标，系统在广东养鸡应用。完成系统在广东养鸡公司90%以上棚面应用，验证效果。

3. 研究开发期限：2024年5月-2026年12月。

4. 研究开发地点：广州、新兴。

丙方：

1. 研究开发内容：提供经费支撑。

2. 工作进度：同乙方。

3. 研究开发期限：同乙方。

4. 研究开发地点：同乙方。

第三条 为确保本合同的全面履行，合作各方确定，采取以下方式对研究开发工作进行组织管理和协调：由乙方和丙方各派出一位负责人进行相关研究工作的协调和组织管理；甲方、丙方负责协调研发成果落地；乙方负责技术研究，形成可落地实施的研发成果。按季度召开项目进展会；按月会同各方负责人协同项目进度。

第四条 合作各方确定，各自为本合同项目的研究开发工作提供以下技术资料和条件：

甲方：负责协调研发成果落地。

乙方：负责调度系统研发与交付使用。

丙方：负责提供项目研发所需的实验调试数据和应用场景。

本合同履行完毕后，上述技术资料和条件按以下方式处理：交付甲

方和丙方使用。

第五条 合作各方确定，按如下方式提供或支付本合同项目的研究开发经费及其他投资：

甲方：

1. 提供或支付方式： 无。
2. 支付或折算为技术投资的金额： 无。
3. 使用方式： 无。

乙方：

1. 提供或支付方式： 无。
2. 支付或折算为技术投资的金额： 无。
3. 使用方式： 经费支付直接性投入金额 120 万元，其中包括材料

费用 15 万元，用于支付三年产生的易耗品材料费用；差旅费和会议费 35 万元，用于支付生产一线调研、项目推进研讨会产生的差旅和会议费用；出版/文献/信息传播/知识产权事务费 20 万元，用于项目相关的论文出版和知识产权申报所产生的费用；专家咨询费 15 万元，用于支付项目实施产生的专家费用；劳务费用 35 万元，支付项目团队成员的劳务费用。除了直接费用外，需支付项目管理费用 30 万元，用于平台项目管理费用。

丙方：

1. 提供或支付方式： 提供项目总经费 150 万元，项目经费分三次
拨付。第一笔经费为 30%，即 45 万元，在合同签订后拨付；第二笔经费
为 40%，即 60 万元，在工作进度的第二阶段，目标是完成 1 个配送智能
调度系统，并实现数据集成，完成与苍穹平台的数据对接，联合温氏集团
申请发明专利 2 项，经会议认定阶段性考核目标完成后，拨付第二笔经费；
第三笔经费为 30%，即 45 万元，在工作进度的第三阶段，目标是完成系
统在广东养鸡公司 90%以上棚面示范应用，具体见技术、经济考核指标，
经会议认定项目考核目标完成，拨付剩余经费。

2. 支付或折算为技术投资的金额：无。

3. 使用方式：无。

第六条 以提供技术为投资的合作方应保证其所提供技术不侵犯任何第三人的合法权益。如发生第三人指控合作一方或多方因实施该项技术而侵权的，提供技术方应当承担相应经济责任，但乙方承担的金额不超过其已收取的金额。

第七条 本合同的变更必须由合作各方协商一致，并以书面形式确定。但有下列情形之一的，合作一方或多方可以向其他合作方提出变更合同权利与义务的请求，其他合作方应当在7日内予以答复；逾期未予答复的，视为同意。

第八条 未经其他合作方同意，合作一方或多方不得将本合同项目部分或全部研究开发工作转让给第三人承担。但有下列情形之一的，合作一方或多方可以不经其他合作方同意，将本合同项目部分或全部研究开发工作转让给第三人承担：

1. 甲方、丙方书面同意乙方转让的研发工作。

第九条 在本合同履行中，因出现在现有技术水平和条件下难以克服的技术困难，导致研究开发失败或部分失败，并造成合作一方或多方损失的，合作各方约定按以下方式承担风险损失：

1. 已支付的乙方不再返还，还未支付的甲、丙方不再支付；
2. _____；
3. _____。

合作各方确定，本合同项目的技术风险按三方协商方式认定。认定技术风险的基本内容应当包括技术风险的存在、范围、程度及损失大小等。认定技术风险的基本条件是：

1. 本合同项目在现有技术水平条件下具有足够的难度；
2. 乙方在主观上无过错且经认定研究开发失败为合理的失败。

一方发现技术风险存在并有可能致使研究开发失败或部分失败的情形时，应当在10日内通知其他合作方并采取适当措施减少损失。逾期未通知并未采取适当措施而致使损失扩大的，应当就扩大的损失承担赔偿责任。

第十条 在本合同履行过程中，因作为研究开发标的的技术已经由他人公开（包括以专利权方式公开），合作一方或多方应在7日内通知其他合作方解除合同。逾期未通知并致使其他合作方产生损失的，其他合作方有权要求予以赔偿。

第十一条 合作各方确定按以下方式交付研究开发成果：

甲方：

1. 研究开发成果交付的形式及数量：无。

2. 研究开发成果交付的时间及地点：无。

乙方：

1. 研究开发成果交付的形式及数量：结题报告1份，活鸡智能调度系统1套；申请发明专利2项；论文3篇。

2. 研究开发成果交付的时间及地点：2026年12月31日，云浮温氏食品集团股份有限公司研究院。

丙方：

1. 研究开发成果交付的形式及数量：无。

2. 研究开发成果交付的时间及地点：无。

第十二条 合作各方确定，按以下标准及方法对合作一方完成的研究开发工作成果进行验收：

甲方：由甲方组织会议验收对乙方完成的研究开发工作成果进行验收；验收标准如下：

(1) 总体目标

1) 集成开发“基于时空大数据的温氏智能协同配送、调度执行系统”

一套；系统实现与集团苍穹平台接口对接，支撑数据实时交互。

2) 实现在广东养鸡公司 90%以上棚面应用。

3) 联合温氏集团申请发明专利 2 项以上，至少授权 1 项；

4) 发表高水平论文 3 篇以上。

(2) 技术考核指标

1) 系统运行效率指标：结果在 30 分钟以内

2) 智能车辆调度算法时间误差：配送车辆的行驶和等待时间用时与算法得出的预计用时合计偏差小于 30 分钟；

3) 智能配送算法匹配率：配送双方契合度，匹配度为 90%或以上；

4) 智能配送算法时间误差：时间匹配度，配送偏差小于 20 分钟。

(3) 经济考核指标

1) 实现在广东养鸡公司 90%以上棚面应用。

2) 节省不少于 10 个人力成本。

3) 调度环节工作效率提高 30%以上

4) 节约人员、车辆及开栏导致的耗损等各种成本部不少于 3000 万。

(4) 在公司的应用价值

1) 节约人工成本-将节约不少于 10 个人工成本/天。

①调度环节：经前期调研了解到目前广东养鸡公司有服务站 10 个以上，以 10 个计，每个服务站每天从下午结单开始，排班时长 3 个工时，调度员用语音通知客户，工作总时长 5 个工时，合计 8 个工时。本系统运行后实行自动化排班和通知，预计每个服务站节约 5 个工时，示范点共节约 50 个以上工时。

②分拣环节：广东养鸡公司日均出鸡量约 60 万只，磅位工作总时长 200 小时以上，因匹配度高，棚面压鸡减少等，整体分拣、结算效率提高，按分拣环节效率提高 5%计算，按每个磅位均需 5 人计算，共节约工时 50 个人工时。

2) 提升经济效益-事业部推广应用后, 节约成本合计不少于 3000 万
在温氏指定区域的生产端、销售端以及运营端等关键环节开展示范应
用, 节约人员、车辆、物料等各种, 整体工作效率提高 5%以上。

①节约调度、排班人员人力成本 100 万元;

②降低开栏数, 降低开栏因鸡的应激反应而导致的各种损耗, 预期降
低开栏数 10%左右, 有效降低开栏销售天数, 0.8 天, 节约成本 2000 万;

③降低运输成本, 通过整体规划, 就近运输, 降低车辆空置率, 防止
寻租等, 节约运费 1000 万;

④降低分拣损耗导致的损失。降低供货在棚面等待时长, 降低耗损 50
万。

3) 提高工作效率--不低于 5%

①本系统可降低客户的等待时长、减少跨服务站取货, 降低客户等待
总时长 10%;

②可降低养户在棚面等待时长、运输时长(就近运输)和抓鸡时长(集
中抓鸡装车), 降低总时长 10%

乙方: 按本合同约定的工作进度, 达成本合同第一条第 1 点的技术目
标。

丙方: _____。

第十三条 合作各方确定, 因履行本合同所产生、并由合作各方
分别独立完成的阶段性技术成果及其相关知识产权权利归属, 按第 1 种
方式处理:

1. 甲乙双方 (甲方/乙方/甲乙双方) 享有申请专利的权利。

专利权取得后的使用和有关利益分配方式如下: 另行约定。

2. 按技术秘密方式处理。有关使用和转让的权利归属及由此产生的
利益按以下约定处理:

(1) 技术秘密的使用权: _____;

(2) 技术秘密的转让权：_____；

(3) 相关利益的分配办法：_____。

合作各方对因履行本合同所产生、并由合作各方分别独立完成的阶段性成果及其相关知识产权权利归属，特别约定如下：知识产权共享。包括程序源代码、技术文档等全部共享至甲方，并完成相关知识转移。

第十四条 合作各方确定，因履行本合同所产生的最终研究开发技术成果及其相关知识产权权利归属，按第 1 种方式处理：

1. 甲乙双方 方享有申请专利的权利。

专利权取得后的使用和有关利益分配方式如下：另行约定。

2. 按技术秘密方式处理。有关使用和转让的权利归属及由此产生的利益按以下约定处理：

(1) 技术秘密的使用权：_____；

(2) 技术秘密的转让权：_____；

(3) 相关利益的分配办法：_____。

合作各方对因履行本合同所产生的最终研究开发技术成果及其相关知识产权权利归属，特别约定如下：知识产权共享。包括程序源代码、技术文档等全部共享至甲方，并完成相关知识转移。

第十五条 合作各方分别独立完成并与履行本合同有关的阶段性技术成果的研究开发人员，享有在有关此阶段性技术成果文件上写明技术成果完成者的权利和取得有关荣誉证书、奖励的权利。

合作各方应以协商方式确定最终研究成果的完成人员名单。此完成人员享有在有关最终技术成果文件上写明技术成果完成者的权利和取得有关荣誉证书、奖励的权利。

第十六条 合作各方确定，任何一方有权利利用本合同项目研究开发所完成的技术成果，进行后续改进。由此产生的具有实质性或创造性技术进步特征的新的技术成果，归 合作各方（完成方、合作各方）方所

有。具体相关利益的分配办法如下：_____

第十七条 为有效履行本合同，合作各方确定，在本合同有效期内，甲方指定黄明亮为甲方项目联系人，乙方指定王金凤为乙方项目联系人，丙方指定王凯为丙方项目联系人。项目联系人承担以下责任：

一方变更项目联系人的，应当及时以书面形式通知其他各方。未及时通知并影响本合同履行或造成损失的，应承担相应的责任。

第十八条 合作各方确定，出现下列情形，致使本合同的履行成为不必要或不可能的，可以解除本合同：

1. 因发生不可抗力和技术风险；

第十九条 合作各方因履行本合同而发生的争议，应协商、调解解决。协商、调解不成的，确定按以下第1种方式处理：

1. 提交云浮仲裁委员会仲裁；
2. 依法向人民法院起诉。

第二十条 本合同一式四份，具有同等法律效力。

第二十一条 本合同经合作各方签字盖章后生效。

以下无正文，转签章页



甲方：温氏食品集团股份有限公司 (盖章)

法定代表人/委托代理人： (签名)



年 月 日



乙方：华南农业大学 (技术合同专用章)

法定代表人/委托代理人： (签名)

项目负责人： (签字)

薛红已
王会凤

2024年 5月 20日



丙方：温氏食品集团股份有限公司研究院 (盖章)

法定代表人/委托代理人： (签名)

项目负责人： (签字)

黄明亮

2024年 5月 20日




3.主持：关于基于大数据挖掘的农村耕地评价及智能决策系统研究项目的立项通知与示范项目的合同

广东省科技计划项目合同书

受理编号: c1530550100086

项目编号: 2015A020209150

文件编号: 粤科规财字[2015]150号



2015A020209150

广东省省级科技计划项目 合同书

项目名称: 基于大数据挖掘的农村耕地评价及智能决策系统研究

计划类别: 农村科技领域

项目起止时间: 2015-06-01 至 2017-12-31

管理单位(甲方): 广东省科学技术厅

承担单位(乙方): 华南农业大学

乙方主管部门(丙方): 华南农业大学

通讯地址: 广东省广州市天河区五山路483号

邮政编码: 510642 单位电话: 020-38632819

项目负责人: 王金凤 联系电话: 020-85280320

项目联系人: 王金凤 联系电话: 020-85280320

广东省科学技术厅
二〇一四年制

三、项目进度和阶段目标

开始日期	结束日期	主要工作内容
2015-06-01	2015-12-31	(1) 进行前期数据整理, 基于模糊积分构建评估模型 (已经展开部分工作); (2) 将初建模型应用到经典数据库, 评价性能, 以改善调整;
2016-01-01	2016-12-31	(1) 完善耕地数据库, 建立指标筛选体系; (2) 进入基层调研, 衡量筛选指标的科学性, 并补充数据; (3) 以上工作进行期间, 同时完善评价模型;
2017-01-01	2017-12-31	(1) 集成耕地利用评估模型和指标筛选技术, 构建决策支持系统平台; (2) 整理研究成果, 发表论文。

六、人员信息

项目负责人情况								
姓名	年龄	性别	职称	职务	学历	在项目中承担的任务	所在单位	签名
王金凤	37	女	讲师	科研秘书、计算机应用教研室主任	博士研究生	负责项目的整体规划和实施, 算法的提出和验证等	华南农业大学	王金凤

主要研究开发人员								
姓名	年龄	性别	职称	职务	学历	在项目中承担的任务	所在单位	签名
刘财兴	53	男	教授	院长	博士	农业耕地大数据建库	华南农业大学	刘财兴
王文中	39	男	讲师	教师	博士	目标规划, 政策建议	华南农业大学	王文中
孙微微	44	女	副教授	教师	博士	耕地指标体系构建	华南农业大学	孙微微
王希	32	男	讲师	教师	硕士	算法实现	华南农业大学	王希
陈润健	23	男	未取得	研究生	本科	编程实现及系统架构	华南农业大学	陈润健
陈长成	24	男	未取得	研究生	本科	耕地数据的收集、预处理和分析	华南农业大学	陈长成

九、本合同签约各方

管理单位(甲方): 广东省科学技术厅 (盖章)

单位地址: 广州市连新路171号大院信息大楼

法定代表人(或授权代表): 黄宁生 (签字)

联系人(经办人)姓名: 林振亮 (签字)

Email: linzj@gdstc.gov.cn

电话: 020-83163905

年 月 日



承担单位(乙方): 华南农业大学 (盖章)

二级部门: 华南农业大学数学与信息(软件)学院

单位地址: 广东省广州市天河区五山路483号

法定代表人(或法人代理): 陈晓阳 (签字)

联系人(项目主管)姓名: 石睿 (签字)

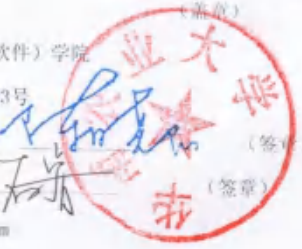
Email: 77909213@qq.com

电话: 020-85283435

开户单位名称: 华南农业大学

开户银行及帐号: 广东广州工行五山支行 3602002609000310520

年 月 日



乙方主管部门(丙方): 华南农业大学 (盖章)

单位地址: 广东省广州市天河区五山路483号

法定代表人(或法人代理): 陈晓阳 (签字)

联系人(项目主管)姓名: 石睿 (签字)

Email: 77909213@qq.com

电话: 020-85283435

开户单位名称: 华南农业大学

开户银行及帐号: 广东广州工行五山支行 3602002609000310520

年 月 日



4.主持：关于面向病毒基因大数据分析的若干问题及关键技术研究项目的立项通知

广州市科技计划项目合同书

项目编号： 201804010353

**广州市科技计划项目
合同书**

项目名称： 面向病毒基因大数据分析的若干问题及关键技术研究研究

计划类别： 科学研究计划

专题名称： 一般项目专题

起止时间： 2018年04月01日 至 2020年03月31日

承担单位： 华南农业大学

组织单位： 华南农业大学

责任处室： 社会发展与基础研究处

填表日期： 2018年02月09日

**广州市科技创新委员会
(2017年版)**

填写说明

一、本合同书的项目编号由市科技创新委员会（以下简称“市科创委”）统一确定。

二、本合同书由申报书在后台自动转换生成，如有错漏之处需修正，请联系市科创委项目责任处室退回承担单位修正。

三、经费预算中的“经费”是指项目执行过程中所发生的所有直接费用和间接费用。一般包括：劳务费、设备费、材料费、燃料动力费、测试化验加工费、会议/差旅/国际合作与交流费、出版/文献/信息传播/知识产权事务费、专家咨询费、其他直接费用和间接费用。

（一）劳务费：指项目研究开发过程中支付给项目组成员及项目组临时聘用人员的人力资源成本费，可包括项目临时聘用人员的社会保险补助。项目组成员所在单位有事业费拨款、工资性收入的部分，不得在财政资助的项目经费中重复列支。

（二）设备费：指项目研究开发过程中所必需的专用仪器、设备、样品、样机购置费及设备试制费，不得用于购置通用与办公用品。项目研发过程应尽量使用广州地区大型科学仪器协作共用网的仪器设备，避免重复购置。确需购置的，须经专家评审论证其必要性。

（三）材料费：是指在项目研究开发过程中需要消耗的各种原材料、辅助材料、低值易耗品、元器件、试剂、实验动物、部件、外购件、包装物的采购，运输、装卸、整理等费用。

（四）燃料动力费：是指在项目研究开发过程中相关大型仪器设备，专用科学装置等运行发生的可以单独计量的水、电、气、燃料消耗费用等。

（五）测试化验加工费：是指在项目研究开发过程中由于承担单位自身的技术、工艺和设备等条件的限制，必须支付给外单位（包括项目承担单位内部独立经济核算单位）的检验、测试、设计、化验及加工等费用。

（六）会议/差旅/国际合作交流费：指项目实施过程中发生的会议费、差旅费和国际合作交流费。在编制预算时，本科目支出预算不超过直接费用预算10%的，不需要编制测算说明。承担单位和科研人员应当按照实事求是、精简高效、厉行节约的原则，严格执行国家、本市和单位的有关规定，统筹安排使用。

（七）出版/文献/信息传播/知识产权事务费：是指在项目研究开发过程中需要支付的出版费、资料费、专用软件购买费、文献检索费、专业通信费、专利申请及其他知识产权事务等费用。

（八）专家咨询费：是指在项目研究开发过程中支付给临时聘请的咨询专家的费用。专家咨询费不得支付给参与项目研究及其管理相关的工作人员。专家咨询费的开支标准应当按照国家及本市有关规定执行。

（九）其他直接费用：是指在项目研究开发过程中发生的除上述费用外的其它直接支出。

(十) 间接费用：是项目承担单位在组织实施项目过程中发生的无法在直接费用中列支的相关费用。主要包括承担单位为项目研究提供的现有仪器设备及房屋、水、电、气、暖消耗，有关管理费用的补助支出，以及绩效支出等。其中绩效支出是指项目承担单位为提高科研工作绩效安排的相关支出。

四、本项目如涉及多家（包含两家）单位参加，乙方应在签定本合同书时与合作单位就任务分工、经费和知识产权分配等问题签订有关合同或协议（仅委托其他单位进行常规试验、提供社会化科技服务和少量辅助科研工作的情况除外），作为本合同书的附件。

五、项目承担单位是具有独立法人资格的企业、高等院校、科研院所或其他单位，高等院校、科研院所不具备独立法人资格的二级单位不得作为项目承担单位。

六、项目组织单位是指项目承担单位的上级主管部门。项目组织单位包括广州市各区科技行政主管部门，以及其他经市科技创新委同意作为项目组织单位的行政事业单位、企业集团、高等院校、科研院所等。除经市科技创新委同意或科技计划类别另有规定的外，项目组织单位为项目承担单位注册地所属的区科技行政主管部门。

七、项目基本信息表中单位特性指项目承担单位的资质或获得的称号，如高新技术企业、软件企业、技术先进型服务企业、创新型企业、科技小巨人企业等。单位类型按以下类型填写：高等院校、科研院所、国有企业、民营企业、股份制企业、中外合资企业、港/澳/台商投资企业、外商投资企业等。

八、本合同书适用于广州市事前资助类科技计划项目，有特殊要求另行制定的除外。

一、项目基本信息

项目名称	面向病毒基因大数据分析的若干问题及关键技术研究			
研究类别/ 所属技术领域	电子信息-软件技术-智能软件技术(含专家系统、机器翻译系统等、决策支持系统)			
承担单位	名称	华南农业大学		
	通信地址	广州市天河区五山路483号		
	邮政编码	510642	传 真	85281885
	单位特性		单位类型	高等院校
	统一社会信用代码 或组织机构代码	455416563		
	法定代表人	陈晓阳	电子邮箱	xychen@scau.edu.cn
	联系人	侯建国	联系电话	02038632413

二、项目组人员信息

姓名	王金凤	证件类型	汉族	身份证	证件号码	130105197810080929	性别	女
出生年月	1978年10月08日	民族	汉族	汉族	国籍	中国	学历	博士研究生
学位	博士	学位授予国家(或地区)	中国香港	中国香港	职务	教研室主任	职称	副教授
所学专业	计算机应用	手机号码	15815860686	15815860686	办公电话	15815860686	电子邮箱	wangphoenix@163.com

项目组成员(含项目负责人)

序号	姓名	证件类型	证件号码	年龄	性别	职务	职称	学历	现从事专业	分工	所在单位	签名
1	王金凤	身份证	130105197810080929	38	女	教研室主任	副教授	博士研究生	计算机专业	项目负责人	华南农业大学	王金凤
2	刘财兴	身份证	440106196210111853	55	男	院长	教授	本科	计算机应用	项目规划	华南农业大学	刘财兴
3	田绪红	身份证	110108196609051473	51	男	系主任	教授	博士研究生	计算机应用	数据库构建	华南农业大学	田绪红
4	梁广锡	其他	E612174(1)	62	男	系主任	教授	博士研究生	计算机应用	算法创新	境外人员	梁广锡
5	王文中	身份证	13060319760220125X	41	男	系副主任	讲师(高校)	博士研究生	数学建模	数据分析建模	华南农业大学	王文中
6	王希	身份证	440106198302111817	34	男	高校教师	讲师(高校)	硕士研究生	计算机应用	数据库构建	华南农业大学	王希
7	陈润健	身份证	440103199209044838	25	男	硕士研究生	未取得	本科	计算机应用	智能平台构建	华南农业大学	陈润健

陈嘉杰	身份证	441900199305102415	24	男	硕士研究生	未取得	本科	计算机应用	程序实现	华南农业大学	陈嘉杰
周展博	身份证	440111199503162413	22	男	硕士研究生	未取得	本科	计算机应用	编程实现	华南农业大学	周展博

201804010353

三、项目实施内容

(一)项目概述

主要内容：第一，面向DNA/RNA序列的大数据分析关键技术研究；第二，创新重要基因位跟踪提取的方法研究；第三，创智能挖掘算法新融合机制，构建流行病智能决策诊断服务平台。

绩效目标：发表高水平论文3-5篇，申报发明专利和软件著作权各1项，培养研究生2-3名。

(二)项目研究内容

(研究目标和内容、拟解决的关键技术问题、主要创新点、采用的方法、技术路线以及工艺流程、其他)

4.1 研究目标和内容

(1) 研究目标

本项目的研究目标有以下几点：

- 1) 构建稀疏存储的病毒基因组数据库；
- 2) 深化模糊积分的扩展机制研究；
- 3) 病毒基因表达式数据挖掘；
- 4) 集成病毒智能诊断服务云平台。

(2) 研究内容

本项目旨在通过机器学习算法解决病毒基因组数据分析和挖掘问题，包括基因大数据获取、数据库建立、病毒数据分析和挖掘算法创新等问题，详细阐述如下：

(1) 病毒基因大数据的收集、存储和清洗

本项目将针对引发生物疾病的病毒基因组进行采集、检索、收集，并提出基因数据的稀疏存储优化算法，构建病毒基因大数据存储云平台，进行去噪、清洗、排序等预处理操作，提出统一的存储和访问标准。

(2) 基因序列中遗传标记查找与定位

识别各类病毒基因表达式序列中的碱基位是本项目的关键，拟采用模糊测度的非可加性反映各个基因位以及基因碱基组合的重要性，通过正则化方法对模糊测度进行求解，以筛减重要基因的个数，为分类器提供变异基因位置的明确信息。

(3) 基因大数据的挖掘模型构建研究

针对基因数据的特性进行扩展和创新，其中包括扩展模糊积分之多重模糊积分和指数模糊积分、关联规则学习以及随机森林法等。不同的基因组合作用可能会产生不同的病害效应，需要进行并行化的多层次信息挖掘，构建基因之间的关系网络。

(4) 智能诊断平台的构建

构建面向基因数据的智能分析和诊断系统，并基于互联网+技术，搭建智能诊断云服务平台，研发同步App软件，真正实现基因检测的云端服务。

4.2 拟解决的关键技术问题

(1) 如何快速学习模糊测度值

前期研究中已经尝试了L1正则化方法进行求解，简单快速，并已经针对此方法申报了发明专利，本项目中拟采用Lp范式正则化方法 ($0 < p \leq 1$)，预期更为精确。

(2) 如何选取扩展模糊积分的形式

何种模糊积分适用解决当前问题，需要进行数据分析以及学习而得被积函数的参数需要通过优化算法学习得到其最优解，以保证模糊积分以最佳状态出现。

(3) 病毒基因表达式序列中重要基因位的确定

序列中哪个基因或基因集合影响病毒产生，单靠生物技术无法判断，因此确定导致疾病的基因碱基位尤为重要，模糊测度恰恰可以表示基因的重要性。

4.3 主要创新点

- (1) 面向基因大数据创新了一系列数据挖掘关键技术；
- (2) 重要基因提取技术创新；

(3) 智能诊断系统平台的构建,完成基于模糊积分融合的面向DNA/RNA微序列的智能诊断系统,并在云计算框架下构建智能诊断云服务平台,提供在线服务支撑。

4.4 采用的方法、技术路线以及工艺流程

(1) 研究方法

本项目的开展从四个方向出发,包括基于扩展模糊积分的数据挖掘关键技术研究,基因大数据的整理与分析,基因提取算法创新,和面向DNA/RNA基因大数据的智能诊断系统平台构建。

- 1) 针对经典模糊积分投影后仍出现类间交叠的情况,提出多重模糊积分方法,用以将误分类数据继续拉伸至不同空间,正确分类;
- 2) 针对模糊积分中线性被积函数投影路径只能为直线的限制,提出了基于指数函数的模糊积分扩展形式,以便能够更为灵活地覆盖复杂的数据分布情况;
- 3) 在基于新方法的基础上进行分类器建模,并针对经典数据集进行测试和验证,理论和实验两条线并行,以期能够使模型更加优化;
- 4) 通过多渠道获取疾病的DNA和RNA基因数据,并进行整理、清洗、存储,继而进行分析挖掘;
- 5) 提出模糊测度的精简学习方法,用于解决基因数据的特征提取问题,已获得初步验证;
- 6) 采用Matlab编程并加入GUI界面,开发一套基于扩展模糊积分的疾病智能诊断系统平台;
- 7) 面向人类和动植物疾病基因大数据进行判断,从准确度(Accuracy)、敏感度(Sensitivity)和特异度(Specificity)三个方面验证算法的有效性,其中Sensitivity对于疾病诊断更为重要。

(2) 技术路线

- a. 首先,构建本项目相关的基因数据库,同时开展模糊积分扩展理论研究的技术论证以及公式验证工作;通过多种途径选取本项目所需要的DNA/RNA数据,之后完成存储、整理、清洗和分析过程;
- b. 继而进入技术体系创新阶段;
- c. 开展模糊积分扩展体系的深入研究,同时融入随机森林法、马尔科夫算法等算法;
- d. 面向长序列,创新重要基因提取算法,涉及到基于L1和Lp正则化方法的模糊测度值求解以及其所对应的重要基因位或基因组合的跟踪;
- e. 构建智能诊断系统需要用到模糊积分分类器以及对多种挖掘算法进行融合;
- f. 构建面向DNA/RNA基因大数据的智能诊断服务云平台,理论方面对我们的扩展模糊积分以及融合机制给予验证。

(三) 项目预期风险及规避措施

(预期风险,规避措施)

7.1 预期风险

(一) 政策风险分析

本项目符合《国家中长期科技发展规划纲要》中重点领域和优先主题的有关要求,结合《中华人民共和国国民经济和社会发展第十二个五年规划》,围绕国家发展目标和战略展开,遵守基因数据的保密规定,一切数据取自公开的基因数据库或者签署协议的数据来源。

(二) 技术风险分析

本项目组成员长期从事相关领域研究,发挥各自优势,以研究目标为导向,以技术难点研究为核心,发挥依托单位优势,保证高质量地完成课题预期目标,无技术风险。

(三) 市场风险分析

目前还没有基于数据挖掘和基因选取的智能诊断系统出现,急需开展相关研究予以支持,市场应用前景十分广阔,无市场风险。

综上所述,本项目有有严格的管理和考核办法,高水平的科研队伍和市场导向和需求,扎实的前期工作基础和充分的科研条件,故基本无风险。

7.2 规避措施

- (1) 加强团队管理,分工明确,以小课题划分形式,责任到人,保证项目的顺利开展与推进;
- (2) 严格遵守科研经费管理办法;
- (3) 组织专家顾问小组对项目实施过程中的技术、财务、进度、质量提供论证、评估、咨询和指导。

四、项目主要验收指标

(一) 主要技术指标

- (1) 发表高水平论文共计4篇：其中SCI检索2篇，国际会议论文2篇；
- (2) 申报专利1项，软件著作权1项；
- (3) 培养硕士研究生4名；
- (4) 完成一个智能诊断系统，软件形式，上线使用。

201804010353

(二) 主要技术成果			
序号	成果形式		成果数量
1	新产品 (或新材料、新装备、新品种/系)	无	0
2	新工艺 (或新方法、新模式、新技术、新服务)	无	0
3	发明专利 (件)	申请	1
		授权	0
4	实用新型专利 (件)	申请	0
		授权	0
5	外观设计专利 (件)	申请	0
		授权	0
6	国外专利 (件)	PCT受理	0
		授权	0
7	技术标准制定 (个)	牵头	0
		参与	0
8	软件著作权 (项)		1
9	论文论著 (篇)	SCI/EI/ISTP	2
		其他	2
10	创新平台 (载体) 项目	技术服务数量 (项)	0
		服务企业数量 (家)	0
11	获得国家级奖项 (项)		0
12	获得省级奖项 (项)		0
13	科技人才奖励 (人)		0
14	引进人才 (人)		0
15	培养人才 (人)	博士	0
		硕士	4
其他成果及形式说明 (创新特色、成果宣传推介措施等)			
无			

(三) 主要经济指标及社会效益		
序号	指标名称	指标值
1	实施期内项目销售收入(万元)	0
2	实施期内项目利税(万元)	0
3	实施期内项目出口创汇(万美元)	0
4	实施期内项目新增就业人数(人)	0
其他经济指标及社会效益说明		
无		

201804010353

五、项目经费预算

(单位: 万元)

项目经费: 20.00					
资金来源	小计	市科创委经费	自筹资金	其他财政经费	
2018年	20.00	20.00	0.00	0.00	
2019年	0.00	0.00	0.00	0.00	
2020年	0.00	0.00	0.00	0.00	
合计	20.00	20.00	0.00	0.00	
支出预算明细					
支出科目	小计	市科创委经费		自筹资金	
		经费额	用途说明	经费额	用途说明
一、劳务费	3.50	3.50	用于支付研究生培养费用及临聘人员劳务费	0.00	无
二、设备费	4.00	4.00	添置工作站两台	0.00	无
1.购置费	4.00	4.00	添置工作站两台	0.00	无
2.试制费	0.00	0.00	无	0.00	无
三、材料费	5.00	5.00	用于项目实施过程中所需的耗材补充费用	0.00	无
四、燃料动力费	0.00	0.00	无	0.00	无
五、测试化验加工费	0.00	0.00	无	0.00	无
六、会议/差旅/国际合作交流费	1.50	1.50	用于国际、国内交流合作产生的交通差旅费用	0.00	无
七、出版/文献/信息传播/知识产权事务费	2.00	2.00	用于论文出版、专利申请等费用	0.00	无
八、专家咨询费	2.00	2.00	用于支付专家评审、咨询费用	0.00	无
九、其它直接费用	0.00	0.00	无	0.00	无
十、间接费用	2.00	2.00	用于项目管理费和房屋、水电支出	0.00	无
合计	20.00	20.00	\	0.00	\

注: 自筹资金由项目申报单位自身筹措, 不包含来自各级政府部门的财政资金, 参照事前资助类项目经费管理, 经费预算按项目申报单位实际需要进行编制。

设备仪器购置明细

(金额单位: 万元)

序号	名称	购置经费来源	数量	预计费用		是否政府采购	备注
			台/套	单价	总价		
1	工作站	市科创委经费	2	2.00	4.00	是	用于数据库搭建及系统后台存储架构
合计			2	—	4.00	—	—

注:单台(套)价值在30万元及以上的设备仪器请及时办理进入“广东省科技资源共享服务平台”手续,以提高使用率。

201804010353

六、工作进度安排（以一年为阶段填报）

序号	起止时间	阶段目标主要内容及成果
1	2018-04-01至2019-03-31	1) 构建相关基因大数据库，完成数据的存储、清洗、聚类、存储等预处理过程； 2) 创新数据挖掘算法，尤其是扩展模糊积分，并构建相应的分类模型； 3) 完成分类器模型的优化和融合； 4) 发表相关论文，申报专利及软件著作权；
2	2019-04-01至2020-03-31	1) 对分类器模型进行测试和完善； 2) 构建疾病基因智能诊断平台； 3) 开发App软件，实现移动端应用； 4) 整理研究成果，完成项目结题。

七、共同条款

第一条 甲、乙、丙三方根据《中华人民共和国合同法》及国家有关法律和规定，经协商一致，特订立本合同，作为甲、乙、丙三方在合同执行中共同遵守的依据。

第二条 甲、乙、丙三方应当严格履行《广州市科技计划项目管理办法》（穗科创规字〔2017〕1号）中规定的职责，严格按照《广州市科技创新发展专项资金管理办法》（穗科创规字〔2017〕4号）实施项目经费管理。

第三条 甲方应在项目执行期满（执行期以本合同“六、工作进度安排”为准，下同）时按相关管理办法组织项目验收。

1. 对通过验收的项目，签发《验收证书》；
2. 对未通过验收的项目，要求其承担单位限期整改，整改后仍不能通过验收的，终止合同，收回尚未使用和使用不符合规定的财政经费。

第四条 乙方应：

1. 按照《合同书》规定的内容组织实施项目，接受并配合甲方、丙方以及各级财政、审计部门，或上述部门委托的机构进行评估、稽查、审计、检查和绩效评价，并按要求提供项目任务与预算执行情况和有关财务资料；
2. 保证自筹资金按时到位和其它配套条件的落实；
3. 在项目研究开发过程中优先考虑使用“广东省科技资源共享服务平台”的仪器设备，项目购置的设备仪器若符合入网条件应及时办理入网手续对社会共用共享，提高设备仪器的使用率。按照《中华人民共和国采购法》要求，对符合政府采购范围的设备仪器，执行政府采购；
4. 项目合同执行期内确需进行变更的，须按照《广州市科技计划项目管理办法》（穗科创规字〔2017〕3号）、《广州市科技创新发展专项资金管理办法》（穗科创规字〔2017〕4号）相关程序办理；
5. 项目合同执行期满3个月内向甲方提出验收申请，提前完成合同规定任务的可提前申请验收；
6. 项目未通过验收的，按相关管理办法限期整改并重新提出验收申请；
7. 办理《验收证书》和科技成果登记手续；
8. 按照科技经费管理相关要求对项目资金单独设帐，按照预算专款专用；
9. 项目验收时，须提交科技报告；
10. 对合作单位承担监管责任，与合作单位签署合作协议，作为本合同的附件，因合作单位违反合作协议或其他导致本合同书项目建设未能按约定完成的行为，应向甲方承担违约责任。

第五条 乙方因某种原因（如技术或市场情况发生变化，项目所依托的技术、资金、设备仪器或人力条件不能落实，原定技术方案及路线不合理等）或不可抗力因素，致使项目计划无法执行，须终止合同的，应向甲方提出申请，经丙方同意，由甲方审核批准，收回尚未使用和使用不符合规定的财政经费；如乙方没有提出终止申请，甲方根据项目研究开发过程监督检查情况，有权终止项目，收回财政经费；乙方在执行期满无故不提交验收申请，经甲方催办仍不提交的，甲方有权终止项目，收回财政经费，因乙方不及时报告或申请所导致的各方损失，由乙方承担。乙方违反约定造成项目工作停滞、延误或失败，未能通过验收的，应承担违约责任。

第六条 丙方应：

1. 协助甲方对项目的实施过程进行跟踪、检查和提供相关信息，并对所提供信息的客观真实性负责；

2. 负责监管乙方严格遵守《合同书》规定的任务；

3. 督促乙方按时到位自筹资金并保证和落实其他配套条件；

第七条 在履行本合同的过程中，当事人一方发现可能导致项目失败或部分失败的情形时，应及时通知另一方，并采取适当措施减少损失，没有及时通知并采取适当措施，致使损失扩大的，应当就扩大的损失承担责任。

第八条 在履行本合同的过程中，如遇到市财政计划改变等不可抗力情况，甲方对所核拨经费的数量和时间可进行相应变更。

第九条 本项目技术成果及知识产权的归属、转让和实施技术成果所产生的经济利益的分享，除另有约定外，按国家和省、市有关规定执行；正式发表的论文、论著应标注“广州市科技计划项目资助”字样及项目编号；项目所取得的技术成果和知识产权应优先广州产业化或推广转让；需向外地转让或产业化的，须事先以书面形式征得甲方同意。

第十条 属技术保密的项目，经协商订立如下技术保密条款：

1. 本合同书保密内容范围为：本合同及其补充协议和附件。乙方因履行本合同所接触或知晓的甲方工作秘密（包括但不限于甲方的任何技术性资料、以及甲方为完成本合同提供的任何其他信息资料并且在提供时未说明是公开信息的）；

2. 本合同书保密期限为：\；

3. 乙方（包括但不限于乙方雇员、代理人、顾问等人员，下同）采取有效的保密措施以避免泄露给任何第三方；在本合同有效存续期间及合同终止后，未经甲方事先的书面同意，不得以任何方式公布、发表、公开、披露、散播、复制此种保密信息的任何部分，或对其加以任何形式的利用或使用；如甲方要求，乙方必须签署甲方提供的保密协议。乙方应与可能知悉保密内容的人员签订技术保密保护协议，保密义务不得低于本合同书的约定；

4. 双方应建立技术保密制度；

5. 属技术保密的项目必须经市负责技术保密部门审查后，方可确定可否发表或用于国际合作与交流。

第十一条 根据项目具体情况，经协商订立如下附加条款作为本合同正式内容的一部分：

1. 甲方同意给予乙方人民币（¥20.00万）的资助，立项后一次性拨付；

2. \

第十二条 违约责任

乙方无正当理由造成项目工作停滞、延误或失败，未能通过验收的，甲方有权终止项目，收回财政经费，由此造成的经济损失由乙方承担；经检查确认项目计划进度不符合合同书约定的，甲方有权警告并责令乙方整改，由此产生的损失由乙方负担；情节严重的，甲方有权终止项目，收回财政经费。

第十三条 廉洁责任

乙方应严格遵守国家、省、市关于科技专项经费使用的有关法律、法规，相关政策以及廉洁建设的各项规定，建立健全促进科研诚信的规章制度，积极开展人员廉洁从业、诚信科研教育，防范科技项目组成员在科研活动中出现下列违法违规行为：

1. 在项目申报、研发过程中提供虚假信息或材料，抄袭、剽窃他人科研成果，捏造、变造或篡改科研数据；

2. 向甲方、组织单位、评审机构及其工作人员赠送礼金、有价证券、任何形式的贵重物品和回扣、好处费、感谢费等；

3. 为甲方、组织单位、评审机构及其工作人员报销应由对方或个人支付的费用；

4. 为甲方、组织单位、评审机构工作人员个人装修住房、婚丧嫁娶、配偶子女的工作安排以及出国（境）、旅游等提供方便；

5. 为甲方、组织单位、评审机构及其工作人员组织有可能影响公正执行公务的宴请、健身、娱乐等活动；

6. 其他：\

乙方及其工作人员有上述行为之一的，一经查实，甲方有权按照科研诚信管理规定采取终止项目合同、不拨付财政经费、限制项目申报等处理；涉嫌犯罪的，移交司法机关追究刑事责任。甲方、组织单位、评审机构及其工作人员有涉及上述行为之一的，乙方应及时向甲方或其上级机关或纪检监察、司法等有关机关检举举报。

第十四条 争议解决

因本合同书所产生的争议，各方应友好协商解决；协商不成的，各方同意由本合同签订地人民法院管辖。

第十五条 通知与送达

甲方在本合同履行过程中向乙方或丙方发出或者提供的所有通知、文件、文书、资料等，均以本合同所列明的乙方或丙方地址送达。乙方或丙方如果迁址，应当书面通知甲方；未履行书面通知义务的，甲方按原地址邮寄相关材料即视为已履行送达义务。当面交付上述材料的，在交付之时视为送达。

本合同一式五份，各份具有同等效力。甲方、市财政局和丙方各存一份，乙方存二份。本合同签订各方均负有相应的法律责任，不受机构、人事变动而影响。

说明：本《合同书》中，凡是三方约定无需填写的条款，在该条款的空白处划（\）。

八、合同书各方签章

签订地点：广州市越秀区

广州市科技创新委员会（甲方）：广州市科技创新委员会

项目经办人（签章）：

陈洁

联系电话：83124145

责任处室负责人（签章）：

华国彭



项目承担单位（乙方）：华南农业大学

二级部门：华南农业大学数学与信息学院

项目负责人（签章）：

郑爱明

财务负责人（签章）：

财务负责人联系电话：02085287402

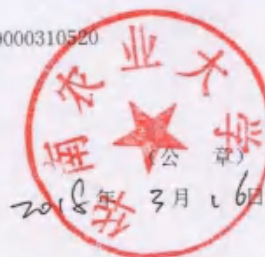
帐户名：华南农业大学

帐号：3602002609000310520

开户银行：广东广州工行五山支行

法定代表人（签章）：

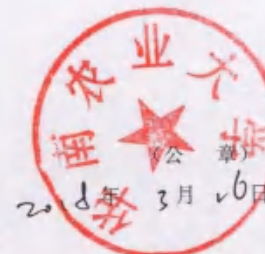
侯建国



组织单位（丙方）：华南农业大学

项目经办人（签章）：

侯建国



5.主持：关于面向多模态抗疫中医药功能大数据的“天盾”平台构建及示范应用项目的立项通知和合同

关于 2024 年度中医药广东省实验室 科研培植项目立项的通知

尊敬的刘良院士：

您好！

根据《中医药广东省实验室（横琴实验室）培植项目管理办法（试行）》（附件1）、《中医药广东省实验室（横琴实验室）2024年度科技研发培植项目申报指南》等文件要求，经申报、形式审查、专家函审、会议审议、立项公示等程序，您申报的培植项目“面向多模态抗疫中医药大数据‘天盾’平台的构建及示范应用”已正式批复立项。

请项目负责人认真落实主体责任，按要求签订**项目任务书**（附件2）与**项目合作协议**（附件3），明确项目实施内容、预期成果和量化指标、项目进度和阶段考核目标、资金预算方案。严格项目过程管理，加强与参与单位之间的衔接与协调。强化项目资金管理，做到财政资金专款专用，确保项目研究开发目标和任务按期完成，提高资金使用绩效。

一、立项项目详情

项目名称	面向多模态抗疫中医药大数据“天盾”平台的构建及示范应用
项目编号	HQL2024PZ001
项目负责人	刘良
项目依托单位	中医药广东省实验室
项目起止时间	2024年9月1日-2026年8月31日
项目经费	200万元

二、签订流程

1. 填写项目任务书:

a) 项目负责人于 **2024年8月16日** 前按照要求**填报并提交电子版**的中医药广东省实验室（横琴实验室）2024年科研培植项目任务书至电子邮箱：**hqlabkjyf@gzucm.edu.cn** 进行审核。

b) 审核通过后，提交科研培植项目任务书**(壹式叁份)**至横琴实验室科技研发部，**提交时间**：2024年8月28日前；**提交地点**：珠海市香洲区厚朴道398号横琴国际科技创新中心7栋608-3。

2. 签订合作协议：请遵照《中医药广东省实验室（横琴实验室）培植项目管理办法（试行）》要求签订合作协议，**合作单位签字盖章版**的中医药广东省实验室（横琴实验室）2024年科研培植项目合作协议**(壹式肆份)**同项目任务书纸质版一并提交。

联系人：黄老师、漆老师

联系电话：15928425723、18679565176

电子邮箱：hqlabkjyf@gzucm.edu.cn

附件：

1. 《中医药广东省实验室（横琴实验室）培植项目管理办法（试行）》
2. 中医药广东省实验室（横琴实验室）2024年科研培植项目任务书
3. 中医药广东省实验室（横琴实验室）2024年科研培植项目合作协议(模版)

横琴实验室科技研发部
2024年8月12日

中医药广东省实验室（横琴实验室）科技研发 培植项目合作协议

甲方（项目依托单位、牵头申报单位）：中医药广东省实验室

乙方（全部合作单位）：（根据项目团队情况和指南限定添加或减少）

项目联合单位 1：广东省中医院

联系人：潘胡丹

项目联合单位 2：华南农业大学

联系人：王金凤

经甲乙双方协商一致，就共同申报中医药广东省实验室（横琴实验室）科技研发培植项目（课题）达成如下协议：

一、任务分工：

1. 申报项目名称：面向多模态抗疫中医药功能大数据的“天盾”平台构建及示范应用
2. 甲方项目负责人为：刘良
3. 甲乙双方在所申报项目中承担角色、研究内容、任务分工：
甲方：项目承担单位。负责项目总体设计、经费管理、审计及结题验收等工作。具体承担抗疫功能靶点与中医药数据智慧模型库构建。
乙方：合作单位。乙方配合甲方做好项目的实施、进展汇报、结题等工作，具体承担基于疫病核心功能靶点构建筛选及中医药抗疫数据库平台的设计与构建。
4. 全体申请人承诺按照指南总体部署参与共同推进项目任务的整体实施，圆满完成项目目标。

二、经费与预算

1. 专项经费的分配方案：根据《中医药广东省实验室（横琴实验室）科技研发培植项目管理办法（试行）》，科研经费分配以广东省内机构为主，原则上在横琴和珠海市研发的项目经费比重不低 60%，省内其他地区研发项目经费比重占 20%~30%，省外合作单位研发经费比重占 10%~20%。甲乙双方商定按照 6:4 的比例分配编制预算，其中乙方项目联合单位 1 分配比例为 10%，项目联合单位 2 分配比例为 30%。

2. 预算编制按国家相关管理规定的比例上限足额预算项目经费中的间接费用，甲方承诺按项目经费分配比例分配给乙方相应间接费用，或双方协商确定间接经费分配方案。

三、知识产权约定

1. 甲乙双方在申请本项目之前各自获得的知识产权及相应权益归各自所有。
2. 横琴实验室资助的培植项目成果，科研成果归属横琴实验室所有，参与人员及参与单位可根据贡献分享成果产生的收益。鼓励申请发明专利及开发具有应用潜力和价值的产品成果。专利须以“中医药广东省实验室（英文：Chinese Medicine Guangdong Laboratory）”为专利权人。
3. 基金项目研究形成的论文、专著、软件、数据库、专利以及鉴定、获奖、成果报道等，须注明中医药广东省实验室项目资助及批准号，若以发表论文为成果，须以“中医药广东省实验室（英文：Chinese Medicine Guangdong Laboratory）”作为第一作者或最后通讯作者的第一单位，且研究成果须与项目合同中的研究目标和研究内容相符。

四、数据递交、人类遗传资源、伦理条款

1. 甲乙双方承诺各领域项目产生的所有科学数据无条件、按期递交到横琴实验室指定的平台，在本项目约定的条件下对本项目各个承担单位，乃至今后面向所有的科技工作者和公众开放共享。
2. 甲乙双方签署数据递交协议后而不在商定的期限内履行数据递交责任的，则由横琴实验室责令整改，拒绝整改者，则由横琴实验室追回项目资金，并予以通报。
3. 甲乙双方承诺本研究涉及人类遗传资源采集、收集、买卖、出口、出境等须遵照《人类遗传资源管理暂行办法》相关规定执行。
4. 甲乙双方承诺本研究涉及实验动物和动物实验，要遵守国家实验动物管理的法律、法规、技术标准及有关规定，使用符合要求的合格动物进行实验，在合格设施内进行动物实验，保证实验过程合法，实验结果真实、有效，保障动物福利，并通过实验动物福利和伦理审查。

五、技术情报和资料的保密

1. 甲乙双方及其有关人员均应遵照《中华人民共和国保守国家秘密法》和科学技术部《科技保密规定》等国家有关法律法规的要求，承担保密责任，采取保密措施。
2. 项目执行过程中，双方共享但未公开的技术资料及数据，甲乙双方均负有保密责任，应采取相应的保密措施。未经双方许可，任何一方不得向第三方泄露数据、技术、图片、视频等相关资料。

六、违约责任

一方违约造成另一方损失的，违约方应对守约方的实际损失予以赔偿，赔付上限不超过项目总额。

七、争议的解决办法

如本协议有未尽事宜，甲乙双方应本着友好协商的原则进行商议，也可请求主管部门进行调解。

八、其他约定

1. 本任务书自双方签字盖章后生效。对本任务书任何条款的修改、补充或更改，双方必须重新签订书面协议并签字盖章。
2. 本协议壹式拾贰份，甲方持捌份，乙方各持贰份，具有同等法律效力。
3. 未尽事宜由双方另行商议。

甲方（盖章）



项目依托单位：中医药广东省实验室

法定代表人/委托代理人：



项目负责人：

Handwritten signature

日期：2024年 8 月 31 日

乙方（盖章）

项目联合单位 1：广东省中医院

法定代表人/委托代理人：

联系人：

日期：2024年 8 月 29 日

项目联合单位 2：华南农业大学

法定代表人/委托代理人：

联系人：

日期：2024年 8 月 22 日



Handwritten signature



Handwritten signature

6.主持：关于面向温氏活鸡销售的智能调度模型研究与开发项目的合同

HXKJHJ2020940

合同编号： _____

技术开发合同

项目名称：面向温氏活鸡销售的智能调度模型研究与开发

委托方：温氏食品集团股份有限公司勒竹分公司
(甲方)

受托方：华南农业大学
(乙方)

签订时间：2020.10.10

签订地点：广州市天河区华南农业大学

有效期限：2020年10月至2021年10月

中华人民共和国科学技术部印制

填写说明

一、本合同为中华人民共和国科学技术部印制的技术服务合同示范文本，各技术合同认定登记机构可推介技术合同当事人参照使用。

二、本合同书适用于一方当事人（受托方）以技术知识为另一方（委托方）解决特定技术问题所订立的合同。

三、签约一方为多个当事人的，可按各自在合同关系中的作用等，在“委托方”、“受托方”项下（增页）分别排列为共同委托人或共同受托人。

四、本合同书未尽事项，可由当事人附页另行约定，并作为本合同的组成部分。

五、当事人使用本合同书时约定无需填写的条款，应在该条款处注明“无”等字样。

技术开发服务合同

委托方（甲方）：温氏食品集团股份有限公司勒竹分公司

住 所 地：广东省云浮市新兴县簕竹镇榄根

法定代表人：梁志勇

项目联系人：梁凯敏

联系方式：13580659982

通讯地址：广东省云浮市新兴县簕竹镇榄根

电 话：13580659982 传真：无

电子信箱：无

受托方（乙方）：华南农业大学

住 所 地：广东省广州市天河区五山路 483 号华南农业大学

法定代表人：刘雅红

项目联系人：王金凤

联系方式：15815860686

通讯地址：广东省广州市天河区五山路 483 号华南农业大学

电 话：15815860686 传真：无

电子信箱：wangjinfeng@scau.edu.cn

本合同甲方和乙方就共同参与研究开发 面向温氏活鸡销售的智能调度模型研究与开发 项目，经过平等协商，在真实、充分地表达各自意愿的基础上，根据《中华人民共和国合同法》的规定，达成如下协议，并由双方共同恪守。

第一条：技术研究开发的内容如下：

1. 技术目标：(1) 构建一套活鸡销售客户智能调度模型；(2) 实现一个基于智能决策的养户供货排班系统；

2. 技术内容：(1) 活鸡配送时空网络流模型及实现 (2) 订单（客户）分配调度模型与实现；(3) 养户活鸡配送调度模型与实现；(4) 活鸡配送反应式调度模型与实现；(5) 服务台及棚面调度模型与实现。

3. 技术方式：(1) 构建智能模型；(2) 完成模拟测试 (3) 自交付使用起，免费维护模型部分一年。

第二条：乙方应按下列要求完成技术服务工作：

1. 技术服务地点：数信学院九楼的温氏大数据中心

2. 技术服务期限：2020.10-2021.08

3. 技术服务进度：(1) 2020.10.1-2020.10.10: 数据准备和模型创建阶段；(2) 2020.10.11-2020.10.31: 完成基本的订单匹配和养户排单模型；(3) 2020.11-2021.02:模型继续优化，以较优方案完成调度。

4. 技术服务质量要求：无

5. 技术服务质量期限要求：无

第三条：为保证乙方有效进行技术服务工作，甲方应当向乙方提供下列工作条件和协作事项：

1. 提供技术资料：

(1) 构建模型所需的相关历史数据

(2) 实际业务环节咨询解答

(3) 业务流程中的特殊情况和需要注意的相关信息

2. 提供工作条件:

(1) 大数据服务器访问权限开放

(2) 工作人员测试平台

3. 其他: 无

4. 甲方提供上述工作条件和协作事项的时间及方式: 项目执行期间可随时沟通协作, 方式为实地调研结合线上会议

第四条: 甲方向乙方支付技术服务报酬及支付方式为:

1. 技术服务费总额为: 贰拾万元

2. 技术服务费由甲方 按预付款, 进度款, 验收结算全款的方式 支付乙方。

双方约定: 合同金额共计 20 万元整, 签订合同之日起 10 日内甲方向乙方支付 6 万元, 作为预付款; 2020 年 10 月 31 日前乙方模型建立并运行 3 日内, 甲方向乙方支付 8 万元作为进度款; 项目验收完成, 甲方一次性支付余额 6 万元。

具体支付方式和时间如下:

甲方开户银行名称、地址和账号为:

开户银行: 中国工商银行股份有限公司新兴支行

地 址: 温氏食品集团股份有限公司勒竹分公司

账 号: 2020 0032 2920 0107 936

支付时间: 模型正常运行后一次性支付——2.1

乙方开户银行名称、地址和账号为：

开户银行：广州工行五山分行

地 址：广东广州市天河区五山 483 号华南农业大学

账 号：3602 0026 0900 0310 520

第五条：双方确定因履行本合同应遵守的保密义务如下：

甲方：

1. 保密内容（包括技术信息和经营信息）：调度模型相关技术和模型使用信息。

2. 涉密人员范围：养殖事业部和信息中心工作人员；

3. 保密期限：2020-10 至 2025-09。

4. 泄密责任：保留追究相关人员的法律责任。

乙方：

1. 保密内容（包括技术信息和经营信息）：温氏广东养鸡公司提供的相关数据和信息，仅用作项目建模使用，不得对外公开和共享。

2. 涉密人员范围：项目组所有成员；

3. 保密期限：2020-10 至 2025-09。

4. 泄密责任：保留追究相关人员的法律责任。

第六条：本合同的变更必须由双方协商一致，并以书面形式确定。但有下列情形之一的，一方可以向另一方提出变更合同权利与义务的请求，另一方应当在3日内予以答复；逾期未予答复的，视为同意：

第七条：双方确定以下列标准和方式对乙方的技术服务工作成果

第十三条：双方确定：本合同及相关附件中所涉及的有关名词和技术术语，其定义和解释如下：

1. 无
2. 无
3. 无
4. 无
5. 无

第十四条：与履行本合同有关的下列技术文件，经双方确认后，无 为本合同的组成部分：

1. 技术背景资料：无
2. 可行性论证报告：无
3. 技术评价报告：无
4. 技术标准和规范：无
5. 原始设计和工艺文件：无
6. 其他：无

第十五条：双方约定本合同其他相关事项为：无

第十六条：本合同一式 叁 份，具有同等法律效力。

第十七条：本合同经双方签字盖章后生效。

甲方：温氏食品集团股份有限公司勒竹分公司

(盖章)

法定代表人/委托代理人：张明 (签名)

2020年10月10日

乙方：华南农业大学 (盖章)

法定代表人/委托代理人：刘维红 (签名)

2020年10月10日

7.主持：关于恩平禽畜薯现代循环经济智慧示范园可行性报告项目的合同

HXKJH72023 13 11

合同编号：

技术咨询合同

项 目 名 称：恩平禽畜薯现代循环经济智慧示范园可行性报告

委托方（甲方）：恩平市瑞帮农业科技有限公司

受托方（乙方）：华南农业大学

签 订 时 间：2023.03

签 订 地 点：华南农业大学

有 效 期 限：2023.03.19-2024.3.18

中华人民共和国科学技术部印刷

填写说明

一、本合同为中华人民共和国科学技术部印制的技术咨询合同示范文本，各技术合同登记机构可推介技术合同当事人参照使用。

二、本合同书适用于一方当事人（受托方）为另一方（委托方）就特定技术项目提供可行性论证、技术预测、专题技术调查、分析评价报告所订立的合同。

三、签约一方为多个当事人的，可按各自在合同关系中的作用等，在“委托方”、“受托方”项下（增页）分别排列为共同委托人或共同受托人。

四、本合同书未尽事项，可由当事人附页另行约定，并可作为本合同的组成部分。

五、当事人使用本合同书时约定无需填写的条款，应在该条款处注明“无”等字样。

技术咨询合同

委托方（甲方）： 恩平市瑞帮农业科技有限公司
项目联系人： 孙丙申
联系方式
电 话： 13363655956
受托方（乙方）： 华南农业大学
住 所 地： 广东省广州市天河区五山路 483 号
项目联系人： 王金凤
联系方式
通讯地址： 广东省广州市天河区五山路 483 号华南农业大学数学与信息学院
电 话： 15815860686
电子信箱： wangjinfeng@scau.edu.cn

本合同甲方委托乙方恩平禽畜薯现代循环经济智慧示范园可行性报告项目进行技术咨询，并支付咨询报酬。双方经过平等协商，在真实、充分地表达各自意愿的基础上，根据《中华人民共和国合同法》的规定，达成如下协议，并由双方共同恪守。

第一条 乙方进行技术咨询的内容、要求和方式：

- 1、咨询内容：恩平禽畜薯现代循环经济智慧示范园可行性，纲要见附件 1。
- 2、咨询要求：完成“恩平禽畜薯现代循环经济智慧示范园可行性投资报告”，不少于 8 万字。
- 3、咨询方式：（1）提供咨询报告；（2）现场交流。

第二条 乙方应当按照下列进度要求进行本合同项目的技术咨询工作：

1、2023.3.19-2023.06.30：调研、资料搜集、整理，细分提纲

2、2023.7.01-2023.9.30：“恩平禽畜薯现代循环经济智慧示范园可行性投资报告”初稿

3、2023.10.01-2023.12.31：“恩平禽畜薯现代循环经济智慧示范园可行性投资报告”讨论稿

4、2024.1.1-2024.2.29：“恩平禽畜薯现代循环经济智慧示范园可行性投资报告”。

第三条 为保证乙方有效进行技术咨询工作，甲方应当向乙方提供下列协作事项：

1、提供技术资料：

(1) 恩平产业（养猪、养鸡、南瓜、马铃薯、水稻）等产业资料；恩平土地规划方案、恩平环保等项目相关资料；

甲方提供上述协作事项的时间及方式：双方协商。

第四条 甲方向乙方支付技术咨询报酬及支付方式为：

1、技术咨询报酬总额为：贰拾万元人民币（¥200,000.00）

2、技术咨询报酬由甲方分两次支付乙方。

具体支付方式和时间如下：

(1) 因前期启动经费需求大，在合同签订后的30个工作日内支付课题经费的百分之五十，即壹拾万元人民币；

(2) 在课题完成并通过鉴定的5个工作日内，支付课题经费的百分之五十，即壹拾万元人民币；

(3) 甲方须在收到乙方提交研究成果30天内组织鉴定，逾期未予以鉴定的，视为通过鉴定，须在逾期日开始的3个工作日内按上述(2)

点支付课题经费余额：

乙方开户银行名称、地址和账号为：

开户银行： 广州工行五山支行

户 名： 华南农业大学

帐 号： 3602002609000310520

第五条 双方确定因履行本合同应遵守的保密义务如下：

甲方：

1、保密内容（包括技术信息和经营信息）：乙方交付甲方咨询材料。

2、涉密人员范围：甲方课题组参与成员。

3、保密期限：3年。

4、泄密责任：支付乙方合同金额 10% 赔偿金。

乙方：

1、保密内容（包括技术信息和经营信息）：甲方交付乙方政府材料。

2、涉密人员范围：乙方课题组参与成员。

3、保密期限：3年。

4、泄密责任：支付甲方合同金额 10% 赔偿金。

第六条 本合同的变更必须由双方协商一致，并以书面形式确定。但有下列情形之一的，一方可以向另一方提出变更合同权利与义务的请求，另一方应当在10日内予以答复；逾期未予答复的，视为同意：

1、甲方合同执行环境发生重大变更；

2、乙方合同执行环境发生重大变更；

第七条 双方确定，按以下标准和方式对乙方提交的技术咨询工作成

果进行验收：

乙方提交技术咨询工作成果的形式：“恩平禽畜薯现代循环经济智慧示范园可行性投资报告”

2、技术咨询工作成果的验收标准：行文简练，通俗易懂，具有科学性、客观性、前瞻性和可操作性，符合政府公文体例，对推动“恩平禽畜薯现代循环经济智慧示范园具有指导意义。定性定量分析详细，脉络清晰，因果关系明确，言之有理，持之有据。

3、技术咨询工作成果的验收方法：研究报告通过专家评审或甲方认可。

4、验收的时间和地点：研究报告通过专家评审或甲方认可后的一个月内，地点由双方商定。

第八条 双方确定，甲方按照乙方符合本合同约定标准和方式完成的技术咨询工作成果做出决策并予以实施所造成的损失，按以下第1种方式处理：

- 1、乙方不承担责任。
- 2、乙方承担部分责任。
- 3、乙方承担全部责任。

第九条 双方确定：

1、在本合同有效期内，甲方利用乙方提交的技术咨询工作成果所完成的新的技术成果，归甲（甲、双）方所有。

2、在本合同有效期内，乙方利用甲方提供的技术资料和工作条件所完成的新的技术成果，归乙（乙、双）方所有。

第十条 双方确定，在本合同有效期内，甲方指定孙丙申为甲方项目联系人，乙方指定王金凤为乙方项目联系人，项目联系人承担以下责任：

- 1、项目实施过程中的沟通协调_____；
- 2、项目实施过程中相关资源的组织调配_____；
- 3、项目实施进度的监督管理_____。

一方变更项目联系人的，应当及时以书面形式通知另一方。未及时通知并影响本合同履行或造成损失的，应承担相应的责任。

第十一条 双方确定，出现下列情形，致使本合同履行成为不必要或不可能的，可以解除本合同：

- 1、发生不可抗力；

第十二条 双方因履行本合同而发生的争议，应协商、调解解决。协商、调解不成的，确定按以下第1种方式处理：

- 1、提交广州市仲裁委员会仲裁；
- 2、依法向人民法院起诉。

第十三条 与履行本合同有关的下列技术文件，经双方以书面或邮件方式确认后，为本合同的组成部分：

- 1、恩平禽畜薯现代循环经济智慧示范园可行性投资报告纲要；

第十四条 双方约定本合同其他相关事项为：

1、课题成果完成后，由甲方认定或组织专家评审，并负责成果扩印和鉴定费用。专家评审会如提出言之有理的修改意见，乙方要据此对课题成果作必要修改。

2、本成果若需出版，费用由甲方负责。

3、在本课题研究中如出现上述条款未规定的内容，由双方按友好原则协商解决。

第十五条 本合同一式肆份，具有同等法律效力，其中甲方两份，乙方两份。

第十六条 本合同经双方签字盖章后生效。



甲方：恩生瑞帮农业科技有限
公司（盖章）

委托代理人：张和申

2023 年 3 月 18 日



乙方：华南农业大学（盖章）

委托代理人：刘维红

2023 年 3 月 18 日

8.主持:关于基于机器视觉的汽车 4S 店智能管理系统研发项目的合同



华南农业大学横向委托项目

提前结题确认书

项目名称: 基于机器视觉的汽车 4S 店智能管理系统研发

委托方(甲方): 杭州智灵瞳人工智能有限公司

受托方(乙方): 华南农业大学

甲乙双方于 2023 年 11 月 28 日签署了开发 基于机器视觉的汽车 4S 店智能管理系统研发 项目的合同书。双方经过平等协商，在真实、充分地表达各自意愿的基础上，达成提前结题协议，并由双方共同签署确认。

1、甲方支付乙方原合同款项（人民币：111682.00 元，大写：拾壹万壹仟陆佰捌拾贰元整）的 35%，即 39088.00 元，大写：叁万玖千零捌拾捌元整，分两次支付：

1.1 本协议签署 3 个工作日内，甲方支付给乙方首付款：27361.00 元，大写：贰万柒仟叁佰陆拾壹元整。

1.2 2024 年 2 月 20 日，甲方支付给乙方尾款：11727.00 元，大写：壹万壹仟柒佰贰拾柒元整。

2、双方确定，乙方应在向甲方交付研究开发成果后，根据甲方的要求，为甲方指定的人员免费提供技术指导和培训，或提供与使用该研究开发成果相关的技术服务。

3、乙方应在首付款银行到账后 5 个工作日内，向甲方提供已完成的相关技术资料，包括：

- a) 前端源码；
- b) 后台数据库源码；
- c) 连接摄像头的接口调试代码；
- d) 相关说明文档一份。

交付形式：上传至甲方指定的共享网盘位置。

本确认书一式三份，甲乙双方各持一份，学校科技管理部门持有一份，以上条款双方签字盖章后生效。

甲方：杭州智灵瞳人工智能有限公司



(盖章)

法定代表人/委托代理人：



(签名)

年 月 日

乙方：华南农业大学



(盖章)

法定代表人/委托代理人：

薛红

(签名)

2023年12月28日



9.主参：高等农业院校中英联合办学项目材料

华南农业大学签订横向科技合同审批表（涉外项目）

项目名称	高等农业院校中英联合办学管理模式和管理策略研究						
项目负责人	肖莉	职称	副教授	所在单位	数学与信息学院	联系电话	13824415007
合作对方单位	英国阿爾斯特大学		联系人	肖莉	联系电话	13824415007	
合作单位性质	<input type="checkbox"/> 政府部门 <input checked="" type="checkbox"/> 大学/科研机构 <input type="checkbox"/> 企业 <input type="checkbox"/> 民间组织 <input type="checkbox"/> 其它			所在国(地区)	英国		
校内项目组成人员(包括项目负责人)							
序号	姓名	性别	所在单位	职称	年龄	签名	
1	王金凤	女	数学与信息学院	副教授	40		
2	刘财兴	男	数学与信息学院	教授	56		
3	房翊	女	材料与能源学院	正科级	50		
4							
5							
合同签订时间	2018年1月1日			计划完成时间	2021年12月30日		
合同金额(选择打“√”)	<input checked="" type="checkbox"/> 有, 31.0873 万元			<input type="checkbox"/> 无			
经费情况(选择打“√”)	<input checked="" type="checkbox"/> 拨入			<input type="checkbox"/> 拨出 <input type="checkbox"/> 无经费			
合同类别(选择打“√”):	<input type="checkbox"/> 技术开发 <input type="checkbox"/> 技术转让 <input type="checkbox"/> 技术服务 <input type="checkbox"/> 技术咨询 <input type="checkbox"/> 共建平台 <input type="checkbox"/> 专利(新品种)许可实施 <input type="checkbox"/> 框架协议 <input type="checkbox"/> 短期交流访问 <input checked="" type="checkbox"/> 其它 技术合作						
项目负责人承诺: 审批表内容及签名真实, 合同涉及的内容真实、合法。 如果本合同正式签订, 本人将严格按照合同约定和国家、学校有关管理规定, 按质、按量、按期执行项目, 做好项目结题和档案归档工作, 合理使用经费; 愿意承担因主观原因造成的损失; 在合同执行过程中维护国家尊严, 维护学校的合法权益, 遵守保密规定, 尊重他人知识产权, 遵守学术规范。 本人已与合作方进行充分的沟通和交流, 了解其开展合作的真实目的, 本合同不涉及威胁国家安全或意识形态安全等内容。 项目负责人: _____ 年 月 日							
所在单位分管外事工作领导审核意见: 主管领导签名: _____ 年 月 日		所在单位分管科技工作领导审核意见: 按照上级及学校科技合同有关规定, 本人已仔细阅读合同全文并理解各条款的含义, 合同内容经初审符合《中华人民共和国合同法》等法律法规和学校有关规定, 项目负责人及项目组成员具有相关的研究基础, 有足够的技术力量完成本合同规定由我方履行的条款, 本单位内无涉及该项目的知识产权纠纷。 同意签订本合同, 若本合同正式签订, 本单位将配合做好项目执行和管理工作。 主管领导签名: _____ 单位盖章 _____ 年 月 日					
国际交流处审批意见: (单位盖章) 审核负责人(签名): _____ 年 月 日							

合同初次提交时间		合同修改返回时间	
科技处业务科室审批意见:			
审核负责人: _____ 年 月 日			
科技处分管领导审批意见:			
_____ 年 月 日			
科技处处长审批意见:			
_____ 年 月 日			
校领导审批意见:			
_____ 年 月 日			

备注:

1. 本审批表请双面打印。
2. 正式签订的合同请于15个工作日内将1份合同原件交科技处成果与知识产权科存档。
3. 涉外项目是指我校与合作方为国际（境）外单位、或其驻中方分支机构委签署的横向科技项目。

10.主参：关于面向多组学数据的高阶融合网络建模项目的立项通知

广东省基础与应用基础研究基金项目任务书

受理编号：c23140500000850

项目编号：2023A1515012885

文件编号：粤基金字（2023）2号

广东省基础与应用基础研究基金项目 任务书

项目名称：面向多组学数据的高阶融合网络建模

项目类别：广东省自然科学基金面上项目

项目起止时间：2023-01-01 至 2025-12-31

管理单位（甲方）：广东省基础与应用基础研究基金委员会

依托单位（乙方）：华南农业大学

通讯地址：广东省广州市天河区五山路483号

邮政编码：510642

单位电话：020-85283435

项目负责人：王海燕

联系电话：15013255687



（广东科技微信公众号）



（查看任务书信息）



（受理纸质材料二维码）

广东省基础与应用基础研究
基金委员会
二〇二〇年制

填写说明

一、项目任务书内容原则上要求与申报书相关内容保持一致，不得无故修改。

二、项目承担单位通过广东省科技业务管理阳光政务平台下载项目任务书，按要求完成签名盖章后扫描上传到广东省科技业务管理阳光政务平台。

三、签名盖章说明。请分别在单位工作分工及经费分配情况页、人员信息页、签约各方页等地方按要求签字或盖章，签章不合规或错漏将不予受理。其中，人员信息页要求所有参与人员本人亲笔签名，代签或印章无效，漏签将不予受理。

四、本任务书自签字并加盖公章之日起生效，各方均应负本任务书的法律责任，不应受机构、人事变动影响。

五、根据《广东省科学技术厅广东省财政厅关于深入推进省基础与应用基础研究基金项目经费使用“负面清单+包干制”改革试点工作的通知》（粤科规范字[2022]2号），2022年度及以后立项资助的全部省基金项目（包括省自然科学基金、省市联合基金、省企联合基金项目等）均适用“负面清单+包干制”，项目提交申请书和任务书时无需编制费用明细科目预算。

一、主要研究内容和要达到的目标

主要研究内容如下：

（一）基于张量二部图的高阶融合低维表示学习模型构建

多组学数据具有多源差异性、信息互补和语义一致的优势，提取组学间的高阶相关统计，加深对数据整体的理解是重要的科学挑战。本项目拟在前期的张量学习方法上（工作基础2），利用张量理论的优势设计有效的高阶融合学习算法。在这一框架下，设计多组学张量二部图学习，深入研究多组学数据间的互补信息和高阶相关性。同时迭代利用所学二部图张量进行Tucker分解，量化各组学的重要性并从各组学中抽取适量信息识别数据的统一低维度表示，提升多组学数据表示学习的性能。

（二）基于流形度量的高阶超图网络融合学习模型构建

本项目拟在前期Grassmann流形度量的学习模型框架下（工作基础1），开展后续的研究工作，基于充分调研流形学习的数学理论和超图网络建模学习的基础上，在对组学数据进行高阶超图网络构建的同时进行共享子空间识别。模型使得超图刻画组学内部具有高阶局部亲和力的样本关系网络；更使得多组学子空间在Grassmann流形上充分对齐，有效避开欧式度量复杂多组学数据产生的不准确性，实现新度量空间下的多组学整合分析方法研究。

（三）基于高阶相似度的组学网络融合学习模型构建

多组学数据在有噪声的情况下，难以刻画准确的样本关系。本项目拟设计基于高阶相似度的组学高阶融合网络模型，用于挖掘多组学数据中可靠的样本关联关系，减小高噪声的生物数据分析挑战，以期实现关键生物标志物挖掘及精准疾病诊断。通过建立组学高阶张量相似度及对应高阶相似度表示，寻找隐藏在组学样本中的高阶关联性，有效缓解由于数据噪声带来的影响。

具体研究目标如下：

- 1) 建立基于张量二部图的高阶融合低维表示学习模型，利用嵌入在单组学二部图中的空间结构和互补信息，通过张量理论充分挖掘组学间的高阶相关关系，实现疾病关联基因模块的准确挖掘。
- 2) 建立基于流形度量的高阶超图网络融合学习模型，利用高阶超图网络学习具有高阶局部亲和力的样本关系网络，同时使得多组学子空间在流形度量空间充分对齐，学习有效的共享子空间学习，实现流形度量空间下的多组学整合分析研究，提高共表达模块识别的鲁棒性，减少疾病诊断的假阳性。
- 3) 建立基于高阶相似度的组学网络融合学习模型，利用四阶张量建模组学样本点对点关系，为点-点关系提供补充，充分挖掘组学样本相似度信息，运用高阶相似度理论挖掘建模学习多组学共享基因模块，实现去噪声和整合等任务，提升临床精准医疗的准确率。

二、项目预期获得的研究成果及形式

论文及专著情况	国家统计局刊物以上刊物 发表论文(篇)		3		科技报告(篇)			
	其中被SCI/EI/ISTP收录 论文数(篇)		3		培养人才(人)			
	专著(册)				引进人才(人)			
专利情况(项)	发明专利		实用新型专利		外观设计专利		国外专利	
	申请	授权	申请	授权	申请	授权	申请	授权
	2							

2023A1515012885

三、项目进度和阶段目标

(一) 项目起止时间： 2023-01-01 至 2025-12-31		
(二) 项目实施进度及阶段主要目标:		
开始日期	结束日期	主要工作内容
2023-01-01	2023-12-31	<p>(1) 研究计划：依托于张量计算理论，通过张量二部图学习、张量Tucker分解的有机结合，建立基于张量二部图的高阶融合低维表示学习模型。</p> <p>(2) 阶段目标： 建立基于张量二部图的高阶融合低维表示学习模型； 深入了解国内外学术界最新发展，参加重要国内外学术会议1-2次； 发表高水平论文1篇。</p>
2024-01-01	2024-12-31	<p>(1) 研究计划：在充分文献方法调研的基础上，结合超图及流形学习度量理论，完成基于流形度量的高阶超图网络融合学习模型构建，给出高效的数值求解方案。</p> <p>(2) 阶段目标： 建立基于流形度量的高阶超图网络融合学习模型； 深入了解国内外学术界最新发展，参加重要国内外学术会议1-2次； 发表高水平论文1篇，申请国家发明专利1项。</p>
2025-01-01	2025-12-31	<p>(1) 研究计划：结合张量谱分析理论，建立基于高阶相似度的组学网络融合学习模型，给出高效的数值求解方案。</p> <p>(2) 阶段目标： 建立基于高阶相似度的组学网络融合学习模型、参加重要国内外学术会议1-2次； 发表高水平论文1篇，申请国家发明专利1项。</p>

四、项目总经费及省基金委经费预算

1. 省基金委经费下达总额：（大写）壹拾万圆整；（小写）10万元；					
2. 省基金委经费年度下达计划：					
年度	2023 年	年	年	年	年
经费(万元)	10.00				

2023A1515012885

五、人员信息

项目负责人								
姓名	证件号码	年龄	性别	职称	学历	在项目中承担的任务	所在单位	签名
王海燕	340621198704047560	36	女	讲师	博士研究生	项目负责人	华南农业大学	

项目组主要成员								
姓名	证件号码	年龄	性别	职称	学历	在项目中承担的任务	所在单位	签名
黄栋	441622198711022098	36	男	副教授	博士研究生	算法检验	华南农业大学	
彭超达	445221199009201918	33	男	副教授	博士研究生	算法检验	华南农业大学	
王金凤	130105197810080929	45	女	副教授	博士研究生	算法检验	华南农业大学	
张广煜	440106199105151814	32	男	讲师	博士研究生	算法检验	华南农业大学	
崔金荣	410927198412094042	39	女	讲师	博士研究生	算法检验	华南农业大学	

六、工作分工及财政经费分配

承担/参与单位名称 (盖章)	工作分工	省级财政科技资金分配 (万元)
华南农业大学	项目负责人负责跟进年度执行目标的达成，并完成主要的科研工作；为保障项目顺利进行，合作老师负责辅助检验算法及实验数据；同时计划引进一名硕士生，负责数据清洗及辅助算法调试。	10.00
	合计	10.00

2023A1515012885

七、任务书条款

第一条 甲方与乙方根据《中华人民共和国民法典》及国家有关法规和规定，按照《广东省科学技术厅关于广东省基础与应用基础研究基金（省自然科学基金、联合基金等）项目管理的实施细则（试行）》《广东省省级科技计划项目验收结题工作规程（试行）》等规定，为顺利完成（2023）年面向多组学数据的高阶融合网络建模专项项目（文件编号：粤基金字〔2023〕2号）经协商一致，特订立本任务书，作为甲乙双方在项目实施管理过程中共同遵守的依据。

第二条 甲方的权利义务：

1. 按任务书规定进行经费核拨的有关工作协调。
2. 根据甲方需要，在不影响乙方工作的前提下，定期或不定期对乙方项目的实施情况和经费使用情况进行检查或抽查。
3. 根据《广东省科研诚信管理办法（试行）》等规定对乙方进行科技计划信用管理。

第三条 乙方的权利义务：

1. 确保落实自筹经费及有关保障条件。
2. 按任务书规定，对甲方核拨的经费实行专款专用，单独列账，并随时配合甲方进行监督检查。
3. 经费使用按照广东省级财政科研项目经费使用等有关规定进行管理。
4. 项目依托单位应制定经费使用“负面清单+包干制”内部管理制度并报甲方备案。
5. 使用财政资金采购设备、原材料等，按照《广东省实施〈中华人民共和国招标投标法〉办法》有关规定，符合招标条件的须进行招标。
6. 项目任务书任务完成后，或任务书规定的任务、指标及经费投入等提前完成的，乙方可提出验收结题申请，并按甲方要求做好项目验收结题工作。
7. 若项目发生需要终止结题的情况，乙方须提出终止结题申请，并按甲方要求做好项目终止结题工作。
8. 在每年规定时间内向甲方如实提交上年度工作情况报告，报告内容包含上年度项目进展情况、经费决算和取得的成果等。
9. 按照国家和省有关规定，提交科技报告及其他材料。
10. 利用甲方的经费获得的研究成果，项目负责人和参与者应当注明获得“广东省基础与应用基础研究基金（英文：Guangdong Basic and Applied Basic Research Foundation）（项目编号）”资助或作有关说明。
11. 乙方要恪守科学道德准则，遵守科研活动规范，践行科研诚信要求，不得抄袭、剽窃他人科研成果或者伪造、篡改研究数据、研究结论；不得购买、代写、代投论文，虚构同行评议专家及评议意见；不得违反论文署名规范，擅自标注或虚假标注获得科技计划（专项、基金等）等资助；不得弄虚作假，骗取科技计划（专项、基金等）项目、科研经费以及奖励、荣誉等；不得有其他违背科研诚信要求的行为。
12. 确保本项目开展的研究工作符合我国科技伦理管理相关规定。

第四条 在履行本任务书的过程中，如出现广东省相关政策法规重大改变等不可抗力情况，甲方有权对所核拨经费的数量和时间进行相应调整。

第五条 在履行本任务书的过程中，当事人一方发现可能导致项目整体或部分失败的情形时，应及时通知另一方，并采取适当措施减少损失，没有及时通知并采取适当措施，致使损失扩大的，应当就扩大的损失承担责任。

第六条 本项目技术成果的归属、转让和实施技术成果所产生的经济利益的分享，除双方另有约定外，按国家和广东省有关法规执行。

第七条 根据项目具体情况，经双方另行协商订立的附加条款，作为本任务书正式内容的一部分，与本任务书具有同等效力。

第八条 本任务书一式三份，各份具有同等效力。甲、乙方及项目负责人各执一份，三方签字、盖章后即生效，有效期至项目结题后一年内。各方均应负责任务书的法律责任，不应受机构、人事变动的影响。

第九条 乙方必须接受甲方聘请的本项目任务书监理单位的监督和管理。监理单位按照甲方赋予的权利对本项目任务书的履行进行审核、进度调查，对项目任务书变更、经费使用情况进行监督管理及组织项目验收。

说明：1. 本任务书中，凡是当事人约定无需填写的内容，应在空白处划（/）。

2. 委托代理人签订本任务书的，应出具合法、有效的委托书。

八、本任务书签约各方

管理单位（甲方）：	广东省基础与应用基础研究基金会（盖章）
法定代表人（或法人代理）：	  （签章）
2023 年 02 月 14 日	
依托单位（乙方）：	华南农业大学（盖章）
法定代表人（或法人代理）：	刘雅红（签章）
联系人（项目主管）姓名：	倪慧群（签章）
Email: kjcgxk@scau.edu.cn	
电话: 020-85283435 / 15920301530	
开户单位名称：	华南农业大学
开户银行名称：	广东广州工行五山支行
开户银行帐号：	3602002609000310520
年 月 日	
联系人（项目负责人）姓名：	王海燕（签名）
Email: cshywang@scau.edu.cn	
电话: 15013255687	
年 月 日	

11.主参：关于渔业大数据系统标准化体系研究项目的立项通知

九、承担/参研单位合作协议

项目合作协议书

甲方：南方海洋科学与工程广东省实验室（湛江）

乙方：华南农业大学

甲乙双方本着互相协作的精神，就 2023.07—2025.06 年度南方海洋科学与工程广东省实验室（湛江）（简称湛江湾实验室）项目《智慧渔业体系构建与关键技术研究》，达成如下合作协议：

一、研究任务

甲方为项目承担单位，乙方为项目参研单位之一。乙方须按以下要求完成研究任务：

1. 主要研究内容：

乙方负责的主要研究内容为：

课题3、渔业大数据系统标准化体系研究

- (1) 智慧渔业数据基础类标准
- (2) 智慧渔业数据开发标准
- (3) 智慧渔业数据安全技术标准
- (4) 智慧渔业数据管理标准
- (5) 智慧渔业大数据应用标准

2. 预期成果

形成渔业大数据系统标准化体系的标准草案 4 份，核心期刊或以上论文不少于 1 篇，课题研究报告 1 份。

3. 考核指标

形成渔业大数据系统标准化体系的标准草案 4 份，并通过专家评审；核心期刊或以上论文不少于 1 篇，提供 DOI 号；课题研究报告 1 份。

二、实验室经费分配

乙方获得资助金额为 30 万（占实验室资助总经费的 6%），各单位获取资助



金额按照实际获得实验室资助金额为准，按照此比例进行分配。

三、该项目产生的知识产权按照参与各方在人、财、物等资源投入数量、投入时间等因素进行单位排序或划分权益分配比例，当共有知识产权的实施有赖于合作方或他人的在先知识产权的实施为前提时，不得侵犯合作方或他人的在先合法权利。

四、双方在项目执行期间产生的歧义与争议，本着友好协商的精神解决，若协商不能达成协议，由有关上级主管部门出面解释和调解。

五、双方根据《广东省人民政府关于印发〈广东省省级财政专项资金管理办法（试行）〉的通知》（粤府〔2018〕120号）、《广东省财政厅广东省审计厅关于省级财政科研项目资金的管理监督办法》（粤财规〔2019〕5号）等有关规定，以及有关法律、政策和管理要求，依据项目立项通知，签署本合作协议。

甲方（盖章）：



法定代表人/委托代理人：

张进伟

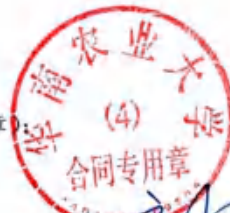
项目负责人：

杨仁友

所在单位：南方海洋科学与工程广东省实验室（湛江）

日期： 年 月 日

乙方（盖章）：



法定代表人/委托代理人：

刘雅红

项目负责人：

黄小

所在单位：华南农业大学

日期： 年 月 日

12.主参：关于基于移动式平台的水下场景立体重构与目标识别项目的立项通知

项目类别： 重点 一般
合同编号： 粤自然资合[2024]18号

**广东省海洋经济发展（海洋六大产业）
专项资金项目合同书**

项目名称：基于移动式平台的水下场景立体重构与目标识别
委托方（甲方）：广东省自然资源厅
受托方（乙方）：南方海洋科学与工程广东省实验室（湛江）

签订时间：2024年03月01日
签订地点：广州

广东省自然资源厅
2024年版

项目类别： 重点 一般
合同编号：

广东省海洋经济发展（海洋六大产业） 专项资金项目合同书

项目名称： 基于移动式平台的水下场景立体重构与目标识别

委托方（甲方）： 广东省自然资源厅

受托方（乙方）： 南方海洋科学与工程广东省实验室（湛江）



签订时间： 2024年03月01日

签订地点： 广州

广东省自然资源厅

2024年版

委托方（甲方）：广东省自然资源厅

办公场所：广州市天河区体育东路 160 号

法定代表人：胡建斌

项目负责人：黄国锐

通信地址：广州市天河区天河北路 369 号

邮政编码：510000

联系电话：83624507

电子邮箱：szzzyt_hjc@gd.gov.cn

受托方（乙方）：南方海洋科学与工程广东省实验室（湛江）

办公场所：广东省湛江市坡头区湛江湾实验室龙王湾园区

法定代表人：颜开

项目负责人：杨仁友

通信地址：广东省湛江市坡头区湛江湾实验室龙王湾园区

邮政编码：524000

联系电话：16625138289

电子邮箱：youngrenyou@zjblab.com

甲乙双方根据《中华人民共和国民法典》和《广东省省级财政专项资金管理办法（修订）》《广东省自然资源厅关于印发 2024 年省级促进经济高质量发展（海洋经济发展）海洋六大产业专项项目申报指南的通知》《海洋经济发展（海洋六大产业）专项资金管理实施细则

(试行)》《广东省自然资源厅关于规范海洋经济发展(海洋六大产业)专项资金项目验收工作的通知》《广东省自然资源厅办公室关于开展海洋经济发展(海洋六大产业)专项资金项目数据汇交工作的通知》等有关规定,在平等协商、真实表达意愿的基础上,订立了本合同,并由双方约定共同遵守。

第一章 委托内容及要求

第一条 本合同委托的项目名称为:“基于移动式平台的水下场景立体重构与目标识别”。

第二条 甲乙双方围绕本项目实施制订总体考核指标,包括成果指标及绩效指标,作为检验项目成效的重要依据。

第三条 乙方应当按照本项目设定的项目实施方案及考核指标要求组织项目实施工作。

第四条 乙方应当在合同到期前90日内向甲方交付成果材料申请验收。因不可抗力因素导致无法按期提交材料的,乙方应当于合同期满提前30日向甲方递交延期验收书面说明并加盖公章。

第五条 甲方对乙方提交的项目成果进行验收,项目验收按《海洋经济发展(海洋六大产业)专项资金管理实施细则(试行)》《广东省自然资源厅关于规范海洋经济发展(海洋六大产业)专项资金项目验收工作的通知》《广东省自然资源厅办公室关于开展海洋经济发展(海洋六大产业)专项资金项目数据汇交工作的通知》等规定执行。

第六条 项目实施及完成后取得的技术成果和知识产权,按照相关法律法规执行,甲方在保护专利前提下有权在满足社会公众利益范围内使用项目成果。乙方应当保证其研究成果不侵犯任何第三方的合法权益,否则应当承担相应的损害赔偿责任。

第七条 项目实施过程中,乙方向甲方申请提供的数据、源代码、技术文档等资料,只能用于该项目研究,未经提供方书面同意,乙方

不得公开发表或用于其他方面。否则，由于乙方过错导致的上述资料泄密的，由乙方承担一切责任，按照甲方实际损失程度进行经济赔偿。项目完成后，甲、乙双方均有责任对本项目的技术保密承担责任。

第二章 合同期限、价款及付款方式

第八条 本合同自签订之日起生效，为期2年。

第九条 本合同项目专项资助经费为人民币¥300.00万元（大写：叁佰万元），牵头单位和参与单位自筹经费¥500万元（大写：伍佰万元）。

第十条 甲乙双方约定，在本合同签订之日起15日内，甲方拨付专项资金的50%（即人民币¥150.00万元）给乙方。

第十一条 乙方于2024年5月31日前向甲方提交第一阶段项目实施进展报告及相应佐证材料，甲方收到报告后在15日内组织评审，评审通过后拨付专项资金的35%（即人民币¥105.00万元）。评审不通过的，乙方应当在收到甲方通知之日起15日内重新提交项目实施进展报告，甲方将重新组织评审，拨款时间相应顺延。重新评审不通过的，甲方有权停止或核减拨付该阶段应拨付的资金。

第十二条 乙方于2024年7月31日前向甲方提交第二阶段项目实施进展报告及相应佐证材料，甲方收到报告后在15日内组织评审，评审通过后拨付专项资金的15%（即人民币¥45.00万元）。评审不通过的，乙方应当在收到甲方通知之日起15日内重新提交项目实施进展报告，甲方将重新组织评审，拨款时间相应顺延。重新评审不通过的，甲方有权停止或核减拨付该阶段应拨付的资金。

第十三条 甲方按第十条、第十一条、第十二条约定向乙方拨付经费。甲方在前款规定的付款时间内向相关财政部门提出办理财政支付申请即视为已经按期付款。甲方将经费拨付至乙方下述指定银行账户：

2025年上半年开展中期评审，中期评审不通过的，乙方将专项资金的50%按照财政专项资金拨付渠道退回，甲方或丙方视项目进展情况再予以拨付。

开户单位名称：南方海洋科学与工程广东省实验室(湛江)

税务代码：103591060807

开户银行：中国农业银行湛江霞山支行

账号：44608001040018958

乙方需要变更指定银行账户的应当提前 10 个工作日书面告知甲方，否则由此引起的后果和纠纷由乙方自行承担。

第三章 双方的权利与义务

第十四条 甲方的权利与义务：

(一) 对项目实施和经费使用按照本专项资金管理办法及其实施细则等相关规定进行监督检查与绩效管理；

(二) 建立项目监管制度，开展项目实施进度（实施方案执行情况）评估，检查经费使用情况，发现违约违规使用经费的行为及时制止并责令整改；

(三) 根据本合同约定的金额及方式向乙方支付项目经费；

(四) 对项目实施提供必要的政策指导和帮助，及时协调处理和解决项目实施中的重大问题；

(五) 项目结题时，组织开展项目验收；

(六) 法律规定的其他权利义务。

第十五条 乙方的权利与义务：

(一) 按照本合同规定组织项目实施，合理规范使用财政资金；

(二) 按甲方要求定期汇报项目实施及研究进展情况，加强项目管理，按时提交成果，保证成果信息来源的真实性、合法性、完整性；

(三) 严格按照项目经费预算，落实项目自筹资金，按相关规定加强财务管理；

(四) 加强对本项目参与单位的监督管理，对项目组织实施、验收和监督管理负全面责任；

5

- 5 -

(五)接受政府相关职能部门或者甲方委托的第三方机构监督管理；

(六)对本合同所有相关信息的真实性、完整性负责，并承担由此产生的一切法律责任；

(七)法律规定的其他权利义务。

第四章 项目资金管理及相关要求

第十六条 项目资金由财政专项补助资金与项目牵头单位和参与单位自筹资金组成，实施独立核算、专账核算及管理。本财政补助资金不得用于与其他各级财政已支持项目的重复内容。

第十七条 项目投资与经费使用计划一经确定，不得随意更改及调整。项目实施过程中，凡涉及到政府采购、招投标等事项的，要严格按照有关规定执行。

第十八条 自筹资金由项目牵头单位和参与单位按项目资金使用计划同步（或提前）筹措到位。

第十九条 项目通过验收后，乙方须在甲方出具通过验收结论之日起30日内，将经审计后的财政专项资金结余部分按原财政专项资金拨付渠道退回。

第五章 合同变更与终止

第二十条 对本合同任何条款的修改或补充，须由双方协商一致并签订书面协议，协议经双方签字盖章后生效，与本合同具备同等法律效力。

第二十一条 具有下列情形之一的，当事人有权解除本合同：

(一)未经甲方书面同意，乙方将其承担的主要工作以任何方式转交第三方完成的，甲方可以解除合同。

(二)甲方无正当理由故意不履行协助义务致使乙方研究工作不

能正常开展，乙方应当催告甲方在合理期限内履行义务；经乙方催告仍不履行的，乙方可以解除合同。

（三）乙方未按照本合同约定的时间、要求交付成果的，甲方应当催告乙方在合理期限内履行义务；经甲方催告仍不履行的，甲方可以解除合同。

（四）乙方将本合同项下的财政补助资金用于与本项目无关的其他方面的，甲方可以解除合同。

（五）甲乙双方协商一致，可以解除合同。

第六章 违约责任及争议的解决

第二十二条 任何一方因无法克服的技术困难、自然灾害等不可抗力不能履行合同义务时，可以免除违约责任，但应当及时通知另一方，并在合理期间内出具因不可抗力导致合同不能履行的证明。在出现不可抗力的情况下，双方均采取适当措施减轻损失。任何一方因未采取措施或者采取措施不当导致损失扩大的，应当对扩大的损失承担责任，但在一方迟延履行期间遭遇不可抗力的，违约方不得免责。

第二十三条 甲乙双方违反法律规定或者本合同约定义务的，应当承担违约责任：

（一）因乙方的原因导致研究工作未能按期完成，或者项目成果未能达到合同约定要求的，或者未经甲方书面同意，乙方将其承担的主要工作以任何方式转交第三方完成的，甲方有权解除合同，要求乙方返还已经支付的财政专项经费，并赔偿因此给甲方造成的损失。

（二）乙方将本合同项下的财政补助资金用于与本项目无关的其他方面的，甲方有权解除合同，要求乙方返还已经支付的财政专项经费，并赔偿因此给甲方造成的损失。

（三）甲方违反本合同约定，逾期未支付财政专项经费，经催告后仍未按时支付的，乙方有权解除合同，并有向甲方追回已开展工作

第三十条 本合同附件是合同内容的组成部分，具备合同效力。

- 附件：1. 项目基本情况
2. 项目实施方案及考核指标
3. 项目预算及说明
4. 项目预算审核报告
5. 项目实施承诺书
6. 项目合作协议

甲方：广东省自然资源厅



(公章)

负责人(或委托式)(签字):



日期：2024年3月26日

乙方：南方海洋科学与工程广东省实验室(湛江)

(公章)

单位法人(签字):



项目负责人(签字):

日期：2024年3月13日

附件 1

项目基本情况

项目名称	基于移动式平台的水下场景立体重构与目标识别					
项目类别	<input type="checkbox"/> 重点		<input checked="" type="checkbox"/> 一般			
所属专题	海洋电子信息专题					
项目牵头单位	单位名称	南方海洋科学与工程广东省实验室（湛江）				
	通讯地址	广东省湛江市坡头区湛江湾实验室龙王湾园区				
	项目负责人	杨仁友	联系电话	16625138289	传真	
			邮箱	youngrenyou@z.jblab.com		
	项目联系人	朱宇珍	联系电话	13802828018		
			邮箱	kyb@z.jblab.com		
	科研管理部门负责人	朱宇珍	联系电话	13802828010	传真	0759-2086298
邮箱			zhuyuzhen@z.jblab.com			
参与单位	单位名称	华南农业大学				
	联系人	王金凤	联系电话	15815860686		
	单位名称	广东紫贝智慧农业有限公司				
	联系人	蔡逢旺	联系电话	13976040894		
	单位名称					
	联系人		联系电话			
	单位名称					
项目组	总人数	正高级	副高级	中级	初级	其他
	35	3	7	8	5	12
经费预算	800 万元	专项资金	300 万元			
		自筹资金	500 万元			
		其他	万元			

姓名	性别	年龄	职称	学历	项目职责	所在单位	签名
杨仁友	男	44	正高级	博士研究生	项目负责人	南方海洋科学与工程广东省实验室(湛江)	杨仁友
欧先伟	男	59	副高级	本科	项目骨干	南方海洋科学与工程广东省实验室(湛江)	欧先伟
黄琼	男	42	正高级	博士研究生	项目骨干	华南农业大学	黄琼
梁云	男	43	正高级/院士	博士研究生	项目骨干	华南农业大学	梁云
蔡逢旺	男	30	其他	专科	项目骨干	广东紫贝智慧农业有限公司	蔡逢旺
蔡逢运	男	48	其他	本科	项目骨干	广东紫贝智慧农业有限公司	蔡逢运
杨靛	女	38	初级	硕士研究生	项目骨干	南方海洋科学与工程广东省实验室(湛江)	杨靛
秦浩	男	28	初级	硕士研究生	项目骨干	南方海洋科学与工程广东省实验室(湛江)	秦浩
郑凯健	男	28	初级	硕士研究生	项目骨干	南方海洋科学与工程广东省实验室(湛江)	郑凯健
朱宇珍	女	41	中级	硕士研究生	项目骨干	南方海洋科学与工程广东省实验室(湛江)	朱宇珍
何琛	男	27	初级	硕士研究生	项目骨干	南方海洋科学与工程广东省实验室(湛江)	何琛
李日富	男	32	中级	硕士研究生	项目骨干	南方海洋科学与工程广东省实验室(湛江)	李日富

						实验室 (湛江)	
李俞锋	男	36	中级	硕士研究生	项目骨干	南方海洋科学与工程广东省实验室(湛江)	李俞锋
崔志坚	男	27	初级	硕士研究生	项目骨干	南方海洋科学与工程广东省实验室(湛江)	崔志坚
黄沛杰	男	44	副高级	博士研究生	项目骨干	华南农业大学	黄沛杰
王金凤	女	46	副高级	博士研究生	项目骨干	华南农业大学	王金凤
尹令	女	47	副高级	博士研究生	项目骨干	华南农业大学	尹令
张素敏	男	47	中级	博士研究生	项目骨干	华南农业大学	张素敏
蔡爵	男	30	其他	专科	项目骨干	广东紫贝智慧农业有限公司	蔡爵
龙朋	女	45	其他	专科	项目骨干	广东紫贝智慧农业有限公司	龙朋
黄兹辉	男	27	其他	本科	项目骨干	广东紫贝智慧农业有限公司	黄兹辉
李杰	女	61	副高级	硕士研究生	其他研究人员	南方海洋科学与工程广东省实验室(湛江)	李杰
陈晓豪	男	35	中级	本科	其他研究人员	南方海洋科学与工程广东省实验室(湛江)	陈晓豪
钟馨聪	男	26	中级	本科	其他研究人员	南方海洋科学与工程广东省实验室(湛江)	钟馨聪
梁浩权	男	26	其他	硕士研究生	其他研究人员	南方海洋科学与工程广东省实验室(湛江)	梁浩权

三、论文、著作等

1.检索证明

SCAULIB202518623

检索证明

根据委托人提供的论文材料，委托人华南农业大学数学与信息学院 王金凤 5篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大分
1	FF-GLAM-net: a fusion framework based on GLAM with channel shuffle for speech emotion recognition	INTERNATIONAL JOURNAL OF MACHINE LEARNING AND CYBERNETICS 出版年: 2024 出版日期: MAR 卷期: 15 3 页码: 287-340 文献类型: Article	第一作者	B类	华南农业大学 数学与信息学院	SCI	IF2-year=2.7 IF5-year=3.1 (2024)	计算机科学 3区 Top期刊: 否 (2025)
2	A calibrated SVM based on weighted smooth $GL(\frac{1}{2})$ for Alzheimer's disease prediction	COMPUTERS IN BIOLOGY AND MEDICINE 出版年: 2023 出版日期: MAY 卷期: 158 页码: - 文献号: 106752 文献类型: Article	第一作者	A类	华南农业大学 数学与信息学院	SCI	IF2-year=7.0 IF5-year=6.7 (2023)	医学 2区 Top期刊: 是 (2023)
3	Driver Fatigue Detection Using Improved Deep Learning and Personalized Framework	INTERNATIONAL JOURNAL ON ARTIFICIAL INTELLIGENCE TOOLS 出版年: 2022	第一作者	B类	华南农业大学 数学与信息学院	SCI	IF2-year=1.1 IF5-year=1.2 (2022)	计算机科学 4区 Top期刊: 否 (2022)

第 1 页/共 2 页

		出版日期: MAR 卷期: 31 02 页码: - 文献号: 2250024 文献类型: Article						
4	Parallel computing of fuzzy integrals: Performance and test	JOURNAL OF INTELLIGENT & FUZZY SYSTEMS 出版年: 2021 卷期: 41 2 页码: 3137-3159 文献类型: Article	第一作者		华南农业大学 数学与信息学院	SCI	IF2-year=1.737 IF5-year=1.664 (2021)	无
5	Fuzzy measure with regularization for gene selection and cancer prediction	INTERNATIONAL JOURNAL OF MACHINE LEARNING AND CYBERNETICS 出版年: 2021 出版日期: AUG 卷期: 12 8 页码: 2389-2405 文献类型: Article	第一作者	B类	华南农业大学 数学与信息学院	SCI	IF2-year=4.377 IF5-year=3.764 (2021)	计算机科学 3区 Top期刊: 否 (2021)

说明: 论文等级和中科院大分分区按《中国学术期刊全文数据库评价方案(试行)》执行。

报告免责声明: 如未盖章, 报告无效

检索员: 田成
华南农业大学图书馆
2025-07-09

第 2 页/共 2 页

检索证明

根据委托人提供的论文材料，委托人华南农业大学数学与信息学院 王金凤 4 篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者工作单位	收录情况	影响因子	中科院大类专业
1	A simple self-supervised learning framework with graph-based data augmentation in diagnosis of Alzheimer's disease	BIOMEDICAL SIGNAL PROCESSING AND CONTROL 出版年: 2024 出版日期: OCT 卷期: 96 页码: 文献号: 10632 文献类型: Article	通讯作者	A类	华南农业大学 数学与信息学院	SCI	IF2-year=4.9 IF5-year=5.0 (2024)	医学 2区 Top期刊: 否 (2025)
2	Fuzzy measure with regularization for gene selection and cancer prediction	INTERNATIONAL JOURNAL OF MACHINE LEARNING AND CYBERNETICS 出版年: 2021 出版日期: AUG 卷期: 12 8 页码: 2389-2405 文献类型: Article	通讯作者	B类	华南农业大学 数学与信息学院	SCI	IF2-year=4.377 IF5-year=3.764 (2025)	计算机科学 1区 Top期刊: 否 (2025)
3	EduLCL: Local-global contrastive learning model for education recommendation	KNOWLEDGE-BASED SYSTEMS 出版年: 2024 出版日期: FEB 28	通讯作者	T2类	华南农业大学 数学与信息学院	SCI	IF2-year=7.6 IF5-year=7.6 (2024)	计算机科学 1区 Top期刊: 是

第 1 页/共 2 页

		卷期: 286 页码: - 文献号: 111357 文献类型: Article						(2025)
4	BFCNet: A Dense Flock of Chickens Counting Network Based on Density Map Regression	ANIMALS 出版年: 2023 出版日期: DEC 卷期: 13 23 页码: - 文献号: 3729 文献类型: Article	通讯作者	A类	华南农业大学 数学与信息学院	SCI	IF2-year=2.7 IF5-year=3.0 (2023)	农林科学 2区 Top期刊: 否 (2023)

说明: 论文等级和中科院大类专业按《华南农业大学学位论文评阅方案(试行)》划分。

报告免盖章: 如未盖章, 报告无效

检索员: 田成
华南农业大学图书馆
信息检索部
2025-07-09

第 2 页/共 2 页

检索证明

根据委托人提供的论文材料，委托人华南农业大学数学与信息学院 王金凤 4 篇论文收录情况如下表。

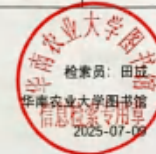
序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者工作单位	收录情况	影响因子	中科院大类分区
1	融入平滑组稀疏化的脑部 MRI 图像分类	中国图象图形学报 出版年: 2022 卷期: 27 3 页码: 885-897 文献号: 文献类型:	通讯作者	B类	华南农业大学 数学与信息学院	北大核心	无	无
2	基于双类框架与局部感受野的实时人脸疲劳检测	计算机应用研究 出版年: 2020 卷期: 37 23 页码: 3795-3798 文献号: 文献类型:	通讯作者	B类	华南农业大学 数学与信息学院	北大核心	无	无
3	基于蚁群算法求解 Cauchy 模糊积分模型	山东大学学报 (工学版) 出版年: 2018 卷期: 48 6 页码: 81-87 文献号: 文献类型:	通讯作者	C类	华南农业大学 数学与信息学院	北大核心	无	无

第 1 页/共 2 页

4	并行计算和稀疏存储在模糊积分上的应用	计算机应用研究 出版年: 2018 卷期: 35 1 页码: 166-171 文献号: 文献类型:	通讯作者	C类	华南农业大学 数学与信息学院	北大核心	无	无
---	--------------------	---	------	----	-------------------	------	---	---

说明: 论文等级和中科院大类分区按《华南农业大学学位论文评价方案(试行)》划分。

报告免责声明: 如未盖章, 报告无效



第 2 页/共 2 页

2.以第一作者发表本专业论文情况

**2.1. A calibrated SVM based on weighted smooth GL(1/2)for
Alzheimer's disease prediction**



Volume 158

May 2023

ISSN 0010-4825

Computers in Biology and Medicine

An International Journal



Actions for 1 selected article
Select all / Deselect all

 Download PDF

 Export citation

Show all article previews

Contents

Editorials

Regular Articles

Reviews

Corrigendum

Jin Zhang, Yungui Zhou, Hui Zhang, Hongyi Wu
Article 106807

 View PDF Article preview ▾

Research article • Full text access
A calibrated SVM based on weighted smooth $GL_{1/2}$ for Alzheimer's disease prediction

Jinfeng Wang, Shuaihui Huang, Zhiwen Wang, Dong Huang, ... Yong Liang
Article 106752

 View PDF Article preview ▾

Research article • Open access
Medical matting: Medical image segmentation with uncertainty from the matting perspective

Lin Wang, Xiufen Ye, Lie Ju, Wanji He, ... Zongyuan Ge
Article 106714

 View PDF Article preview ▾

Research article • Full text access
Multi CNN based automatic detection of mitotic nuclei in breast histopathological images

Abdul Rahim Shihabuddin, Sabeena Beevi K.
Article 106815

 View PDF Article preview ▾



A calibrated SVM based on weighted smooth $GL_{1/2}$ for Alzheimer's disease prediction

Jinfeng Wang^{a,*}, Shuaihui Huang^a, Zhiwen Wang^a, Dong Huang^b, Jing Qin^c, Hui Wang^c,
Wenzhong Wang^d, Yong Liang^e

^a College of Mathematics and Information, South China Agricultural University, Guangzhou 510642, Guangdong, China

^b Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

^c School of BECS, Queen's University Belfast, Belfast, UK

^d College of Economics and Management, South China Agricultural University, Guangzhou 510642, Guangdong, China

^e Ping Cheng Laboratory, 312003, Shenzhen, Guangdong, China

ARTICLE INFO

Keywords:
Alzheimer's disease
Calibrated SVM
Sparse regularization
Weighted smooth $GL_{1/2}$

ABSTRACT

Alzheimer's disease (AD) is currently one of the mainstream senile diseases in the world. It is a key problem predicting the early stage of AD. Low accuracy recognition of AD and high redundancy brain lesions are the main obstacles. Traditionally, Group Lasso method can achieve good sparseness. But, redundancy inside group is ignored. This paper proposes an improved smooth classification framework which combines the weighted smooth $GL_{1/2}$ ($wSGL_{1/2}$) as feature selection method and a calibrated support vector machine (cSVM) as the classifier. $wSGL_{1/2}$ can make intra-group and inter-group features sparse, in which the group weights can further improve the efficiency of the model. cSVM can enhance the speed and stability of model by adding calibrated hinge function. Before feature selecting, an anatomical boundary-based clustering, called as ac-SLK-AMI, is designed to make adjacent similar voxels into one group for accommodating the overall differences of all data. The cSVM model is fast convergence speed, high accuracy and good interpretability on AD classification, AD early diagnosis and MCI transition prediction. In experiments, all steps are tested respectively, including classifier comparison, feature selection verification, generalization verification and comparing with state-of-the-art methods. The results are supportive and satisfactory. The superior of the proposed model are verified globally. At the same time, the algorithm can point out the important brain areas in the MRI, which has important reference value for the doctor's predictive work. The source code and data is available at <https://github.com/14345363/COMFOMRI>.

1. Introduction

Alzheimer's Disease (AD) [1] is an irreversible neurodegenerative disease caused by the death of nerve cells in the brain. It is very common in old age and is difficult to detect at the beginning of the onset. It is mainly clinically reflected as intellectual damage. AD is the most common type of dementia. According to the degree of disease, it is mainly divided into Cognitive Normal (CN), Mild Cognitive Impairment (MCI) and AD. From the current point of view, the pathology and causes of AD are complicated. There is no accurate conclusion so far. Some of the drugs used by patients in the treatment of AD can only delay the deterioration of the disease to a certain extent. Therefore, the early diagnosis of AD plays an important role in the prediction and prevention of AD.

1.1. Early AD detection

It is well-known that early detection of AD symptoms can greatly improve the prevention and treatment effects of AD. If patients can receive treatment as early as possible, it will be able to prevent the deterioration of the condition from causing irreparable damage to the patient's body. At the same time, early detection can also allow patients to make arrangements for their future lives. It is of great significance for family members and clinicopathological research. At the present stage, doctors often use the observation method as clinical diagnosis and further treatment of AD because of no systematic diagnosis method. But this method is obviously affected by human factors, which is not only inefficient, but also prone to misjudgement [2]. Therefore, automatic

* Corresponding author.

E-mail address: wangjinfeng@scia.cn (J. Wang).

<https://doi.org/10.1016/j.combiomed.2021.106752>

Received 27 October 2022; Received in revised form 17 January 2023; Accepted 6 March 2023

Available online 24 March 2023

0010-4825/© 2023 Elsevier Ltd. All rights reserved.

differentiation of AD is an important step to understand the progress of AD and help clinicians make decisions.

In order to detect AD more efficiently, many scientific research groups have developed different types of biomarkers from different perspectives, such as structural Magnetic Resonance Imaging (sMRI) [1], functional Magnetic Resonance Imaging (fMRI) [1], fluorodeoxyglucose Positron Emission Tomography (FDG-PET) [5], or Diffusion Tensor Imaging (DTI) [6] and so on. Among all these markers, MRI is a common biomarker. It has a series of advantages, such as non-invasive, usability, high sensitivity to brain changes after the onset of disease [3], and visualization of brain morphology with high spatial resolution [7]. It is an ideal tool to study various brain structures and morphological changes caused by AD. A systematic structural morphometry analysis on gray matter morphology was implemented [8]. Therefore, it is usually cited as a part of the standard clinical evaluation for the diagnosis of AD.

3.2. Clustering in MRI

Based on the structure of MRI, many researchers proposed method for automatic detection of AD [9–11]. For MRI, the number of images that can be acquired is usually much smaller than the feature dimension of the image. Many features in the image are not related to the learning task, so MRI has the characteristics of small samples and high dimension. The direct common solution is to delimit some adaptive regions of interest (ROI) [12] and analyze the trends of these regions. It can reduce the feature dimension and enhance the classification performance. For example, relevant studies [13] have shown that hippocampal atrophy is closely related to Alzheimer's disease. Many researchers [14–16] use multi-scale segmentation (MAS) for hippocampal segmentation, and obtain the label of the image by registering multiple atlas images to the image. But the segmentation accuracy of this method depends on the registration error. Therefore, Peng [17] et al. proposed a new method based on iterative local cross mapping (ILCM) to achieve accurate and robust hippocampal segmentation without registration.

However, these methods heavily rely on the preservation of the structure. It requires a certain understanding of the area affected by the disease and the acquisition of relevant prior knowledge. If the prior knowledge is insufficient, it is a good choice to consider using the characteristics of the whole structure of the brain [18]. This method compresses the description of each structure into a scalar or low-dimensional representation and ignores the detailed information inside the structure. Then the characteristics of the whole structure are obtained. A new supervised learning framework was presented to combine inter-subject similarity features and intra-subject variability features, which helped to predict Alzheimer's disease [19]. Further, if the brain can be combined with some standard matching template, each subject image is mapped to the same standard space by the registered template, and then the target features are extracted from them respectively. This will greatly improve the classification within the structure to derive a more reasonable degree distribution. These standard brain templates can usually divide the brain into several regions, such as the 10-region Anatomical Automatic Labeling (AAL) [20].

3.2.1. Feature selection and classification on ROI

In the current research, there are many algorithms for AD classification [21–23]. Among them, these SVM based on voxel features is widely used in AD classification because of its ease of use and understanding. However, SVM has its limitations in the classification. It ignores the similarity in the structure of the same brain and treats each voxel feature as a completely independent individual. AD is a disease in the time brain structure, so it is also related to AD [24]. Moreover, among all of its weights are not zero, so it is difficult to distinguish regions highly related to AD. In order to solve these problems, researchers usually choose to use some feature selection

methods to screen out the important parts of the features. In this regard, regularization is a good choice. For et al. proposed a Lasso SVM method [25], which replaces the maximum margin term in standard linear SVM with L_1 norm regularization. This method can encourage the sparsity of weights and obtain regions highly correlated with AD. It has certain effect on the classification of AD, but it does not reflect the correlation between the structures. The selected voxel features are very scattered making interpretation difficult. Sun et al. [26] proposed to replace the L_1 sparse term in Lasso SVM with GroupLasso (G_1) [27] for sparse selection. G_1 can make feature groups sparse and extract voxels from areas that are highly related to the disease. This method has got a better classification effect, since $L_{1/2}$ regularization [28, 29] has a better sparse effect than Lasso, some researchers have proposed the GroupLasso with $L_{1/2}$ ($G_{1/2}$) in order to reduce the redundant features of the relevant areas [28, 31]. So a better feature selection capability than G_1 - $G_{1/2}$ can not only select the structure with high correlation, but also remove some redundant voxel features in the selected structure. In this way, some segmentation errors can be fully processed when extracting structural regions to enhance classification performance. But it contains a non-zero absolute value (norm) similar to $L_{1/2}$. So it has difficulty in convergence since in numerical calculation iterations $G_{1/2}$ solves this problem very well [32]. It was used for training neural network hidden layer and achieving good results. According to the current knowledge, no one has used this method in the field of image recognition. However, standard $G_{1/2}$ imposes the same penalty on each group equally, which fails to mark those more important groups for classification. Therefore, a concept of group weight is put forward to break this limitation.

3.2.2. Summary

In the diagnosis of diseases by MRI gray matter images, different subjects may experience brain atrophy due to many factors such as age, health, etc. However, the degree of brain atrophy of patients is higher than that of normal people generally, especially for those who are related to cognitive ability. Therefore, dividing brain regions into groups and making predictions at group level will be able to define more critical brain regions. The selected voxels will be more concentrated. As all we known, $G_{1/2}$ can take on this function. The sparseness among groups can select more critical brain areas. The sparseness within groups can reduce the impact of normal atrophy caused by natural factors. It is similar to the effect of $G_{1/2}$. Meanwhile, $G_{1/2}$ also increases the stability and improves the efficiency comparing with G_1 . The main contributions of this paper are listed as follows:

- (1) Firstly, an innovative template called as wS-LCDAAAL is adopted for dividing more detailed brain regions. Specifically, it uses the Simultaneous Sparse Learning (SSL) method to keep anatomical landmarks and segment adaptively registered AAL template. Small regions (less than 1000 voxels) are merged into the nearest large region with the same AAL label. It can provide a fine division of brain regions based on overall sparseness and reduce fragmentation by hierarchical clustering.
- (2) Secondly, a weighted $G_{1/2}$ method called as wSG $_{1/2}$ is proposed. A group sparsity weight for $G_{1/2}$ is designed to improve the efficiency of group sparsity. In the process of removing redundant groups and redundant voxels, the influence of important brain areas on the sparse effect is improved.
- (3) Finally, an improved classification framework is constructed, which is based on a supervised SVM (s-SVM) combined with wSG $_{1/2}$, called as wSFL - wSG $_{1/2}$. It can greatly improve the speed of convergence and extract important feature regions.

The rest of the paper is organized as follows. In Section 2, the related works is introduced, including grouping templates and different levels of sparse methods. In Section 3, MRI datasets and data preprocessing are presented in detail. In Section 4, the proposed model is proposed and constructed. In Section 5, our proposed method is evaluated and compared with other models and further participated in the comparison with ensemble-based methods. The paper is finally concluded in Section 6.

2. Related work

2.1. Feature grouping template

The grouping of voxel features in sMRI needs to be implemented based on the division of brain tissue regions. Different brain areas have different functions for human body. Therefore, brain tissue segmentation provides the detailed quantitative brain analysis for Alzheimer disease diagnosis, abnormal detection and classification. Anatomical Automatic Labeling (AAL) [17] is a software package which contains the atlas of the human brain. Researchers can use it to obtain the anatomical location of a specific brain area. The most common AAL template is AAL10. Its anatomical data comes from the spatially normalized high-resolution T1 anatomical segmentation of a single subject provided by the Montreal Neurological Institute (MNI). The AAL template used in this article is AAL16 which adds 36 cerebellar regions to AAL10. In order to make the template more suitable for the test data, AAL16 was registered with standard template to make the template space consistent with the performance of subject's image.

However, it may not be the best way to group only according to the anatomical region. The division of the template brain area is manually divided, so there may exist some errors. Some important information may be hidden in the sub-region of anatomy. It will divide a small part of uncorrelated areas into highly correlated areas, or vice versa. It will affect the classification performance. Therefore, a good choice is to cluster the highly correlated voxels together to form a group according to the correlation among voxels [18]. Using supervised can improve the accuracy of the classifier. However, the watershed-based method ignored the anatomical boundary. If this supervised segmentation method is not restricted, it will gather some voxels which are highly correlated but belong to different anatomical regions. Different anatomical regions of a human body may undertake different functions. Therefore, a novel anatomically constrained SDC (ac-SDC) by Sun et al. [2] was proposed.

In this paper, ac-SDC method is used to automatically cluster the regions of AAL template. This method greatly penalizes voxels that span an anatomical area, assuming that voxels in the same area have similar effect. It considers the correlation between the feature $x(i)$ which means gray matter density at point i and the disease label y_i . Voxels are grouped based on image intensity. In this study, image intensity is derived from the Pearson Correlation Coefficient (PCC) map of the average voxel characteristics in different subject groups. PCC is often used to measure the degree of correlation between two eigenvalues. As shown in Fig. 1, the blue area represents a negative correlation and the red one represents a positive correlation. The lighter the color is, the greater the gap between them is. Therefore, PCC for each voxel i is computed as follows:

$$PCC(i) = \frac{\text{cov}(x(i), y)}{\text{std}(x(i)) \cdot \text{std}(y)} \quad (1)$$

Similar to ac-SDC method, our clustering method uses spatial distance d_s , color distance d_c and anatomical distance d_a to evaluate the distance D between pixel i and cluster center C_k .

In the original SDC method, the number k of cluster center is defined and the related voxels are gathered by calculating the distance between pixel i and cluster center C_k . The distance D is divided into two parts, i.e. spatial distance and color distance. After the weighted summation of each part, the distance between voxel and cluster center is finally obtained. Spatial distance d_s , color distance d_c , anatomical distance d_a and distance D are defined as follows:

$$d_s(i, C_k) = \sqrt{(x_{i_1} - x_{k_1})^2 + (y_{i_1} - y_{k_1})^2 + (z_{i_1} - z_{k_1})^2} \quad (2)$$

$$d_c(i, C_k) = \sqrt{|\sin(\theta_{i_1}) - \sin(\theta_{k_1})|} \quad (3)$$

$$d_a(i, C_k) = \begin{cases} 0 & \text{if } i \text{ and } C_k \text{ belong to the same anatomical area} \\ \infty & \text{otherwise} \end{cases} \quad (4)$$



Fig. 1. PCC diagram of different voxel groups. (a) All voxels, (b) $AD > NC$, (c) $AD > MCI$, (d) $AD > DLB$.



Fig. 2. Comparison of clustering results. (a) Original AAL, (b) ac-SDC, (c) ac-SDC with anatomical constraint.

$$D(i, C_k) = \sqrt{\alpha d_s^2 + \beta d_c^2 + \gamma d_a^2 + d_c \cdot d_s \cdot d_a} \quad (5)$$

$(x_{i_1}, y_{i_1}, z_{i_1})$ and $(x_{k_1}, y_{k_1}, z_{k_1})$ are spatial coordinates of pixel i and cluster center C_k , respectively. The color of pixel i is represented in the CIE-LAB color space [14, 19]. α is a positive coefficient which is used to weigh the spatial distance and the color distance. A small α tends to select the spherical supervised region. The bigger the intensity distribution of gray content is, the bigger the effect of α is, and vice versa. So the processed sMRI in this study adopts gray intensity $u(i, j)$ to reflect the color distance. $Gray(i)$ represents the gray matter intensity of voxel i . It is worth noted that the input image here uses the gray intensities of PCC images between different types of subjects rather than the gray intensity of each subject. In addition, $d_a(i, C_k)$ is the anatomical distance from voxel point i to cluster center C_k in anatomical template (i.e. AAL16). If i and C_k belong to the same anatomical area, anatomical penalty is 0, else anatomical penalty is infinite. This setting ensures that features in same group will not cross the anatomical area when clustering. In addition, because group weight k is set according to the results of classification for each group area, small area has too little influence. It will increase the number of groups and the complexity of solution. Therefore, small areas with less than 1000 voxels are merged into the nearest large area with same AAL label to reduce fragmentation. The improved AAL template does not cross the anatomical boundary as shown in Fig. 2. It can be seen that automatic segmentation follows the anatomical boundary defined by

AA(L) template. The complexity of standard SBC method has a linear relationship with the number of voxels in the image and the number of voxels to do with the number of supervoxels [10]. The ac-SLQAAI used in this article has the same complexity as SBC, so it has good improvement in grouping efficiency.

2.2. Feature selection based on sparse method

All datasets have small size with high-dimension. Selecting all voxel features will increase the complexity of algorithm. The intervention of redundant features lead to a decrease in accuracy of AAI classification. So, it is necessary to reduce dimension using feature selection [11].

GL is a sparse method at group level in the traditional regularization method for group sparse. It can remove some redundant groups. But this method cannot remove redundant features in one group. $L_{1/2}$ was proposed to solve this problem. It can not only remove redundant groups, but also select those important features in the selected group. Meanwhile, it can improve the prediction accuracy. But this method still has its limitations. It contains a non-smooth absolute value function which may cause oscillations in numerical calculations and difficulty in converging. SGL_{1/2} method is adopted in this paper, which uses a smooth function to approximate the absolute value function in $L_{1/2}$. So it can achieve better convergence. In addition, in order to improve the influence of important groups on model, weights are set for each group. The setting of group weight speeds up the solving and locates one of key brain regions. The grouping template adopts the registered AA(L) template or the ac-SLQAAI template. They may mirror density value of each area is extracted as the original feature of respective group. On this basis, the regularization effect at group level is realized.

2.2.1. Non-group level norms

The sparse norms at non-group level are performed in all features. They are usually used for sparse and feature selection in the whole brain region. In this study, the main regularization at non-group level include L_1 , $L_{1/2}$, $L_1 + L_2$ elastic net and $L_1 + L_{1/2}$ elastic net.

L_1 norm [12] represents the sum of the absolute values of elements in the vector, so it is also called "Lasso regularization". Since it was proposed [13], Lasso penalty term has been widely used for feature selection in high-dimensional spaces. L_1 is the optimal convex approximation of L_0 norm [14] which has no NP-hard problem. It is easier to optimize and solve than L_0 . The function of L_1 norm is defined as:

$$\|f\|_1 = \sum_{i=1}^n |x_i|, i = (1, 2, 3, \dots, n) \tag{6}$$

L_1 norm is a convex function, which is not differentiable everywhere. That is, L_1 has absolute value, and $|x|$ is not differentiable at $x = 0$. Therefore, gradient descent cannot be used to solve L_1 . need to be approximated as a differentiable function as:

$$\rho_{\text{min}} = \text{max}(0, |x|) = \text{soft}(x, \mu) \tag{7}$$

where μ represents the regularization parameter; and μ represents the step size.

$L_{1/2}$ norm is a regularizer with non-convex penalty proposed by Xu et al. [15]. It has good characteristics with an unbiasedness and sparseness. On this basis, a weighted iterative algorithm was proposed [16], which can transform the solution of $L_{1/2}$ into a series of L_1 solutions. The solution of $L_{1/2}$ is sparser than that of L_1 . It has good sparsity and effectiveness. $L_{1/2}$ can be used as a representative of regularization of L_0 ($0 < \gamma < 1$), which is defined as:

$$\|f\|_{1/2} = \sum_{i=1}^n (|x_i|)^{\gamma}, \gamma = (1/2, \dots, 0) \tag{8}$$

The basis of $L_1 + L_0$ is a common elastic network [17]. This combination allows learning a model with only a few parameters which

are nonzero. But it will retain the characteristics of L_2 . A proportion parameter was used to weigh these two parts. The elastic network can not only get good and effective solutions, but also converge the model fast and stably. Compared with Lasso which is likely to consider only one feature randomly for selection, elastic network tends to choose more features. $L_1 + L_2$ elastic network can be defined as:

$$\|f\|_{L_1+L_2} = \alpha\|f\|_1 + \frac{\alpha(1-\alpha)}{2}\|f\|_2 \tag{9}$$

where α is the hyper-parameter of elastic net and α is a weight L_1 and L_2 with $0 \leq \alpha \leq 1$. When $\alpha = 0$, elastic net becomes L_2 ; when $\alpha = 1$ elastic net becomes L_1 . $L_{1/2} + L_2$ elastic net [18] is similar to $L_1 + L_2$ in general. It expresses the tradeoff between L_2 and $L_{1/2}$. This kind of elastic net can achieve sparser feature sets under the premise of ensuring that an effective solution is found, so that fewer features are selected. It is very efficient for those datasets with large dimensions. $L_{1/2} + L_2$ elastic network can be defined as:

$$\|f\|_{L_{1/2}+L_2} = \beta\|f\|_{L_{1/2}} + \frac{\beta(1-\beta)}{2}\|f\|_2 \tag{10}$$

where α and β have the same meaning as $L_1 + L_2$ elastic network.

2.2.2. Group level norms

The norm at group level refers to the grouping of features. The feature group was processed as a unit. Group level norm can perform sparseness among groups and select the relevant groups. In this paper, since regularization methods at group-level, i.e., GL, $L_{1/2}$ and SGL_{1/2}, are used for experiments.

GL is different from Lasso, it just deals with the sparseness relationship between groups. It can reduce feature dimensions and increase classification performance by removing some irrelevant groups. The function of GL is defined as:

$$\|f\|_G = \sum_{g=1}^G \|x_g\|_2 \tag{11}$$

where g represents different group and x_g is the feature in the g th group. At the same time, its derivative function can be approximated as:

$$\rho_g = \text{max}(0, t) = \frac{\mu}{2} |x_g| \tag{12}$$

in which μ is the regularization parameter and μ is the step size of approximate gradient.

$L_{1/2}$ adopts $L_{1/2}$ norm to replace L_2 norm in GL. Therefore, it can also sparsely select features in one group. $L_{1/2}$ can remove non-zero redundant groups but also redundant features in remaining groups. Then, key features in sparsely explained groups can be obtained. The function of $L_{1/2}$ is defined as:

$$\|f\|_{L_{1/2}} = \sum_{g=1}^G \sum_{i=1}^{k_g} |x_{gi}|^{\gamma} \tag{13}$$

in which k is feature number in each group. The derivative function can be approximated as:

$$\rho_{L_{1/2}} = \text{max}(0, t) = \frac{\lambda \mu}{2|x_{gi}|^{\gamma}} |x_{gi}| \tag{14}$$

where $\rho_{L_{1/2}}$ is a γ -symbolic function related to feature x_{gi} in group g .

As mentioned early, it is difficult for $L_{1/2}$ to converge because of its non-smooth absolute value function. It is similar to $L_{1/2}$ according to the following definition:

$$\|f\|_{\frac{1}{2}} = \sum_{i=1}^n |x_i|^{\frac{1}{2}} \tag{15}$$

SGL_{1/2} was proposed to approximate the absolute value function using a smooth function in $L_{1/2}$, so that it can converge fast and calculate

Table 1
Statistics of subjects distributed in two different datasets (ADNI: ADNI Medical Image Consortium).

Dataset	Sex	Quantity	Age	Gender	Weight
ADNI-3	M	86	74.6 ± 8.4(23-91)	508/1168	81.1 ± 6.1(2-101)
	MCI	43	72.2 ± 10.0(46-92)	280/228	81.3 ± 21.0(2-38)
	CV	72	76.7 ± 4.3(64-89)	488/238	77.1 ± 25.0(4-38)
Cognitive	CV	34	76.3 ± 3.1(69-82)	386/48	81.0 ± 11.0(4-90)
	MCI	60	76.0 ± 7.1(58-91)	476/70	81.0 ± 26.0(4-92)
	MCI	76	74.5 ± 7.4(56-88)	481/238	82.0 ± 29.0(2-90)
	CV	134	74.5 ± 7.1(56-91)	994/339	77.2 ± 17.0(4-88)
CV	MCI	162	76.1 ± 4.4(69-82)	754/868	80.2 ± 10.0(2-38)

smoothly. The method of $SGI_{1/2}$ is defined as:

$$GI_{1/2}(x) = \sum_{i=1}^n \sum_{j=1}^m f(x_{ij})^2 \quad (16)$$

Then the derivative function of $SGI_{1/2}$ can be transferred as:

$$\partial_{x_{ij}} G_{1/2} = \text{grad}(f) = \frac{\partial f}{\partial x_{ij}} = f'(x_{ij}), \quad (17)$$

where $f'(x_{ij})$ is the derivative function of $f(x_{ij})$. $f(x_{ij})$ in $GI_{1/2}$ is replaced with $f(x)$ in $SGI_{1/2}$. $f(x)$ is a smooth function defined as:

$$f(x) = \begin{cases} f(x) & |x| \leq \frac{\epsilon}{2} \\ -\frac{2}{\epsilon^2}x^2 + \frac{2}{\epsilon}x & \frac{\epsilon}{2} < |x| \leq \epsilon \end{cases} \quad (18)$$

where ϵ is a small positive constant. $f'(x) = \begin{cases} f'(x) & |x| \leq \frac{\epsilon}{2} \\ -\frac{4}{\epsilon}x + \frac{2}{\epsilon} & \frac{\epsilon}{2} < |x| \leq \epsilon \end{cases}$ and $f'(x) = 0$ if $|x| > \frac{\epsilon}{2}$.

3. Materials

In this section, structural MRI (sMRI) datasets are used in image preprocessing pipeline. All data in this paper was obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (<http://adni.loni.usc.edu/>).

3.1. Data selection

Two datasets are selected. One dataset is sMRI data selected from ADNI-3, the other is referred from [13] with the same distribution as first part. Each dataset is divided into three categories: MCI (Alzheimer's disease), MCI (mild cognitive impairment) and CN (control normal). Moreover, MCI is further divided into two categories: MCI (being stable in early stage) and MCI (ending in transition to AD). Some parameters are expressed as $Mn \pm [I]Range$.

The ADNI-3 dataset contains total of 317 data selected from different subjects. There are 86 AD patients, 71 MCI patients, 65 MCI patients and 94 CN patients. The Cognitive dataset includes 137 AD, 134 MCI, 76 MCI and 162 CN subjects. The details of the selected dataset are shown in Table 1. They are both original MRI images of ADNI-3 and Cognitive dataset. So these datasets used in the study has been uniformly preprocessed. SPM (S7) is a tool to extract gray matter voxel features from MRI images for calculation. In this paper, SPM12 toolbox is adopted with default parameters (<http://www.fil.ion.ucl.ac.uk/spm/doc/>). CAT12 is used to segment gray matter of brain with default parameters (<http://www.neuro.uni-jena.de/cat/>). Only those voxels inside brain are used for feature construction.

3.2. Preprocessing and settings

Use different brain structures for different individuals, preprocessing is needed before feature selection.

Firstly, images are calibrated by AC-PC line. AC-PC is a commissure line from the midpoint of the posterior edge of anterior commissure

(AC) to the midpoint of the anterior edge of posterior commissure (PC), also known as ACPC line. In brain facts, tomographic anatomy of brain, this line is often used as a baseline. The original images are set to the center and registered to the MNI space in this step.

Secondly, affine registration is finish on the corrected brain images. An affine registration was performed to linearly align each sMRI to the standard template to remove global linear differences. In this study, the Colin27 [14] was used as standard template. After finishing image registration, MRI images have a size of 181 × 217 × 181 voxels.

Thirdly, the registered images are processed by skull removal and segmentation. The skull was stripped to obtain a clearer and higher resolution spatial template. These images without non-brain tissues are obtained. Then brain gray matter voxel images are obtained by tissue segmentation.

Finally, these gray images are resampled after registration. In this paper, all sMRI are resampled to have identical spatial resolution (i.e., 1 × 1 × 1 mm³). The final gray matter voxel features are obtained.

In brain medical images, voxel features are common bio-markers which are usually used to represent local features of certain regions of brain. This paper selects the density characteristics of gray matter voxels from T1-weighted sMRI, which is a common method for AD classification. It operates on the given T1-weighted images $I_{i,j} \in \{1, \dots, n\}$ of a series of subjects. Each image is segmented with SPM to obtain the corresponding regional organization, and the space is normalized to MNI. A modulated density map $V_{i,j} \in \{1, \dots, n\}$ of gray matter tissue is computed by multiplying the spatially normalized gray matter map with the determined Jacobian determinant. After obtaining the gray matter density map of each subject, it was matched with AAL template or ac-SLIC-AAL template. Then the voxel features of each anatomical area are obtained. Each anatomical area is divided into several groups and the voxel features $x_{i,j}$ in each group are extracted. Here, i represents the i th group, and j represents the j th feature of the group. The subjects are labeled by classification sign $y_{i,j} \in \{-1, 1\}$ to distinguish different categories, such as CN and AD.

RA-SVM [15] is employed to automatically register brain segmentations from AAL template space. For ac-SLIC-AAL template, the neighborhood is restricted with 26 nearest neighbors in each direction of 3D space as searching candidates which are ensured not crossing anatomical boundaries. The initial grid size for ac-SLIC-AAL is set to 10 × 10 × 10. The grid may be changed after multiple iterative clustering. The clustering will be stopped at maximum of iterations or none of voxels being changed. The threshold of iteration is set to 15 in this paper. In order to better balance between spatial distance and color distance, the hyper parameter γ is set as 1.

4. Method

In this section, our proposed method will be introduced in detail, including the structure of SVM, a $SGI_{1/2}$ method with group weighting ($GSVM_{1/2}$) as well as the setting of group weighting strategy, and the architecture of the overall proposed model.

4.1. Combined SVM

Support Vector Machine (SVM) [16] is a generalized linear classifier that performs binary classification through supervised learning. Its decision boundary is the maximum margin hyperplane which can be learnt for training samples. SVM uses loss function with maximum margin term. The optimal weight w_{opt} and offset parameter b_{opt} are calculated by solving minimization problem. The form is defined as follows:

$$\min_{w,b} P_{svm}(w,b) = \min_{w,b} \sum_{i=1}^n \max(0, \xi_i) + \frac{1}{2} \|w\|_2^2 \quad (19)$$

$$s.t. \quad y_i(w \cdot x_i + b) \geq 1 \quad (20)$$

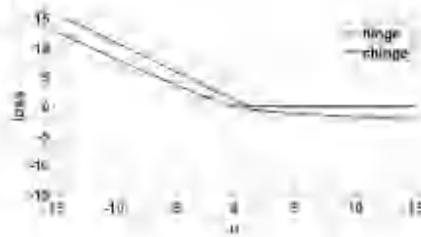


Fig. 3. Comparison of hinge loss (Change loss) and Change loss (hinge loss).

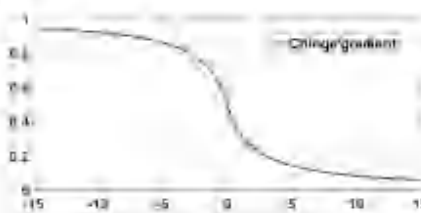


Fig. 4. Change's gradient function.

where $\lambda \|w\|_2^2$ is the maximum margin penalty term and f_{margin} is the loss function. It is suitable for many machine learning tasks which have little prior knowledge about the relationship among features. λ_1 is the parameter of maximum margin penalty which is a non-negative number. The dimension of weight is the same as that of feature vector x . In the standard linear SVM, hinge function is used as loss function $f_{margin}(w) = f_{margin}(w) = \max(0, 1 - w)$. Hinge function uses a soft threshold boundary. The boundary is zero if samples are located on correct side of hyperplane. When samples are classified on wrong side, the boundary is proportional to the distance to hyperplane. So it ensures that the prediction results near to hyperplane tend to be on the correct side. However, hinge function makes the penalty for those results with values being greater than 1, which leads to having no good probability explanation for value distribution.

For overcoming the shortcoming of hinge function, the calibrated hinge (Change) is adopted:

$$f_{margin}(w) = f_{margin}(w) = \max(0, w) = \frac{1}{2} \left(|w| + \frac{w}{|w|} \right) \quad (21)$$

Its gradient function is

$$grad_{f_{margin}}(w) = \frac{1 + \max(0, w)}{2 + |w|} \quad (22)$$

As shown in Fig. 4, if $w \leq 0$ is the part of wrong prediction, Change has a loss penalty trend similar to hinge; if $w > 0$ is the part of correct prediction, Change provides a compensation strategy opposite to punishment. The compensation strategy is reflected by the loss of negative value. The greater the value w , the higher the loss compensation is. Fig. 4 shows Change's gradient function. As can be seen, the accelerated downward trend for wrong prediction, so it can converge quickly and improve the solution efficiency, its downward trend is gradually slowed down, so the growth trend of loss compensation decreases with the increase of the correctly predicted value to prevent overfitting. Change appears as a continuous smooth curve. It can use the gradient descent method to update the weight and improve the efficiency in advance.

4.2. Group weighting strategy

In this paper, a $\lambda_1 GL_{1/2}$ method with group weight ($\lambda_1 GL_{1/2}$) is proposed. A reasonable Group sparsity weight is designed for $\lambda_1 GL_{1/2}$.

to improve the efficiency of group sparsity. In the process of removing redundant groups and redundant voxels, the influence of important brain areas on the sparse effect is improved. In addition, it should be noted that the grouping method is an indispensable part of group sparse method. In this paper, two group templates for grouping are adopted, i.e. traditional SGL template and acellular template designed in this article. The $\lambda_1 GL_{1/2}$ model is defined as follow:

$$GL_{1/2}(x) = \sum_{g=1}^G \beta_g \left(\sum_{i \in g} |x_{i,j}| \right) \quad (23)$$

where β_g represents the group weight. It is used to control the overall sparsity of each group of features. Therefore, a good group weight will significantly improve the efficiency of solving the group sparse model, and even affect the final classification performance. In this paper, a group weight is designed for original SGL [7].

Group weight is an index to measure the importance of each group. The solving process can be accelerated and the key brain regions can be located faster by changing group weight. But how to set group weight reasonably is a key problem. Generally speaking, the commonly used strategies are listed as follows:

- S1 Each group has the same importance, and their group weights are set as 1 by default.
- S2 Based on prior knowledge, researchers need to understand the important brain areas related to disease and assign different group weights according to their importance.
- S3 The group weight is set based on the number of voxel features in each group.
- S4 A separate testing is performed for each group separately. The weight is set by using prediction accuracy.

Among these strategies, the first one ignores the differences of AD reflected in various brain regions. It will increase complexity of solution. The second strategy highly relies on the researcher's understanding background of disease. It is difficult to determine accurate weights for important brain regions. The third strategy is to set group weight according to group area size. In (if a small area greatly affects AD or a large area slightly affects AD, it will greatly affect the solution result. The fourth strategy relies on the accuracy of each group used for classification separately. The weight value is interpretable. Therefore, the fourth strategy is easy to be implemented relatively and is more reasonable. The absolute differences between group weights will not be too large. It can not only increase the initial weight of important groups, but also avoid the decisive influence of group weight on the result and prevent ignoring the group connectivity.

The function of group weight is similar to the setting of regularization hyper-parameters. It affects the sparseness threshold of group as a whole. Therefore, the greater the group weight is, the greater the impact of the group on the overall model is. So the weights of important groups should be larger. In order to verify the impact of the proposed strategy on the classification, the accuracy for each group individually is used as weight after performing 0-1 scaling. The stability of accuracy of each area is tested and analyzed by using cSVM. Ten times of 10-fold cross training is conducted, and finally the average value of each classification result is obtained. Finally, the group weight is determined based on the fourth strategy.

4.3. Proposed model

After determining classification model and feature selection model, a complete architecture of final model is proposed. The semi-norm function of the proposed model is defined as:

$$GL_{1/2}(x) = \frac{1}{2} \sum_{g=1}^G \beta_g \max(|x| < 0, -x) + \frac{1}{2} \|x\|_1 + \frac{1}{2} \|x\|_2^2 + \lambda_2 \sum_{g=1}^G \beta_g \left(\sum_{i \in g} |x_{i,j}| \right) \quad (24)$$

in which λ_1 is the maximum margin penalty, and λ_2 is the hyper-parameter of $uSG L_{1/2}$ to adjust the degree of group sparsity.

The model combines the advantages of cSVM and $uSG L_{1/2}$. It can select key voxels in key brain regions and have better discriminative power for samples near hyperplane. At the same time, it can ensure accurate classification and good interpretability. $f_{SG L_{1/2}}$ is a smooth function which can be minimized by gradient descent method. The complexity of solving model can be greatly reduced. Algorithm 1 presents the complete pseudo-code of the proposed model.

Algorithm 1. The pseudo-code of the proposed model.

```

Input: feature  $x$ , subject label  $y$ , initial weight vector  $w_0$ , initial bias  $b_0$ , nonnegative parameters  $\lambda_1, \lambda_2$ , stepsize  $\alpha$ , group number  $g$ , group weight  $\beta, \tau_1 = 1, \delta = 1$ 
Output: final weight vector  $w$ , final bias  $b$ 
while not converged do
  //Gradient descent
   $k_1 \leftarrow 1 + \alpha_1 \nabla_{w_0}$ 
   $k_2 \leftarrow 1 + \alpha_2 \nabla_{b_0}$ 
  //for each group
  if  $G_1$  then
     $w_{k+1} = \text{max}(w) = \frac{\partial f(w_{k+1})}{\partial w_{k+1}} = 0$ 
  else if  $G_2$  then
     $w_{k+1} = \text{max}(w) = \frac{\partial f(w_{k+1})}{\partial w_{k+1}}$ 
  else if  $G_3$  then
     $w_{k+1} = \text{max}(w) = \frac{\partial f(w_{k+1})}{\partial w_{k+1}}$ 
  else if  $G_4$  then
     $w_{k+1} = \text{max}(w) = \frac{\partial f(w_{k+1})}{\partial w_{k+1}}$ 
  end if
  //Update
   $\lambda_{k+1} = \lambda_k + \alpha_3 \left( \frac{\partial f(w_{k+1})}{\partial \lambda_{k+1}} \right)$ 
   $w_{k+1} \leftarrow w_{k+1} + \frac{\alpha_4}{\tau_1} (w_{k+1} - w_k)$ 
end while
return  $w$  and  $b$ 

```

5. Experiment and analysis

In this section, the impact of different modules of the proposed model is tested and verified on AD diagnosis, which include the following parts.

(1) Classifier selection. Classifier is determined by comparing with several classical methods.

(2) Feature selection verification. The performance of cSVM models with different sparse norms (e.g. $L_1, L_{1/2}, L_1 + L_2$ and $L_{1/2} + L_2$) at group level and non-group level will be verified. The effect of different sparse norms (including $L_1, G, L_{1/2}, SG, L_{1/2}$ and $uSG, L_{1/2}$) is tested on the feature selection and classification.

(3) The generalization of the proposed model is verified by being applied to Cingulat dataset with larger scale.

(4) Our proposed model is compared with several state-of-the-art methods for presenting the performance.

Our proposed method was validated on three tasks of AD classification (i.e. AD vs. CN), Early diagnosis of AD (i.e. CN vs. MCI) and MCI conversion prediction (i.e. MCI vs. MCIc). In this study, the selected ADNI3 data set is used for different model comparison experiments, and Cingulat data set is used to verify the generalization performance. Twenty percent of the data is selected randomly as the test set, and the remaining data is used as the training set and verification set. In order to avoid over-fitting in training process, K -fold cross-validation method is adopted. This method randomly divides the data equally into

Table 2

Classification effect of different classifiers.

Classifier	AUC	AOC	SEN	SPC	Specificity (%)
LC	0.8736	0.8473	0.8628	0.8467	79.6
L1	0.9167	0.9008	0.7443	0.7872	1.49
SVM	0.8386	0.8343	0.8140	0.8467	88.6
cSVM	0.8887	0.8779	0.8670	0.8723	21.9

K parts. $K - 1$ parts are used to train the model, and the remaining one is used to test the model. This process is repeated K times so that each part is tested. In this paper, K is set as 5, i.e. 5-fold cross-validation. The experimental environment is built on a local system in which the processor is Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHz (2201 MHz), the memory size is 16 GB, and programming framework is MATLAB2019a and Python.

In the experiment, five indexes are used to evaluate the classification performance, including Accuracy (ACC), Area Under Curve (AUC), Sensitivity (SEN), Specificity (SPC), and F1-score F_1 . TP and FP indicate the correct and incorrect prediction when the case is positive. FN and FN indicate the incorrect prediction for those negative cases. The definitions of the indexes except for AUC are presented as following.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{FP + TN}$$

$$F_1 = \text{score} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

5.1. Classifier selection

In order to verify the superiority of cSVM over other classifier, four methods are compared to test accuracy of AD vs. CN groups in the ADNI3 dataset, which includes linear classification (LC), logistic regression (LR), SVM and cSVM. The comparison results are shown in Table 2. It can be seen from the results that cSVM has excellent training efficiency while ensuring performance. The accuracy of cSVM is as high as 88.99%. Compared with LinearSVM and LR, the classification effect of cSVM is significantly improved, and the average solution time is also greatly reduced for SVM and LC. Based on the results in Table 2, cSVM is determined as the main classification model combining with different sparse methods.

5.2. Feature selection verification

To investigate the superior of regularization at group level, cSVM is combined with different norms at non-group level and group level respectively for testing performance.

5.2.1. Nongroup level

The sparse norms at non-group level can realize the selection of features. In this section, four regularization methods with different norms (i.e. Lasso, $L_{1/2}, L_1 + L_2$ and $L_1 + L_{1/2}$) are adopted for feature selection.

The weight distribution maps of these models are shown in Fig. 3–6. The selected features using cSVM with Lasso are reduced a lot of our data. cSVM with $L_{1/2}$ selects fewer voxels than that with Lasso, which appears as a sparser representation in weight map. In Figs. 3 and 4, the selection range gradually narrowed as λ_2 changes from left to right. When λ_2 is 0.1 for Fig. 3, the selected voxels are concentrated near hippocampus, parahippocampal gyrus and amygdala. This is similar to the distribution when λ_2 is 0.01 in Fig. 6. Therefore, it can be judged that the morphological area of patient is most different from normal

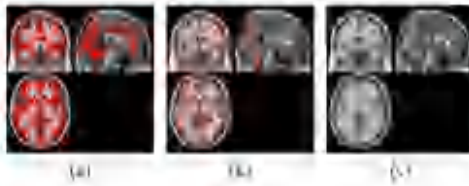


Fig. 5. The distributions of weight map with L_1 (The regularization parameter α increases from left to right: (a) 0.001; (b) 0.01; (c) 0.1).

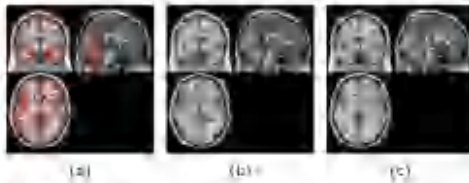


Fig. 6. The distributions of weight map with L_2 (The regularization parameter α increases from left to right: (a) 0.001; (b) 0.01; (c) 0.1).

Table 3

Regularization	AOC	AUC	ACC	SNR	SI
L_1 (elastic)	0.8005	0.8628	0.8448	0.0020	0.0747
L_2	0.8055	0.8665	0.8500	0.0020	0.0840
L_1+L_2	0.8105	0.8719	0.8549	0.0020	0.0914
$L_{1/2}+L_2$	0.8111	0.8729	0.8550	0.0020	0.0917

person in these areas, which may be caused by regional atrophy related to AD. In addition, cSVM with L_1+L_2 and $L_1+L_{1/2}$ also has good sparse effect. Although these sparse methods show the most important brain locations in weight map, these locations are quite isolated and have no good structure. It is difficult to determine which brain regions are actually highly related to AD. As shown in Figs. 5 and 6, the sparse model based on elastic network is more sensitive to the morphological differences of brain; so the selected areas are more comprehensive. After gradually increasing the sparse proportion in elastic network, the selection range is gradually narrowed from left to right. Finally, it is also concentrated in the vicinity of hippocampus, parahippocampal gyrus and amygdala. But there are more selected vessels and more concentrated distribution.

The results of four different sparse methods classifying AD vs. CN groups are shown in Table 3. We can see that cSVM combined with L_1+L_2 has better performance than other norms. L_1+L_2 elastic network integrates the advantages of L_1 and L_2 and balances their respective contributions to achieve a better effect. The accuracy of cSVM with L_1+L_2 is slightly higher than that of the model with $L_{1/2}+L_2$. It is due to the lack of some key features derived from the larger sparseness of elastic net with $L_1+L_{1/2}$. In general, classification results of the two elastic network methods are significantly higher than that of models with L_1 or $L_{1/2}$. As shown in Fig. 7, the sparse model based on $L_{1/2}+L_2$ elastic network is highly sensitive to the morphological differences of the brain, so the selected feature contains the entire brain region. After gradually increasing the sparse proportion in $L_{1/2}+L_2$ elastic network, the selection range is gradually narrowed from left to right. Finally, it is also concentrated in the vicinity of hippocampus, parahippocampal gyrus and amygdala. Even so, it is difficult to locate the brain regions highly related to the disease while ensuring the high classification effect.



Fig. 7. The distributions of weight map with L_1+L_2 (The regularization parameter α increases from left to right: (a) $\alpha=1$, $\alpha=0.01$; (b) $\alpha=1$, $\alpha=0.01$; (c) $\alpha=0.1$, $\alpha=0.1$).

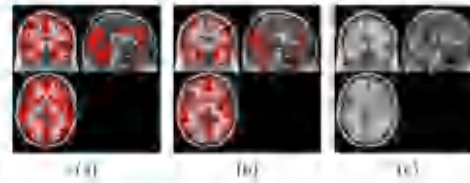


Fig. 8. The distributions of weight map with $L_1+L_{1/2}$ (The regularization parameter α increases from left to right: (a) $\alpha=1$, $\alpha=0.001$; (b) $\alpha=1$, $\alpha=0.001$; (c) $\alpha=0.1$, $\alpha=0.1$).

5.2.2. Regularization at group level

The weight map at non-group level shows the main brain locations being quite isolated. They do not well reflect the distribution of brain structures related to disease, which makes it difficult to determine which brain regions are actually involved in classification of disease. The regularization at group level can select larger regions. The vessels in each region are relatively concentrated, which can better display the brain regions related to AD diseases. Four group regularization methods (i.e. uL_1 , $uGL_{1/2}$, $uGL_{1/2}$ and $uSGL_{1/2}$) are tested on AD vs. CN for comparison. The uSGLC-AAL is just adapted to group template. The group weight is set according to classification impact of each region. Each brain area affects AD in different degree. In order to highlight the more important brain areas affected by the disease, the sparseness threshold of weight is set. Weights less than the threshold will be set as 0. This setting aims to highlight the distribution of important areas.

Those highly correlated vessels can be selected by $uL_{1/2}$, $uGL_{1/2}$ and $uSGL_{1/2}$ in highly correlated brain regions. There is the same sparsity principle for each group and intra group features. $uGL_{1/2}$ improves the efficiency and stability by adding a smoothing function on $uL_{1/2}$. $uSGL_{1/2}$ further speeds up the selection efficiency and improves the effect of feature selection by setting group weights. In addition, $uSG_{1/2}$ and $uSGL_{1/2}$ can make the selected intra area sparse by removing some irrelevant vessels. Therefore, as increasing the penalty term in norm, only those vessels with low correlation in highly correlated brain regions will be removed.

The results of cSVM combined with different group methods are shown in Table 4. It can be seen that $uSGL_{1/2}$ has the best performance. The weight maps of $uSGL_{1/2}$ in cSVM model are shown in Figs. 9 and 10. The areas divided by the uSGLC-AAL template is more detailed and can show the overall differences of different task images. Therefore, the regions located in the weight distribution are more concentrated and sparse. In $uSGL_{1/2}$ model, the sparsity threshold is set as 10^{-20} to filter any weights. Therefore, it can be seen from Fig. 9 that the red marked regions are approximately symmetrical, which are more likely to announce some key brain regions.

5.2.3. Proposed model for different tasks

According to the previous results, sparsity at group level is feasible. It can make the selected features more being concentrated on areas

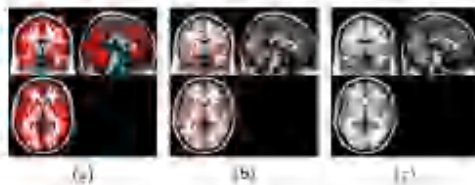


Fig. 9. The distribution of weights map with $aSLG_{1,2}$. The registration parameters: (a) 0.0; (b) 0.1; (c) 0.2.

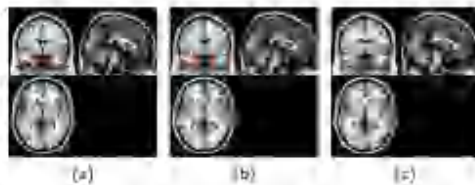


Fig. 10. The distribution of weights map with $aSLG_{1,2}$. The registration parameters: (a) 0.0; (b) 0.1; (c) 0.2.

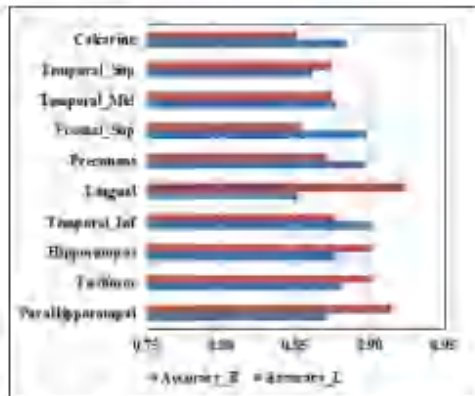


Fig. 11. The top 10 regions on the registered AAL template.

Table 4
Performance of $aSLG_{1,2}$ with different template registration on AD vs. CN.

Template	Area	ACC	AUC	sD	SP	SI
AAL	Groupwise	0.9207	0.9823	0.9286	0.9340	0.9272
	$aSLG_{1,2}$	0.9317	0.9852	0.9370	0.9320	0.9223
	$aSLG_{1,2}$	0.9282	0.9728	0.9285	0.9320	0.9286
ac-SLIC-AAL	Groupwise	0.9167	0.9598	0.9185	0.9340	0.9167
	$aSLG_{1,2}$	0.9222	0.9715	0.9280	0.9340	0.9192
	$aSLG_{1,2}$	0.9278	0.9798	0.9330	0.9320	0.9219
ac-SLIC-AAL	Groupwise	0.9044	0.9390	0.9119	0.9320	0.9119
	$aSLG_{1,2}$	0.9100	0.9500	0.9150	0.9320	0.9150
	$aSLG_{1,2}$	0.9150	0.9600	0.9200	0.9320	0.9200

related highly in disease. The eSVM is combined with the $aSLG_{1,2}$ sparse method to form the final prediction framework, $aSLG_{1,2}$ -eSVM. But in practice, many patients are usually in MCI state between AD and CN. Their clinical manifestations are usually not as serious as AD patients. MCI is further divided into MCIa and MCIc. It is more

Table 5
Performance of $aSLG_{1,2}$ -eSVM with different templates on different tasks.

Task	Template	ACC	AUC	SP	SI	SI
AD vs. CN	AAL	0.9275	0.9318	0.9419	0.9149	0.9257
	ac-SLIC-AAL	0.9144	0.9291	0.9419	0.9465	0.9400
CN vs. MCIa	AAL	0.7100	0.7207	0.7074	0.6203	0.7190
	ac-SLIC-AAL	0.7238	0.8073	0.8005	0.8806	0.7676
MCIa vs. MCIc	AAL	0.6421	0.7082	0.6627	0.6134	0.6627
	ac-SLIC-AAL	0.6842	0.8121	0.8044	0.8208	0.6840

difficult to distinguish early diagnosis of AD (i.e. CN vs. MCIa) and prediction of MCI conversion (i.e. MCIa vs. MCIc). In order to verify the effectiveness of the proposed model for two types of tasks, it will be tested based on different grouping templates.

The results are shown in Table 5. For early diagnosis of AD, the accuracy of the proposed model based on AAL and ac-SLIC-AAL are 71.00% and 73.20% respectively. The accuracy using ac-SLIC-AAL is 66.82% for MCI conversion prediction task. It is significantly higher than $aSLG_{1,2}$ -eSVM with AAL. According to the evaluation indicators of the classification, the model based on ac-SLIC-AAL template is significantly better than the model based on AAL, especially in the AD early diagnosis task. In general, the performance of ac-SLIC-AAL template is generally better than that of the registered AAL template. It segmented each brain region more finely with fewer voxels and was more suitable for correlation testing by individual region. Therefore, modified templates will have their advantages. The experimental results prove that $aSLG_{1,2}$ -eSVM using ac-SLIC-AAL template can also reflect the best effect for other tasks. This is a strong support for the reliability of the proposed model.

5.2.4. Important groups analysis

In order to verify the effect of each group of features from different group templates, the data will be respectively classified according to each relevant area selected by the proposed model. In showing the effect of each brain region related to AD, classification is implemented for each group in left and right brains based on AAL template. The performance of each part are shown in Fig. 11 with the top 10 regions. It can be seen that the top 6 regions include Lingual, ParaHippocampal, Fusiform, Hippocampus, Frontal, Temporal_Mid, and Frontal_Sup.

The ac-SLIC-AAL template performs adaptive area division based on the registered AAL template in order to better cluster similar voxels which do not cross anatomical boundary. It is able to divide brain area finely by clustering similar feature distributions into one group. So it prevents those regions from different anatomical areas being integrated to one group. ParaHippocampal has the highest accuracy reaching up to 94.71%. The right parahippocampal gyrus region corresponds to this group exactly, which was just selected in previous section.

The key areas selected by the two templates are roughly the same. The results are shown in Fig. 10(b). They both include right lingual gyrus, right parahippocampal gyrus, right fusiform gyrus, right hippocampus, left inferior temporal gyrus and left middle temporal gyrus. The results indicate that these areas affects AD more or more degree than other areas.

5.3. Generalization validation

For verifying the generalization of the proposed model, it is applied to Cambridge dataset which was divided into different control groups (AD vs. CN, MCI vs. CN, MCI vs. AD, MCIa vs. MCIc and CN vs. MCIc). The training data is re-optimized by adjusting the hyper parameters of registration. Other parameters are consistent with the proposed model. A eSVM model without group sparseness is added into comparison. The experimental results of these models on Cambridge dataset are shown in Table 6, in which the best ones was marked in bold and

Table 6
Comparison of important brain regions between AAL-based model and ω -SGLC-AAL-based model.

AAL	ω -SGLC-AAL	Proposed model
Lingual_R	Fusiform_R	Hippocampus_L
Parahippocampal_R	Lingual_L	Hippocampus_R
Fusiform_R	Parahippocampal_R	Parahippocampal_L
Fusiform_L	Hippocampus_L	Parahippocampal_R
Hippocampus_R	Temporal_Inf_L	Amygdala_L
Temporal_Inf_L	Temporal_Pole_Sup_L	Lingual_R
Frontal_Sup_L	Temporal_Mid_L	Fusiform_L
Temporal_Pole_Sup_L	Oxyton_Pole_L	Fusiform_R
Fusiform_L	Hippocampus_R	Temporal_Mid_L
Temporal_Mid_L	Parahippocampal_L	Temporal_Inf_L

Table 7
Classifier performance in Cuiqing dataset.

Tasks	Reps	ACC	AUC	SEN	SPE	F1
AD vs CN	ω -SGL	0.8997	0.9254	0.8970	0.9412	0.9213
	GL	0.9164	0.9382	0.9124	0.9198	0.9163
	ω -SGL _{1/2}	0.9231	0.9706	0.9124	0.9321	0.9322
AD vs MCI	ω -SGL	0.6295	0.6561	0.6788	0.6143	0.5981
	GL	0.6517	0.7991	0.6861	0.6286	0.6567
	ω -SGL _{1/2}	0.6671	0.7108	0.7988	0.6238	0.6636
MCI vs CN	ω -SGL	0.6801	0.7221	0.7346	0.6381	0.6667
	GL	0.6855	0.7886	0.7222	0.7573	0.6889
	ω -SGL _{1/2}	0.6882	0.7838	0.7100	0.6667	0.6909
MCI vs MCIc	ω -SGL	0.6285	0.6572	0.6942	0.5970	0.5714
	GL	0.6618	0.6838	0.7237	0.6268	0.6735
	ω -SGL _{1/2}	0.6619	0.6814	0.7105	0.6343	0.6713
CN vs MCIc	ω -SGL	0.6528	0.6888	0.6952	0.6719	0.6821
	GL	0.6757	0.7239	0.6832	0.6842	0.6981
	ω -SGL _{1/2}	0.6868	0.7272	0.7222	0.6478	0.7156

those slightly lower but very similar results are marked in italics. It can be seen that the proposed model has the highest accuracy and the dominating F1-score for all profiling tasks. F1-score indicates the performance of model, the proposed model is with thousandth difference with the top one for MCIc vs. MCIc. In the same situation, ω -SGL_{1/2} still has the highest accuracy and specificity, and the other two indicators are very close to GL. CN vs. MCIc is most concerned by people and is most meaningful and difficult for prevention and intervention. The performance of proposed model is satisfying and acceptable for this situation. In addition, it is obvious for Cuiqing dataset that the performance of ω -SGL_{1/2}-SVM is better than that of SVM without regularization definitely. It fully demonstrates the importance of regularization and the generalization of this model.

5.4. Methods comparison

In this section, the proposed method is compared with some state-of-the-art models which includes machine learning methods and deep learning methods. The brief introduction and setting of these methods are shown as follows.

5.4.1. Methods' introduction

SVM+SP+AP(2012) [11]. This method proposes a general SVM framework for introducing spatial priors (SP) and anatomical priors (AP) into support vector machines. It will refer to the Laplacian of the graph encoding spatial proximity(L_s). As a result, SVMs combined with both regularization terms are to consider spatial and anatomical optimization problem.

SVM+GL+SAR(2018) [23]. A SVM based on Group Lasso (GL) integrating spatial anatomical regularization (SAR) is proposed. SAR is to regularize the weight map, such that two points, which are spatial neighbors and additionally belong to the same anatomical region, should have similar weights. In this way, spatial-neighbor features in

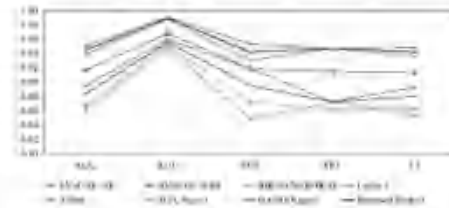


Fig. 12. Comparison of classification effects of different methods in all diagnosis tasks.

the same anatomical region are encouraged to have similar weights in the SVM model. Additionally, structural sparsity is induced to obtain key regions involved in disease.

RBF-SVM+ROICSE(2021) [42]. A ROI-based contourlet subband energy (ROICSE) feature is proposed to represent sMRI images in the frequency domain for AD classification. Specifically, a preprocessed sMRI image is firstly segmented into multiple ROIs by a constructed group template. Instead of extracting features from ROIs in the spatial domain, a contourlet transform is performed on each ROI to obtain its energy subbands. Then, the subband energy (SE) feature vector is constructed to obtain its energy distribution and contour information. Afterwards, the SE feature vectors of each ROI are concatenated to form the ROICSE feature of the sMRI image. Finally, classification is performed on different tasks using a radial basis function (RBF) SVM classifier.

LetNet-5(1998) [43]. LetNet-5 is a convolutional neural network with simple structure and high efficiency. It can make good use of the structural information of the image. In the original LetNet-5, the input image first passes through the convolution layer twice to the pooling layer, then through the full connection layer, and finally reaches the output layer.

ZNet(2014) [44]. ZNet introduces a novel visualization technique (multilayer deconvolution network). It projects the feature activation part back to the input pixel space, which can deeply understand the function of the middle feature layer and the operation of the classifier. The evolution of features can be observed during training and potential problems in the model can be diagnosed.

H-FCN(2018) [45]. A hierarchical deep learning framework is proposed. It combines differential atrophy localization with feature extraction and classifier construction for sMRI-based AD diagnosis. Specifically, a hierarchical fully convolutional network (H-FCN) is proposed to identify patch-level and regional-level discriminative locations automatically and hierarchically in whole-brain sMRI.

DA-MIDL(2021) [46]. A dual-attention multiple-instance deep learning(DA-MIDL) model is proposed to identify discriminative pathological sites for AD diagnosis. Specifically, DA-MIDL consists of three main components, i.e. the Patch-Nets, the attention multi-instance learning (MIL) pooling module and the attention-aware global classifier. DA-MIDL can learn discriminative structural features from multiple local sMRI patches distributed in the brain through a patch network with spatial attention blocks.

3.4.2. Comparison and discussion

In this section, a comparative experiment is implemented for AD and CN tasks of ADNI-3 dataset. And a variety of evaluation indicators are used to evaluate the results which are shown in Fig. 12. It is obvious that the proposed model has better performance than most of them except for two deep learning methods.

The results are shown in Table 8 for machine learning methods. SVM+SP+AP takes into account the proximity of brain space and anatomy, but this method may ignore the differences in individual

- [19] C. Ieracitano, N. Mammone, A. Hussain, F.C. Morabito, A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia, *Neural Netw.* 123 (2020) 175–190.
- [20] V.S. Nari, C.A. Hans, W.H. Crown, R. Au, W.J. Burke, D.M. Sanghani, J. Bleicher, Machine learning models to predict onset of dementia: A label learning approach, *Alzheimer's Demantia Transl. Res. Clin. Intervent.* 5 (2019) 918–925.
- [21] R. Cotinset, J.A. Glasbey, M. Chupin, H. Bernal, O. Collin, Spatial and anatomical regularization of SVM: A general framework for neuroimaging data, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (3) (2012) 682–696.
- [22] R.-E. Fui, K.-W. Chung, C.-J. Hsieh, X.-R. Wang, C.-J. Lin, LIBLINEAR: A library for large linear classification, *J. Mach. Learn. Res.* 9 (2008) 1871–1874.
- [23] Z. Sun, Y. Qiao, B.P. Lelieveldt, M. Staring, A.D.N. Initiative, et al., Integrating spatial-anatomical regularization and structure sparsity into SVM: Improving interpretation of Alzheimer's disease classification, *NeuroImage* 178 (2018) 445–460.
- [24] L. Meier, S. Van De Geer, P. Bühlmann, The group lasso for logistic regression, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 70 (1) (2008) 53–71.
- [25] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 68 (1) (2006) 49–67.
- [26] Z.-S. Xu, H.-L. Guo, Y. Wang, H. Zhang, Representative of $L_{1/2}$ regularization among $L_{1/2}$ ($0 < \rho < 1$) regularizations: An experimental study based on phase diagrams, *Acta Numer. Sinica*. 20124 38 (7) (2012) 1225–1228.
- [27] Z. Xu, X. Cheng, F. Xu, H. Zhang, $L_{1/2}$ Regularization: A thresholding representation theory and a fast solver, *IEEE Trans. Neural Netw. Learn. Syst.* 23 (7) (2012) 1013–1027.
- [28] H.Z. Alemi, J. Zhao, F. Li, W. Wu, Group $L_{1/2}$ regularization for pruning hidden layer nodes of feedforward neural networks, *IEEE Access* 7 (2019) 9546–9557.
- [29] F. Li, J.M. Zúñiga, W. Wu, Smooth group $L_{1/2}$ regularization for input layer of feedforward neural networks, *Neurocomputing* 314 (2018) 109–119.
- [30] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Sussangkarn, SLIC superpixels compared to state-of-the-art superpixel methods, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (11) (2012) 2274–2282.
- [31] G. Pang, J. Stoeckel, SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information, *Knowl. Inf. Syst.* 11 (2) (2007) 243–258.
- [32] Q.-B. Guo, D.-P. Jia, Improved PageRank algorithm of merging feedback information and topical relationship, *Comput. Eng. Des.* 32 (12) (2011) 4071–4074.
- [33] R. Tibshirani, Regression shrinkage and selection via the Lasso, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 58 (1) (1996) 267–288.
- [34] I. Doulberides, M. Defris, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Commun. Pure Appl. Math. 7. Issued Commun. Int. Math. Sci.* 57 (11) (2004) 1413–1457.
- [35] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 67 (3) (2005) 301–308.
- [36] H.-L. Hoang, Y. Jiang, Hybrid $L_{1/2} + L_2$ method for gene selection in the Cox proportional hazards model, *Comput. Methods Programs Biomed.* 164 (2018) 65–73.
- [37] J. Polzella, H.U. Voss, K. Tabelow, Structural adaptive segmentation for statistical parametric mapping, *NeuroImage* 52 (2) (2010) 515–523.
- [38] J. Talariach, P. Tournoux, Co-planar stereotaxic atlas of the human brain: an approach to medical cerebral imaging Thieme Medical Publishers, Stuttgart, 1988.
- [39] B. Aubert-Broche, A.C. Evans, I. Collins, A new improved version of the realistic digital brain phantom, *NeuroImage* 32 (1) (2006) 138–145.
- [40] Y. Garcia-Vizquez, S. Reig, J. Janssen, J. Pascua, A. Rodríguez-Ruano, A. Urdas, J. Chamorro, J.J. Vaquer, M. Descu, Use of IBASPM atlas-based automatic segmentation toolbox in pathological brains: Effect of template selection, in: 2008 IEEE Nuclear Science Symposium Conference Record, IEEE, 2008, pp. 4270–4272.
- [41] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, *ACM Trans. Intell. Syst. Technol.* 2 (3) (2011) 1–27.
- [42] J. Feng, S. Zhang, L. Chen, Extracting ROI-based contourlet subband energy feature from the sMRI image for Alzheimer's disease classification, *IEEE/ACM Trans. Comput. Biol. Bioinform.* (2021).
- [43] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324.
- [44] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: *European Conference on Computer Vision*, Springer, 2014, pp. 818–833.
- [45] C. Liu, M. Liu, J. Zhang, D. Shen, Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (4) (2018) 880–893.
- [46] W. Zhu, L. Sun, J. Huang, L. Han, D. Zhang, Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI, *IEEE Trans. Med. Imaging* 40 (9) (2021) 2354–2366.

2.2. Parallel computing of fuzzy integrals: Performance and test



IOS Press | IOS Press Content Library | Help | About us | Contact us

Home | Journals | Cart | Log In / Register

Search: Search

Published between: and [Search syntax help](#)

Your search for: 'Parallel computing of fuzzy integrals: Performance and test' has returned **2,793** results. (0.376s) [Save search](#)

Sort by: **Relevance** | Show: **10** | 1 | 2 | 3 | 4 | »

Mark all | [Add marked to cart](#) | [Export marked citations](#)

Parallel computing of fuzzy integrals: Performance and test [Cite](#)

Authors: Wang, Jinfeng | Huang, Shuaihu | Jiang, Fajian | Zheng, Zhishen | Du, Jianbin | Chen, Hao | Chen, Runjian | Wang, Wenzhong

Article Type: Research Article

Abstract: *Fuzzy integral* in data mining is an excellent information fusion tool. It has obvious advantages in solving the combination of features and has more successful applications in classification problems. However, with the increase of the number of features, the time complexity and space complexity of *fuzzy integral* will also increase exponentially. This problem limits the development of *fuzzy integral*. This article proposes a high-efficiency *fuzzy integral-Parallel and Sparse Frame Based Fuzzy Integral (PSFI)* for reducing time complexity and space complexity in the calculation of... [Show more](#)

Keywords: *Parallel computing*, sparse storage, *fuzzy integral*, *fuzzy measure*

DOI: 10.3233/JIFS-210372

Citation: Journal of Intelligent & Fuzzy Systems, vol. 41, no. 2, pp. 3137-3159, 2021

[Get PDF](#)

Filter by publication date

- Since 2023 (206)
- Since 2022 (570)
- Since 2021 (1,006)
- Since 2020 (1,374)

[Show more](#)

Filter by journal

- Journal of Intelligent & Fuzzy Systems (1,745)
- Integrated Computer-Aided Engineering (118)
- Intelligent Data Analysis (115)
- Intelligent Decision Technologies (62)

[Show more](#)

Parallel computing of fuzzy integrals: Performance and test

Jinfeng Wang^{a,b,*}, Shuaihui Huang^a, Fajian Jiang^a, Zhishen Zheng^a, Jianbin Ou^a, Hao Chen^a,
Runjian Chen^c and Wenzhong Wang^{d,†}

^aCollege of Mathematics and Informatics, South China Agricultural University, Guangzhou, China

^bGuangzhou Key Laboratory of Smart Agriculture, Guangzhou, China

^cGuangdong Electronic Certification Authority Co., LTD, Guangzhou, China

^dCollege of Economics and Management, South China Agricultural University, Guangzhou, China

Abstract. Fuzzy integral in data mining is an excellent information fusion tool. It has obvious advantages in solving the combination of features and has more successful applications in classification problems. However, with the increase of the number of features, the time complexity and space complexity of fuzzy integral will also increase exponentially. This problem limits the development of fuzzy integral. This article proposes a high-efficiency fuzzy integral—Parallel and Sparse Frame Based Fuzzy Integral (PSFI) for reducing time complexity and space complexity in the calculation of fuzzy integrals, which is based on the distributed parallel computing framework-Spark combined with the concept of sparse storage. Aiming at the efficiency problem of the Python language, Cython programming technology is introduced in the meanwhile. Our algorithm is packaged into an algorithm library to realize a more efficient PSFI. The experiments verified the impact of the number of parallel nodes on the performance of the algorithm, test the performance of PSFI in classification, and apply PSFI on regression problems and imbalanced big data classification. The results have shown that PSFI reduces the variable storage space requirements of datasets with plenty of features by thousands of times with the increase of computing resources. Furthermore, it is proved that PSFI has higher prediction accuracy than the classic fuzzy integral running on a single processor.

Keywords: Parallel computing, sparse storage, fuzzy integral, fuzzy measure

1. Introduction

Fuzzy integral is a good classification tool. Its integral feature supports the calculation of all combined feature values between features. Compared with the decision tree method, it has more obvious advantages in feature combination. It can also be used to fuse multiple classifiers. Fuzzy integral is an excellent nonlinear classifier, but its time complexity and space complexity are both $O(2^n)$. The exponential complexity makes fuzzy integral be unable to solve the combined features of all features in a limited time, so its advantages have not been fully utilized.

As early as 1974, Sugeno proposed fuzzy integral firstly [1]. Subsequently, the Choquet integral was proposed based on Choquet capacity principle [2]. Choquet and Sugeno fuzzy integrals belong to nonlinear integrals, which can consider the interaction between features, thereby improving the accuracy of the classifier. In recent years, the theory of fuzzy integral has been enriched and expanded. Some scholars have applied fuzzy integral to various fields of industry. Choquet and Sugeno fuzzy integrals have achieved good results in wood quality evaluation, public attitude analysis on nuclear energy, color printing picture evaluation, speaker design and human reliability analysis [3–8]. In 2000, a more intuitive and interpretable definition of fuzzy integral was proposed which is called Generalized fuzzy integral for short [9]. Later, the research on WCIPP

*Corresponding authors: Jinfeng Wang, South China Agricultural University, E-mail: wangjinfeng@scau.edu.cn and Wenzhong Wang, E-mail: wangwenzhong@scau.edu.cn

(weighted Choquet integral based projection pursuit) was launched [10, 11]. A new fuzzy integral calculation method was introduced to convert the nonlinear integral to a linear function and use the genetic algorithm to solve the fuzzy measure in Ref. [12]. In 2009, the lasso regression method was used to learn the fuzzy measure on the fuzzy integral in Ref. [13]. The experiment showed that the lasso regression played a certain role in the extraction of fuzzy integral feature. Fuzzy integral is also often used in model fusion, which considers different combinations of multiple models to achieve better classification results [14, 15]. In recent years, researchers have used fuzzy integral fusion classifiers for imbalanced big data classification and achieved good results [16, 17]. This type of research combined the advantages of other classification models with the advantages of fuzzy integral considering the interaction. So the infusion of classifiers can achieve better classification results. The number of features is the number of classifiers. The calculation amount of fuzzy integral is not large.

In the research on fuzzy integrals at home and abroad, a large number of studies have mentioned that obtaining fuzzy measures is a important and difficult problem. As the number of features increases, the time complexity of fuzzy integrals increases exponentially [18–20]. In the current research on fuzzy integral, it is necessary to adopt feature extraction on datasets with a lot of features. This operation often results in fuzzy integral not being able to exert its advantages, because the number of selected features is small and the loss of feature information maybe large. Fuzzy integral is an excellent information fusion tool. It has obvious advantages in solving the combination of features. However, the time complexity problem limits the development of fuzzy integral. Therefore, we hope to reduce the computational complexity of fuzzy integrals with the help of parallel computing, while at the same time improving performance.

In recent years, many scholars have launched parallel research on data mining based on Hadoop MapReduce. On the classification, parallel algorithms of kNN, Decision Tree and Naive Bayes [21–22], Parallel Random Forest algorithm [23, 24], SVM and its application on genetic data classification [25, 26] were implemented on MapReduce. Although many Hadoop parallel algorithms have been proposed, Hadoop is not as efficient as the newer Spark framework in multi-step iterative computing. Hadoop's parallel model is designed save the inter-

mediate results of MapReduce to disk, making tasks wait for I/O for a longer time. Spark uses RDD based on memory calculations which saves intermediate results to memory, so it is less affected by IO in the iterative algorithm. Parallel Logistic regression based on Spark and Hadoop were compared [27]. Experiments show that Spark is more ten times than Hadoop in multiple iteration calculations. In addition, studies have shown that in parallel computing of different algorithms such as SVM, association rules, and clustering, the Spark parallel computing framework is 2–3 times (or even dozens) of times faster than the Hadoop MapReduce framework [28–32]. The above research has achieved good results in distributed parallel algorithms. It showed that more and more traditional algorithms are combined with big data frameworks such as Hadoop and Spark to improve their application capabilities in big data processing. For iterative algorithms, more researchers choose the Spark framework because Spark is more efficient. Therefore, the purpose of this paper is to improve the computing efficiency of fuzzy integral based on Spark. Parallel computing and sparse storage are used for the learning of fuzzy measures in fuzzy integral to reduce the space and runtime so that the advantages of fuzzy integral can be applied to more scenarios. At present, no other research groups have proposed parallel algorithms for fuzzy integral.

The structure of this article is as follows. The second part describes the overview of fuzzy integral used in this article. The third part first describes the related technologies of parallel computing and sparse storage applied to fuzzy integrals, and then proposes parallel computing methods for solving fuzzy measures. The fourth part is the experimental part, which implements experiments on fuzzy integrals with parallel computing and sparse storage, and analyzes and discusses the results. The last part summarizes the whole work.

2. Overview of fuzzy integral

This work was proposed for solving the problem of computing complexity for fuzzy integrals. So, the background knowledge about Fuzzy integral will be introduced firstly. Fuzzy integral not only considers the role of a single feature, but also considers the interaction among features. This section will introduce the basic theory of fuzzy integral and an example for illustrating it.

2.1. Definition of choquet integral

Choquet integral was defined based on fuzzy measure which can describe the interaction between features [33]. So we firstly present the mechanism of fuzzy measure.

When the data set $X = \{x_1, x_2, \dots, x_n\}$ is finite, the power set of X is denoted by $P(X)$ \mathcal{F} represents a σ -algebra of subsets of $P(X)$.

Definition 1. Let (X, \mathcal{F}) be a measurable data space. A discrete fuzzy measure μ defined on the measurable space $P(X)$ is a set function $\mu : P(X) \rightarrow (0, 1)$ satisfying the following axioms:

- 1) $\mu(\emptyset) = 0, \mu(X) = 1$. This is the usual convention, although in fuzzy measure can be any positive finite (or infinite) quantity.
- 2) $A \subseteq B \Rightarrow \mu(A) \leq \mu(B), A, B \in \mathcal{F}$.

Definition 2. A λ -fuzzy measure [1], g_λ , is a special kind of fuzzy measure defined on $P(X)$ and satisfying the finite λ -rule. That is, $A, B \in \mathcal{F}, A \cap B = \emptyset$,

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A)g_\lambda(B),$$

where $\lambda \in (-1, \infty)$.

However, the non-negativity of fuzzy measure is too restrictive for real applications. So the generalized fuzzy measure was proposed.

Definition 3. Set function $\mu : P(X) \rightarrow (-\infty, \infty)$, is called as a signed fuzzy measure if it is monotone non-decreasingly, that is, $A \subseteq B$ implies $\mu(A) \leq \mu(B)$ for any $A, B \in \mathcal{F}$.

Due to the nonadditivity of the fuzzy measure, the global influence of features to the decision is not just the sum of their individual contributions.

Definition 4. Let μ be the fuzzy measure defined on X and f be the non-negative real-valued measurable function defined on X , then the Choquet integral of f with respect to μ

$$(c) \int f d\mu = \int_0^\infty \mu(\{x | f(x) > \alpha\}) d\alpha. \quad (1)$$

Among them, the part on the right side of the Equation (1) is the integral of a function of α with respect to the Lebesgue measure.

If X is a finite set, the function f is a discrete-valued function, and the function values is $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$,

Table 1
Scores of tourist attractions

	x_1	x_2	x_3	average score	Fuzzy integral
Attraction A	6	9	9	8.0	7.5
Attraction B	7	8	9	8.0	7.7
Attraction C	8	8	8	8.0	8.0

Table 2
Fuzzy measure

Fuzzy measure	values
$\mu(\{x_1\})$	0.2
$\mu(\{x_2\})$	0.2
$\mu(\{x_1, x_2\})$	0.5
$\mu(\{x_3\})$	0.2
$\mu(\{x_1, x_3\})$	0.5
$\mu(\{x_2, x_3\})$	0.5
$\mu(\{x_1, x_2, x_3\})$	1.0

without loss of generality, assuming $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_n$, then

$$\begin{aligned} (c) \int f d\mu &= \int_0^\infty \mu(\{x | f(x) > \alpha\}) d\alpha \\ &= \int_0^{\alpha_1} \mu(\{x | f(x) \geq \alpha_1\}) d\alpha + \int_{\alpha_1}^{\alpha_2} \mu(\{x | f(x) \geq \alpha_2\}) d\alpha \\ &\quad + \dots + \int_{\alpha_{n-1}}^{\alpha_n} \mu(\{x | f(x) \geq \alpha_n\}) d\alpha \end{aligned} \quad (2)$$

Here, each integration interval in Equation (2) is left closed and right open, then fuzzy integral of f with respect to μ defined as

$$(c) \int f d\mu = \sum_{i=1}^n (\alpha_i - \alpha_{i-1}) \mu(A_i), \quad (3)$$

in which $\alpha_0 = 0, A_i = \{x_i, x_{i+1}, \dots, x_n\}$.

2.2. An example of fuzzy integral

For illustrating Choquet integral, we take an example. A company plans to choose one of the three scenic spots A, B, and C as the scenic spot for this trip. The travelers participating in the scoring include 3 people. We record the scores of the travelers as feature values which are represented by x_1, x_2 , and x_3 , respectively. Table 1 lists these scoring records. The fifth and sixth columns are the calculated average score and Choquet integral. The fuzzy measure of Choquet integral is shown in Table 2.

Table 2 shows that the fuzzy measures of a single traveler, between two travelers, and between three travelers in the travel problem are 0.2, 0.5, and 1, respectively. The score of each traveler is regarded

as a feature value. The fuzzy measure between the features greater than the sum of the fuzzy measure of the features, for example:

$$\begin{aligned} \mu(x_2, x_3) &= 0.5 \\ \mu(x_2) + \mu(x_3) &= 0.2 + 0.2 = 0.4 \\ \mu(x_2, x_3) &> \mu(x_2) + \mu(x_3) \end{aligned}$$

We sort the values for combined features from small to large. Since the values in Table 2 happen to be $f(x_1) < f(x_2) < f(x_3)$, the conditions are satisfied. So the Choquet integral of each scenic spot can be calculated according to Equation (3).

The Choquet integral value of scenic spot A are calculated as follows

$$\begin{aligned} (c) \int f d\mu &= \mu(x_1, x_2, x_3)(f(x_3) - 0) \\ &+ \mu(x_2, x_3)(f(x_2) - f(x_1)) \\ &+ \mu(x_2, x_3)(f(x_3) - f(x_2)) \\ &= 1.0 * 6 + 0.5 * 3 + 0.5 * 0 = 7.5 \end{aligned}$$

The Choquet integral value for scenic spot B are calculated as follows

$$\begin{aligned} (c) \int f d\mu &= \mu(x_1, x_2, x_3)(f(x_3) - 0) \\ &+ \mu(x_2, x_3)(f(x_2) - f(x_1)) \\ &+ \mu(x_1)(f(x_3) - f(x_2)) \\ &= 7 * 1.0 + 0.5 * 1 + 0.2 * 1.0 = 7.7 \end{aligned}$$

The Choquet integral value for scenic spot C are calculated as follows

$$\begin{aligned} (c) \int f d\mu &= \mu(x_1, x_2, x_3)(f(x_3) - 0) \\ &+ \mu(x_1, x_2, x_3)(f(x_2) - f(x_1)) \\ &+ \mu(x_1, x_2, x_3)(f(x_3) - f(x_2)) \\ &= 1.0 * 8 + 1.0 * 0 + 1.0 * 0 = 8 \end{aligned}$$

The calculation results of Choquet integral show that scenic spot C has the highest score followed by scenic spot B. Scenic spot A has the lowest score. The calculation result is different from the average score and classical integrals. Fuzzy integrals have non-additive property. In this example, the fuzzy integral values take into account the interaction between travelers so that the attraction which has everyone's high scores has the final higher score.

2.3. Generalized fuzzy integral

This section gives the basic definition of Generalized fuzzy integral used in our algorithm [9].

Definition 3. Given a set function $\mu: \mathcal{F} \rightarrow [0, +\infty)$, Satisfy $\mu(\emptyset) = 0, A \in \mathcal{F}$, function $f: X \rightarrow [0, +\infty)$, the integral of f on A respect to μ is represented by the symbol $(w) \int f d\mu$, defined as follows:

$$\begin{aligned} (w) \int_A f d\mu &= \sup \left\{ \sum_{j=1}^k \lambda_j \mu(E_j) \mid f \geq \sum_{j=1}^k \lambda_j \chi_{E_j}, \right. \\ &\left. k \geq 0, E_j \subseteq \mathcal{F} \cap A, \lambda_j \geq 0, j = 1, 2, \dots, k \right\}, \end{aligned} \tag{4}$$

in which $\mathcal{F} \cap A = \{E \cap A \mid E \in \mathcal{F}\}$, χ represents the characteristic function, and \mathcal{F} is the set class on X .

The Equation (4) in definition 3 can be obtained in the following equivalent form:

$$\begin{aligned} (w) \int_A f d\mu &= \sup \left\{ \sum_{j=1}^k \lambda_j \mu(E_j \cap A) \mid f \geq \sum_{j=1}^k \lambda_j \chi_{E_j \cap A}, \right. \\ &\left. E_j \cap A, k \geq 0, E_j \subseteq \mathcal{F}, \lambda_j \geq 0, j = 1, 2, \dots, k \right\}. \end{aligned} \tag{5}$$

When μ is a classical measure and f is measurable, generalized fuzzy integral and Lebesgue integral are consistent. But f is not necessarily measurable. Generalized fuzzy integral is an extension of Lebesgue integral.

Let X be the feature set, and $X = \{x_1, x_2, \dots, x_n\}$ is a finite set, and its power set $P(X)$ is \mathcal{F} . At this time, any function defined on X , is measurable, and since all singleton set are contained in $P(X)$, all supremums must be in the equation $f = \sum_{j=1}^k \lambda_j \chi_{E_j}$ is established, so the expression of integral can be simplified based on Equation (4) as follows:

$$\begin{aligned} (w) \int_A f d\mu &= \max \left\{ \sum_{j=1}^{2^n-1} \lambda_j \mu(E_j \cap A) \mid f \right. \\ &= \left. \sum_{j=1}^{2^n-1} \lambda_j \chi_{E_j \cap A}, \lambda_j \geq 0 \right\}, \end{aligned} \tag{6}$$

which $E_j = \{x_i | \frac{x_i}{2^j} - \frac{i}{2^j} \geq \frac{1}{2}, 1 \leq i \leq n\} \subseteq X, j = 1, 2, \dots, 2^n - 1$.

Generalized fuzzy integral regards the classifier as a multiple-input single-output system, transitioning from the classical integral of the weighted average of features to the integral of features and the interaction between features. Ref. [16] uses the lasso regression method to solve the fuzzy measure in Equation (6) and concludes the linear function of the Generalized fuzzy integral:

$$\int f du = \sum_{j=1}^{2^n-1} z_j u_j, \tag{7}$$

in which z_j is the coefficient of the equation, and u_j is the fuzzy measure to be solved. The coefficient can be defined as follows.

$$z_j = \begin{cases} \max(\frac{1}{2})^{n+1} & \max_{i \in E_j} f(x_i) - \max_{i \in E_{j-1}} f(x_i), \text{ if } > 0 \text{ or} \\ 0, & \text{otherwise} \end{cases} \quad j = 2^n - 1, \tag{8}$$

for $j = 1, 2, \dots, 2^n - 1$ with a convention that the maximum on the empty set is zero. Here, $\text{frc}(\frac{1}{2})$ denotes the fractional part of $\frac{1}{2}$. In Equation (8), if we express j in the binary form $k_n k_{n-1} \dots k_1$, then $\{i | \text{frc}(\frac{1}{2}) \in [\frac{i}{2^j}, 1)\} = \{i | k_i = 1\}$ and $\{i | \text{frc}(\frac{1}{2}) \in [0, \frac{i}{2^j})\} = \{i | k_i = 0\}$. Assumed that a dataset contains 3 features, $f(x_i)$ is used to represent the feature value, $i = 1, 2, 3$. The value of z_j is determined according to Equation (8) by the following function:

$$z_1 = \min\{f(x_1)\} - \max\{f(x_2), f(x_3)\}$$

$$z_2 = \min\{f(x_2)\} - \max\{f(x_1), f(x_3)\}$$

$$z_3 = \min\{f(x_1), f(x_2)\} - \max\{f(x_3)\}$$

$$z_4 = \min\{f(x_3)\} - \max\{f(x_1), f(x_2)\}$$

$$z_5 = \min\{f(x_1), f(x_3)\} - \max\{f(x_2)\}$$

$$z_6 = \min\{f(x_2), f(x_3)\} - \max\{f(x_1)\}$$

$$z_7 = \min\{f(x_1), f(x_2), f(x_3)\} - \max\{0\}$$

The above functions can also be seen as consisting of two $2^n - 1$ dimensional vectors, τ and

Table 3
Complexity for computing fuzzy measure

	Number of features					
	10	11	12	13	14	15
Number of Fuzzy measure	1023	2047	4095	8191	16383	32767

μ , where $z_j = 0$ when $z_j < 0$. When $(f(x_1), f(x_2), f(x_3)) = (60, 40, 10)$, z can be calculated as $(20, 0, 30, 0, 0, 0, 0, 10)$. When the number of features increases, the zero values in the vector z become more and more.

The calculation of the fuzzy measure needs to solve the function value of all feature combinations. Table 3 shows the relationship between the number of features and the number of fuzzy measure coefficients. When the number of features is 10, the number of equation coefficients is 1023. When the number of features is 11, the number of equation coefficients is 2047. With the increase of features, the number of equation coefficients increases exponentially. Thus, the time complexity of this algorithm is $O(2^n)$, so as the spatial complexity. The time complexity of $O(2^n)$ is very huge. When the number of features is 10 or more, operations such as feature extraction are generally performed first, and then a fuzzy integral classification model is established.

3. Fuzzy integral based on parallel computing and sparse storage

Due to the high complexity of solving fuzzy measure, an efficient method need to be explored for dealing with computing problem. A direct way is to reform fuzzy integral model to parallelized form. In the meantime, memory storage is another limit condition for fuzzy measure. Aiming to efficient access, sparse storage strategy will be introduced.

3.1. Parallel computing

There are two main decomposition modes for parallel computing, i.e. task decomposition and data decomposition. Task decomposition refers to divide calculations into a set of independent tasks. Multiple threads can perform these tasks arbitrarily. Data decomposition means that a large dataset needs to be processed and independent calculations can be performed on each record in the dataset. Supposed that there are 1000 records in a dataset with 10 features, the dataset is divided into 4 blocks by data decom-

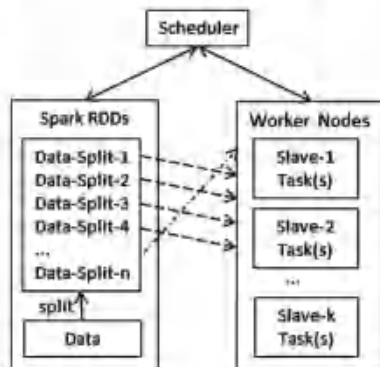


Fig. 1. Parallel computing architecture diagram.

position. All tasks can be executed at the same time, and each task processes 250 records separately. If task decomposition is used, the task of a record is decomposed into multiple tasks for parallel processing. In the calculation of fuzzy measure, the coefficient of fuzzy measure must be calculated for each record in the dataset. In the whole calculation process, the calculation of each record is independent, so the dataset can be divided into multiple blocks through data decomposition. Parallel calculation can be used to reduce the calculation time of the program.

Parallel computing frameworks mainly include MPI, Hadoop's MapReduce and Spark. MPI supports Fortran, C and C++ languages. Programs can be compiled to generate class libraries. However, compared with object-oriented languages such as Java, C and C++ have weak data processing capabilities. Compared with MPI, Spark has high fault tolerance. Once a task fails, Spark will automatically recalculate the task instead of recalculating the entire dataset. Hadoop MapReduce runs faster and supports more programming languages than Spark.

Figure 1 shows the parallel programming architecture diagram of data partitioning in Spark. The left side represents the data block. A dataset is divided into n data blocks. Each data block is considered as a computing task. All tasks enter a queue to be processed and are dispatched to the computing node by the scheduler. The default scheduling method of the Spark scheduler is FIFO (First In First Out). The scheduler assigns tasks to all available computing units in a first-in first-out manner. When the first computing unit calculates the assigned tasks and becomes

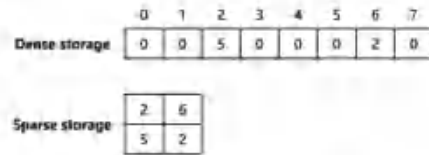


Fig. 2. Sparse storage and dense storage (Dense storage needs 8 bits and sparse storage just needs 4 bits).

an idle unit, the scheduler takes out new tasks from the queue and assigns to the idle units until all tasks are completed. Generally, the number of data blocks is greater than the number of computing units. A working node has multiple computing units, and the number of computing units is less than or equal to the number of CPU cores.

3.2. Sparse storage

Sparse storage means that only non-zero values are stored and zero values are not stored by default. Sparse storage is mainly used to store the data with a lot of zero-values. In the calculation of the fuzzy measure, the coefficient vector z of the Equation (3) has plenty of zero values. When using dense storage, the storage space increases exponentially as the number of features increases. Only non-zero values need to be stored with sparse storage, which can save a lot of storage space.

We take a vector (0, 0, 5, 0, 0, 0, 2, 0) as an example to illustrate the advantages of sparse storage as Fig. 2. This vector contains 8 feature values, including 6 zero values and 2 non-zero values. Sparse storage only records these two non-zero values or values that have practical meaning, and does not record zero values. Therefore, sparse storage uses two arrays (2, 6) and (5, 2) to represent the original vector, which means starting from the 0th bit, the 20th and 60th bits are 5 and 2, respectively, and the rest are 0. It can be seen that sparse storage and dense storage represent the same content, but its meaning has not been changed. When the feature vector contains many zeros, sparse storage can save a lot of space.

3.3. Sparse calculation of fuzzy measures

For describing classification problem, $X = \{x_1, x_2, \dots, x_n\}$ represents its features, Y represents category, l represents the number of records in the dataset, and n represents the number of fea-

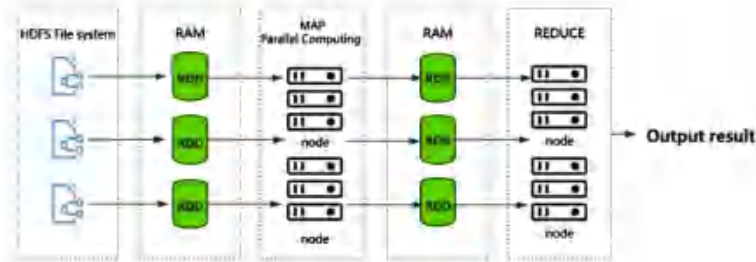


Fig. 3. The principle of parallelized algorithm.

tures. Use $\{f_j(x_1), f_j(x_2), \dots, f_j(x_n), Y_j\}, j = 1, 2, \dots, l$ to represent the feature values and category of the j^{th} record which is labeled as $line_j$.

Let E_j represent the output of generalized fuzzy integral as follows.

$$E_j = \int f d\mu = \sum_{i=1}^{2^n-1} z_i \mu_i \quad (9)$$

Supposed that the error of the model is denoted by e , the model with minimizing the error e defined in Equation (10) is adopted to solve the optimal fuzzy measure in (9).

$$e = \sum_{j=1}^l \sqrt{\frac{1}{l} \sum_{j=1}^l (E_j - Y_j)} \quad (10)$$

In order to increase the generalization ability and reduce the complexity, the sparse solution obtained by the method will be an important feature of the model at the same time. We add L_1 regularization to the regression model as follows:

$$\text{Min}(e) = \frac{1}{2 * n} \left\| \sum_{i=1}^{2^n-1} z_i \mu_i - Y \right\|_2^2 + \lambda, \quad (11)$$

where λ is called the penalty coefficient.

Spark MLlib comes with a LASSO regression implementation library and supports sparse vectors and sparse learning [34]. LASSO uses L1-Norm regularization which can help feature extraction. Combining L1-Norm regularization to solve the Equation (11), the fuzzy measure can contain a lot of zero values.

3.4. Solving fuzzy measure of the combined features

The time complexity of fuzzy integral is $O(2^n)$, where n is the number of features. For a dataset, it is required to solve fuzzy measure of all the combined features. Firstly, the value of combined feature must be calculated. We tried multiple loops and recursion with Python programming. However, whether it is multiple loops or recursion or not, there exists a stack overflow problem due to the excessive number of features in some dataset. The use of multiple nested loop stacks can alleviate the problem to a certain extent. But this method cannot be applicable to the changes in the number of features. Therefore, a method based on the combination of binary accumulator and sparse storage is proposed to realize the calculation of fuzzy measures of all combined features in this article.

The fuzzy measure solving of the combined feature for each sample is independent. According to the task parallelism of data decomposition, all samples are divided into blocks to achieve parallel computing. Single machine is limited and cannot be expanded indefinitely. The exponential time and space complexity of fuzzy integral makes it difficult to run on a single machine. Spark is a distributed parallel computing framework which can use cheap computers to form clusters and support horizontal expansion of computing nodes. The algorithm is extended to the Spark framework to implement distributed parallel computing of fuzzy integral. The parallel principle of the new algorithm is shown in Fig. 3.

The algorithm for solving the coefficients is given as Algorithm 1. In the calculation process, each record is independent. According to the task of data decomposition in parallel, all records are input into Algorithm 1 in blocks. The program finally outputs

the key array, value array and the label of all records. The binary data in the algorithm stores binary values to implement a binary accumulator which can traverse all the combined numbers more efficiently than multiple loops and recursive algorithms. The key array only stores the index values of the non-zero coefficients, and the value array stores the true values of the non-zero coefficients. The two arrays are used to implement sparse storage of coefficients.

Algorithm 1 combines parallel computing and sparse storage, allowing the fuzzy integral to have a greater optimization in time and memory space occupation. PSFI outputs the values of combined feature. The corresponding fuzzy measure values (denoted as μ) are solved by the optimization method. The values of μ represent the importance of the features and their combined features.

Algorithm 1: Solving equation coefficient

Input: $line_j$. Enter record by record.
 Initialize variables: $n = \text{length}(line_j)$, $n_2 = 2^n - 1$
 Initialize the maximum and minimum values: $max = 0$, $min = -\infty$
 Initialize storage space: $key = \text{new array}()$, $value = \text{new array}()$
 Initialize binary array: $binary = \text{new array}(n) = \{0\}$
 $label = Y_j$
 $line = [f_1(x_1), f_2(x_2), \dots, f_n(x_n)]$
for $i = 0$ **to** n_2 **do**
 $binary = 1$ // Binary accumulator
 $max = 0$, $min = -\infty$
 for $j = 0$ **to** n **do**
 if $binary[j]$ equal 0
 if $max < line[j]$
 $max = line[j]$
 end
 else if $min > line[j]$
 $min = line[j]$
 end
end
if $min - max > 0$:
 $key.insert(i)$
 $value.insert(min - max)$
end
Output: key , $value$ and $label$.

3.5. Construction of classification model based on PSFI

The focus of PSFI is to solve the coefficients of the combined feature, then use a supervised machine learning method to solve the fuzzy integral, and build a classification model for classification. The construction of the entire classification model is divided into two parts. The first part is the classification model

based on the sparseness of features, and the second one is the fusion model based on fuzzy integral.

3.5.1. Feature combination classification model based on Fuzzy integral

The process of constructing a classification model based on Fuzzy integral is listed as follows.

- (1) The dataset is preprocessed to represent the features and deal with the strings and null values.
- (2) The dataset is divided into the training dataset and testing dataset.
- (3) PSFI is used to map the feature space to the combined feature space. The combined feature value is solved.
- (4) The combined feature values of the training dataset are input into the machine learning model. The fuzzy measure values are learned. A fuzzy integral classification model is built.
- (5) The fuzzy integral model is used to classify and predict the combined feature values of the testing dataset. The construction process of the classification model based on PSFI is shown in Fig. 4. The machine learning model can adopt Fisher discriminant model, logistic regression model or other available classification models.

3.5.2. Fusion model based on fuzzy integral

Fusion model can be constructed, which is divided into three parts similar to the process of classification model. The construction process of the fusion model is shown in Fig. 5.

The dataset is preprocessed.

The dataset is divided to n training datasets ($n \geq 2$), one verification dataset and one testing dataset.

The third step includes four sub-steps.

- 1) Training n machine learning models on n training datasets;
- 2) Using these n models to predict the data in the validation set, and inputting the prediction results into PSFI;
- 3) Using PSFI to solve the combined model feature value between the models;
- 4) Using the machine learning method to learn the fuzzy measure of the combined model feature to build the fusion model.
- 5) A similar method is used to classify and predict the testing dataset. n models are firstly used to predict the classification value of the test data,

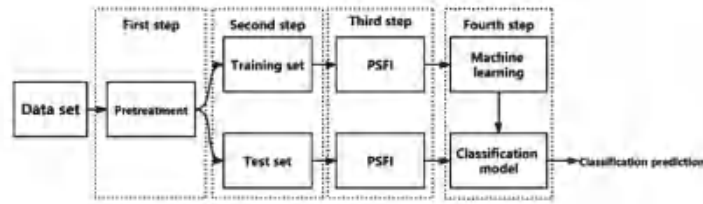


Fig. 4. Classification model based on PSFI.

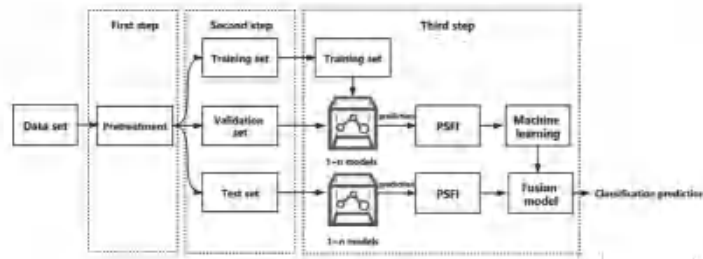


Fig. 5. Fuzzy integral fusion model Experiments and analysis.

and the prediction results are input into PSFI. Then, the fusion model is used to predict the final classification.

4. Experiments and analysis

In this work, we carried out two parts of experiments to test the impact of parallel computing and sparse storage on model performance and the application of PSFI to classification problems on high-dimensional datasets.

4.1. Impact testing on PSFI

Four datasets selected from UCI are used in the experiments, namely Shuttle, Poker, Letter and Mushroom, which are with a large number of data records and features and suitable for parallel computing [35]. Table 4 describes the number of records, features and categories of datasets.

The experiment was implemented in two kinds of configuration environments.

Environment A installs Spark and Hadoop clusters on the basis of E3-1240 CPU and OpenVZ virtualization [36], with one master node as the scheduling end and two slave nodes as the computing nodes. The performance of its computing nodes is different.

Table 4 Descriptions for datasets

Datasets	Records	Features	Categories
Shuttle	43500	9	7
Poker	1000000	10	10
Letter	20000	16	26
Mushroom	8124	22	2

Environment B uses L5630 CPU and XenServer virtualization. Like environment A, it has one master node and two slave nodes, but the performance of its slave nodes are the same or very close. Table 5 shows the server configuration of two environments and Table 6 shows the performance of slave nodes,

4.1.1. Cython's impact on PSFI

"A workman must sharpen his tools if he is to do his work well".

Cython is a programming language which can convert Python language code into C language code and compile it into an efficient executable class library [37]. In the Cython language environment, C language variables can be defined, and Python libraries can be used. Cython compiler converts Cython code into C language code, compiles and generates executable class library files. We modified Algorithm 1 by using C language type to define variables and using Numpy array to optimize the binary accumulator. The

Table 5
Server configuration

Configuration items	Environment A	Environment B
Virtualization	OpenVZ	XenServer
OS	CentOS 6.8 x86_64	CentOS 6.8 x86_64
Master node	Intel Xeon E3-1240 @ 3.30GHz; Quad core CPU+2 GB RAM	Intel Xeon L5630 @ 2.13GHz; Triple core CPU+2 GB RAM
Slave node (2)	Intel Xeon E3-1240 @ 3.30GHz; Quad core CPU+2 GB RAM	Intel Xeon L5630 @ 2.13GHz; Triple core CPU+2 GB RAM
Hadoop version	Hadoop-2.6.4	Hadoop-2.6.4
Spark version	Spark-1.6.2	Spark-1.6.2
Slave-node performance difference	Big	Close to zero

Table 6
Node performance test (unit: second)

Test items	Environment A			Environment B
	node1	node2	Promote rate	Cluster node
50 million $(1+1)$ operations by Python	3.197	2.129	0.334	4.334
20 million $\sqrt{22}$ operations by Python	4.226	2.377	0.438	5.957
50 million $\sum_{i=1}^{10} i$ operations by Cython	0.026	0.017	0.354	0.046
5 million $\sqrt{1, \dots, \sqrt{10}}$ operations by Cython	5.255	3.299	0.372	6.217

following experiments verify the efficiency improvement of the fuzzy integral optimized by Cython language compared to that with Python language. The following experiment compares the running time of Cython with that of Python in calculating the values of combined features of the first 5000 data in Poker dataset.

Table 6 shows that node 2 has 33.4% and 43.8% improvements in Python based integer and floating-point arithmetic than node 1 respectively. For integer and floating-point operation in Cython, running time on node 2 has 35.4% and 47.2% improvements than that on node 1 respectively. Cython has faster computing efficiency on node1, node2 and cluster node than Python.

It can be seen from Table 7 that the efficiency of code optimized by Cython is increased by 44 times.

Table 7
Comparison of PSFI efficiency under Cython

	Python	Cython	Efficiency ratio
Runtime (s)	14.723	0.334	44.08

Table 8
Comparison of storage space of fuzzy integral applied to different datasets

Dataset	Feature	Record	Dense storage	Sparse storage	Efficiency ratio
Shuttle	9	14500	28.42 MB	680.25 KB	33.06
Poker	10	25010	97.79 MB	1.64 MB	59.63
Letter	16	80	20 MB	7.3 KB	2805.48
Mushroom	22	4	64 MB	656 B	102300.1

The operating efficiency of Cython code is much higher than that of Python code in PSFI. It is a big improvement.

4.1.2. Impact test of sparse storage on PSFI

For verifying the algorithm proposed in this article, we compared the space occupied by the fuzzy integral in sparse storage and dense storage, and compared the running time for training the lasso regression model on Spark. Experiments show that sparse storage improves the time and space efficiency of the fuzzy integral.

In fuzzy integral calculation, the dimension of the combined feature increases exponentially and fuzzy measures contain more zero values as the number of features increases. Table 8 lists the comparison of storage space between sparse storage and dense storage for the fuzzy integral. Experimental results show that sparse storage reduces a lot of storage space compared to dense storage. Sparse storage can reduce storage space by tens of times for a dataset with a small number of features. Mushroom dataset has 22 features and only 4 records. Dense storage for mushroom requires 64 MB of space, while sparse storage only takes up 656 B of space. The storage space is saved 102,300 times. The storage space for Shuttle using sparse storage can be saved 33 times. In the Spark framework, the storage space will be loaded into the memory-based RDD and reused, so saving storage space also means saving the memory required for calculation.

Sparse storage for the fuzzy integral reduces the storage space very significantly. As the number of features increases, the storage space increases exponentially. If dense storage is used, whether it is memory or disk, the space overhead is very large,

Table 9
Comparison of training time of Shuttle dataset under different storage models

	Dense storage	Sparse storage
Training time(s)	2.282	0.631

Table 10
Comparison of the number of cores used in nodes with different number of parallel tasks

Number of parallel tasks	1	2	3	4	5	6	7	8
Number of cores in node 1	1	1	2	2	3	3	4	4
Number of cores in node 2	0	1	1	2	2	3	3	4

which often leads to insufficient storage space and program interruption. Therefore, the application of sparse storage on fuzzy integral is very important. Sparse storage in the calculation of fuzzy integrals can not only reduce storage space, but also speed up model training. Table 9 shows the comparison of the running time of the Shuttle dataset in sparse storage and dense storage. The training model for sparse storage takes 0.631 seconds, which is 3.6 times faster than dense storage. Therefore, sparse storage in storage space and model training time significantly improves the performance of PSFI.

4.1.3. Impact text of parallel computing on PSFI

In this section, we compared the running time of the Spark-based PSFI with different computing resources, and explored the impact of nodes with similar performance and different performance on Spark task scheduling. Experiments (1)-(2) use environment A and experiment (3) uses environment B. Environment A is a cluster composed of two 4-core nodes, with a total of 8 computing cores. Table 10 lists the relationship between the number of parallel tasks in the cluster and the number of cores used by slave nodes. When the cluster is configured with 3 parallel tasks, node 1 contributes 2 cores and node 2 contributes 1 core. When the total number of parallel tasks is an odd number, node 1 contributes 1 more core than node 2 contributes.

- (1) In clusters with performance differences in slave nodes, the number of parallel tasks in the cluster is not as high as possible. The completion time of the entire task depends on the completion time of the slowest task.

The running time of PSFI on the Poker is shown in Table 11. The experimental result shows that the running time of data conversion decreases as the number

Table 11
Comparison of running time for Poker dataset

Number of parallel tasks	1	2	3	4
Time (s)	120.045	52.499	38.518	28.937
Total time (s)	120.045	104.998	115.554	115.748

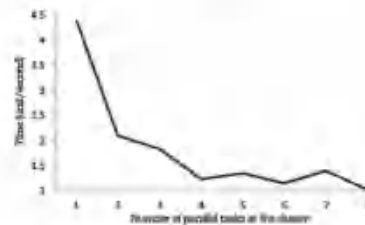


Fig. 6. Comparison of the running time with different numbers of parallel tasks for Shuttle.

of parallel tasks in the cluster increases. However, the running time of two cores is more than that of one core. In theory, it is unreasonable. Compared with a single core, the two cores will have an additional cost of inter-process communication for program execution. Therefore, the running time of the two cores should be nearly two times faster than the running time of a single core, but less than two times. This shows that there are differences between the two nodes used in the experimental cluster. When there are two CPU cores, the CPU core of node 2 is added to the cluster, which improves the computing performance of the cluster. Comparing the number of parallel tasks 3 and 4, it can also be found that the addition of node 2 improves the computing performance of the cluster, so node 2 has better performance than node 1, and the running time of node 2 to execute a single task will be shorter than that of node 1. The node difference between clusters will have a certain impact on the running time of the program. If there are nodes with poor performance in the cluster, this poor node may become the bottleneck of cluster computing. When the last task happens to be executed by the poor node, the completion time of the entire computing task will be based on the completion time of the last task of the worse node.

Figure 6 shows the calculating time of PSFI for the Shuttle. It proves that the running time of computing tasks with even number decreases faster. The reason is that when the number of parallel tasks is odd, the number of cores in node 1 is more 1 than that of node 2 in the cluster. And the computing power of node

Table 12
The time, speedup and efficiency of parallel computing in environment A

Number of parallel tasks	Poker			Letter			Mushroom		
	Running time(s)	Speedup ratio	Efficiency	Running time(s)	Speedup ratio	Efficiency	Running time(s)	Speedup ratio	Efficiency
1	81.62	/	/	215.26	/	/	6364.70	/	/
2	52.50	1.55	0.78	114.11	1.89	0.94	3573.25	1.78	0.89
4	28.94	2.82	0.71	61.00	3.53	0.88	1953.01	3.26	0.81
6	23.08	3.54	0.59	44.50	4.84	0.81	1307.39	4.65	0.78
8	18.96	4.31	0.54	34.88	6.17	0.77	1041.35	6.11	0.76

1 is weaker than that of node 2. When the number of parallel tasks in the cluster is equal to 5 or 7, the running time of computing tasks increases instead. It shows that the number of parallel tasks in the cluster is not as high as possible. When the dataset is small, increasing the number of parallel tasks may not necessarily reduce the running time. The greater the number of tasks is, the greater the communication overhead of the process is. In the meanwhile, the computing tasks will also be affected by nodes with weaker performance.

When the processing overhead or task waiting time is greater than the efficiency brought by parallel computing, the running time increases instead. When the number of parallel tasks is equal to 8, the running time is the shortest, because the last batch of tasks happens to be completed by node 2 in parallel execution. If the last batch of tasks happens to be executed in parallel by node 1 and node 2 and is nearly completed at the same time, then the running time will be faster because it makes better use of the computing resources of node 1. Therefore, for small and medium-sized datasets, the number of cores and tasks should be selected reasonably, especially considering the time for the slowest node to complete tasks, so as to improve the overall operating efficiency of the cluster and save cluster resources.

- (2) In general, increasing the number of parallel tasks will reduce the overall running time of the parallel program, increase the speedup ratio and decrease the efficiency.

Let T_s represent the running time of serial computing, T_p represent the running time of parallel computing, N indicate the number of CPUs, S be speedup ratio, and E indicate the efficiency of parallel programs. The speedup ratio of parallel program can be calculated as:

$$S = \frac{T_s}{T_p}$$

The efficiency of parallel program can be calculated as:

$$E = \frac{S}{N}$$

In environment A, PSFI is applied on the three datasets of Poker, Letter and Mushroom with 1, 2, 4, 6, and 8 parallel tasks respectively. The running time, speedup and efficiency of the experiment are shown in Table 12, where the number of 1 parallel task represents the running time of the serial calculation in node 2 with better performance.

Table 12 shows that the calculation time of fuzzy integrals decreases almost linearly with the increase computing resources. When the number of parallel tasks reaches 6 or 8, the running time decreases slowly. The reason is that the communication cost among processes increases and the resource usage reaches the peak of the cluster, which has a certain impact on the program operation. The greater the number of parallel tasks in the cluster is, the more the process cost is. The larger the amount of data is, the greater the process cost is.

The speedup ratio of parallel computing increases and the efficiency decreases as the number of parallel tasks increases. The efficiency of parallel computing decreases because of the increased cost of system resources. But for large datasets, the reduction in running time of parallel tasks is much higher than the time spent in process overhead. When the amount of data increases, we can increase the number of nodes in the cluster to improve task parallelism. PSFI makes the calculation time of the fuzzy integral to be decreased linearly with the increase of the degree of parallelism. It has a higher efficiency when the appropriate combined features are extracted.

- (3) In cluster with similar slave node performance, the efficiency loss of parallel computing is reduced.

Table 13
Experimental results of Poker and Letter in environment B

Number of parallel tasks	Poker			Letter		
	Computing time	Speedup ratio	efficiency	Computing time	Speedup ratio	efficiency
1	780.744	1.000	1.000	166.580	1.000	1.000
2	96.685	1.869	0.935	86.711	1.921	0.961
3	66.762	2.707	0.902	60.215	2.766	0.922
4	51.778	3.493	0.873	46.172	3.608	0.902

This experiment adopts environment B which uses a low-frequency CPU and 6 computing cores with similar node performance. It has a small impact on the speedup and efficiency of parallel computing. The experiment mainly explores the operating efficiency of PSFI in cluster computing the values of combined features with different numbers of parallel tasks. Before the experiment, the data is pre-processed to convert those character data to numerical type. The Poker dataset itself is numerical data and does not need to be processed. For the Letter dataset, 26 letters need to be processed in advance, so the running time is shorter than the previous experiment. PSFI is applied to the Poker and Letter datasets with different parallel tasks respectively shown as Table 13.

Experimental results show that the speedup and efficiency drop of parallel computing is relatively stable when testing on clusters with similar node performance. The speedup of the parallel program with 4 parallel tasks equals to 4 in the ideal state which have dropped by 0.51 and 0.39 for the Poker and Letter datasets respectively. The efficiency was maintained as 0.9 as dealing with 3 and 4 tasks in parallel for Poker and Letter respectively. The loss of efficiency is consumed in the communication cost of the system and processes. When the number of parallel tasks is 2, the number of the used nodes is 2. The cost and the efficiency loss of the parallel system are great. The Poker and Letter datasets lost efficiencies of 0.065 and 0.039 respectively when the two tasks are in parallel. At this time, the idle resources of the system should be used as much as possible for calculation. Due to insufficient cluster resources, experiments with higher parallelism cannot be carried out. Generally speaking, the greater the number of cluster cores is, the greater the task parallelism setting is, and the more obvious the efficiency improvement of the algorithm is. In terms of the efficiency of parallel computing, the system's network and hard disk can be optimized. Cluster scheduling parameters can be adjusted to reduce the cost of the system and processes, so that more CPU resources can be used for

calculations and efficiency loss of the parallel can be reduced.

The experiment of Ref. [30] shows that the speedup of dataset with 512 KB size is higher than that of dataset with 128 KB size. For clustering algorithms, the larger the dataset is, the more intensive calculations in per iteration are, higher the speedup is. PSFI and the clustering algorithm mentioned in Ref. [30] are implemented on Spark. The speedup and efficiency are very high. It is because they are computationally intensive algorithms. The more intensive the computation of algorithms is, the higher the running efficiency of parallel programs is. The Letter dataset has 16 features, so each data is calculated 2^{16} times. The Poker dataset has 10 features, so each data is calculated 2^{10} times. Relatively speaking, if the times of calculations are constant, then the times of data exchanges for Letter dataset is less. Therefore, parallel computing has important applications in the calculation of fuzzy integrals or other intensive algorithms. When the amount of data or calculations reaches the situation that a single machine cannot handle, the efficiency improvement of parallel computing is obvious, and the system overhead only accounts for a small part of the total resource usage.

4.2. Application of PSFI in classification and regression problems

This section includes three experiments. Firstly, PSFI will be compared with Multiple nonlinear integral (MNI) [38] on several small datasets. Secondly, the influence of PSFI on logistic regression will be verified. Finally, the application of PSFI combined with Xgboost model [39] will be explored on imbalanced big data classification.

Ten datasets of UCI are selected for this part of experiments. The first 3 datasets are Poker, Letter and Mushroom, which have been used in previous experiments. The last seven datasets are Monk1, Monk2, Monk5, Heart, Pima, Wdbc, and Breast, which are small datasets. Monk1-3 datasets are also called monk datasets. They are two-category datasets

Table 14
Databases Descriptions

Datasets	Samples	Features	Categories
Poker	1000000	10	10
Letter	20000	16	26
Mushroom	8124	22	2
Monk1	556	6	2
Monk2	601	6	2
Monk3	554	6	2
Heart	270	13	2
Pima	768	8	2
Wdbc	569	30	2
Breast	699	9	2

Table 15
Experimental environment configuration

Items	Configurations
Virtualization	Xenserver
Operating System	CentOS 6.8 x86_64
Master node	Intel Xeon E5-2680 v2 @ 2.80 GHz; 8 Core CPU+8 GB RAM
Slave node	No slave node, the computing node is on the master node
Spark version	spark-1.6.2
Slave node performance difference	Does not exist, the computing node is also on the master node

which contain 6 categorical features. The difference is that the Monk3 dataset contains 5% noise data. Heart is a heart disease dataset with 13 features, including numerical features and categorical features. Pima is a diabetes dataset, which contains 8 features, and the feature types are mainly numeric. Currently UCI website does not provide download of Pima dataset, we can provide corresponding testing dataset with this article. Wdbc contains 30 features, and its categories are malignant and benign of cancer. Breast is also a disease dataset. It has 9 features, including numerical features and nominal features. Table 14 describes the number of records, features, and categories of each dataset.

This experiment uses the Xenserver virtualization environment which is a stand-alone virtual server with 8-core CPU and 8G memory. The node does not install the HDFS file system. The node configuration is shown in Table 15. Here, Spark tasks are running on the same machine and the computing efficiency of each task is similar. There is no need for cross-machine communication between processes. Table 16 shows the performance testing data of the experimental environment.

4.2.1. Comparison of PSFI and MNI

In this section of the experiment, 7 datasets with a small amount of data such as Monk1 are used, and

Table 16
Experimental environment performance test (unit: second)

Testing items	Performance of Cluster node
50 million $(1+1)$ operations by Python	2.147
20 million $\sqrt{22}$ operations by Python	3.098
50 million $\sum_{i=1}^n i$ operations by Cython	0.022
5 million $\sqrt{1}, \dots, \sqrt{10}$ operations by Cython	3.481

Table 17
Number of the selected features for each dataset

Datasets	T number of the selected features
Monk1	6
Monk2	6
Monk3	6
Heart	7
Pima	7
Wdbc	18
Breast	4

certain feature selection processing is performed on some of the datasets. The number of selected features is shown in Table 17. The experiment uses ten-fold cross-validation, that is, the data is divided into 10 parts, one of them is taken as the testing dataset each time, and the remaining 9 parts are used as the training dataset. Repeat 10 times to ensure that all data can be used in the testing dataset, and take these 10 times. The average of the experimental results is used as the final accuracy rate. The experiment uses PSFI to solve the feature combination, uses the lasso regression model to solve the fuzzy measure, and finally uses the Fisher discriminant to classify and predict the data. The comparison of the experimental results of PSFI and the Multiple nonlinear integral (MNI) is shown in Table 18, in which the values in Bold means the better results.

The experimental results in Table 18 show that the testing accuracy of the Monks1, Monks2, Heart and Pima datasets of PSFI has been greatly improved, especially the Monks2 and Heart datasets. The prediction effects of other datasets are relatively close. In general, PSFI has certain advantages in prediction accuracy.

4.2.2. PSFI impact test on logistic regression

This section uses the logistic regression based on L1/L2 regularization on 9 datasets such as Poker, and then adds PSFI to the model, and compares the experimental results. The experiment divided the 9 datasets

Table 18
Comparison of experimental results of PSFI and MNI algorithms

Algorithms	Index	MNI	PSFI
Monk1	Training accuracy	0.978	0.961
	Testing accuracy	0.918	0.937
Monk2	Training accuracy	0.704	0.988
	Testing accuracy	0.688	0.988
Monk3	Training accuracy	0.966	0.972
	Testing accuracy	0.982	0.976
Heart	Training accuracy	0.867	0.861
	Testing accuracy	0.844	0.867
Pima	Training accuracy	0.782	0.781
	Testing accuracy	0.752	0.764
Wdbc	Training accuracy	0.930	0.940
	Testing accuracy	0.958	0.947
Breast	Training accuracy	0.964	0.957
	Testing accuracy	0.956	0.958
Average	Training accuracy	0.884	0.923
	Testing accuracy	0.871	0.920

into 1 training dataset and 2 testing datasets, among which the training datasets of Poker, Letter and Mushroom accounted for 30%, testing dataset 1 accounted for 20%, and testing dataset 2 accounted for 50%. For the remained datasets, the training dataset accounts for 60% and the testing dataset accounts for 20%. The logistic regression model without PSFI algorithm and the logistic regression model with PSFI are shown in Figs. 7 and 8 respectively.

Tables 19 and 20 show the experimental results of the regularized logistic regression model on 9 datasets. Tables 21 and 22 show the experimental results of the model added to PSFI. The results with bold style in all tables represent the best values with same conditions. The experiment uses different regularization penalty coefficients λ , when $\lambda = 0$, the model is a logistic regression model without regularization.

It can be seen from Table 19 and Table 20 that the regularized logistic regression model achieves higher classification accuracy on the Letter, Breast and Monk2 datasets than the model without regularization. On other datasets, the classification accuracy of the regularized logistic regression model is not lower than that of the model without regularization. Therefore, regularization can improve the prediction accuracy of the model and reduce the risk of model overfitting. It can be seen from the table that the better values of λ are concentrated in 1×10^{-5} and 0.01 . When the value of λ is larger, such as $\lambda = 0.1$, the accuracy of the model decreases. Therefore, λ with too large value may lead to the risk of under-fitting of the model, and the selection of λ value has a greater effect on the accuracy of the model.

The results of Tables 19 and 20 show that in the logistic regression model, the prediction accuracy

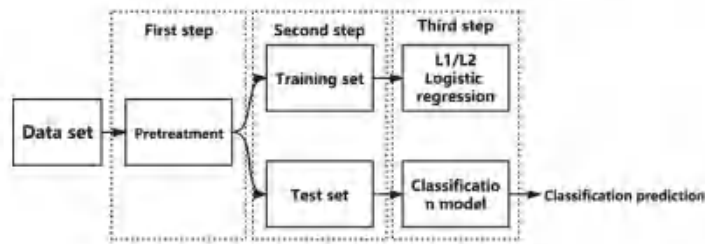


Fig. 7. Regularized logistic regression model.

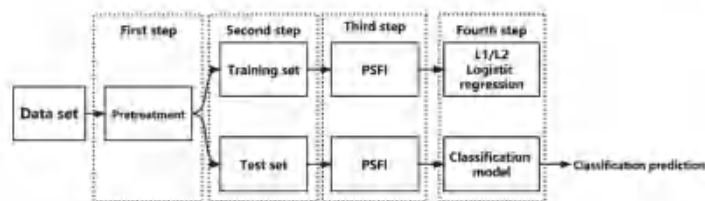


Fig. 8. Logistic regression model optimized by PSFI.

Table 19
Logistic regression algorithm based on L1 regularization

Datasets	Index	λ					
		0	1e-05	1e-04	1e-03	0.01	0.1
Poker	Training accuracy	0.502	0.502	0.502	0.502	0.502	0.422
	Testing accuracy 1	0.501	0.501	0.501	0.501	0.501	0.423
Letter	Training accuracy	0.501	0.501	0.501	0.501	0.501	0.423
	Testing accuracy 2	0.501	0.501	0.501	0.501	0.501	0.423
Mushroom	Training accuracy	0.773	0.768	0.772	0.739	0.575	0.040
	Testing accuracy 1	0.748	0.751	0.748	0.714	0.569	0.034
Heart	Training accuracy	0.741	0.741	0.742	0.733	0.554	0.041
	Testing accuracy 2	0.742	0.742	0.742	0.742	0.942	0.913
Breast	Training accuracy	0.934	0.934	0.934	0.934	0.934	0.906
	Testing accuracy 1	0.935	0.935	0.935	0.935	0.935	0.902
Pima	Training accuracy	0.904	0.904	0.904	0.904	0.897	0.878
	Testing accuracy 2	0.785	0.785	0.785	0.785	0.800	0.785
Diagn	Training accuracy	0.796	0.796	0.796	0.796	0.776	0.735
	Testing accuracy 1	0.704	0.704	0.704	0.704	0.697	0.641
Monk1	Training accuracy	0.685	0.685	0.685	0.678	0.671	0.638
	Testing accuracy 2	0.684	0.684	0.684	0.684	0.668	0.679
Monk2	Training accuracy	0.882	0.882	0.882	0.882	0.882	0.346
	Testing accuracy 1	0.877	0.877	0.877	0.880	0.880	0.300
Monk3	Training accuracy	0.855	0.855	0.855	0.855	0.863	0.589
	Testing accuracy 2	0.750	0.750	0.750	0.750	0.753	0.798
Average	Training accuracy	0.615	0.615	0.615	0.615	0.615	0.635
	Testing accuracy 1	0.638	0.638	0.638	0.638	0.629	0.647
Average	Training accuracy	0.659	0.659	0.659	0.659	0.662	0.662
	Testing accuracy 2	0.632	0.632	0.632	0.632	0.640	0.640
Average	Training accuracy	0.661	0.661	0.661	0.661	0.661	0.661
	Testing accuracy 1	0.798	0.798	0.798	0.798	0.798	0.358
Average	Training accuracy	0.703	0.703	0.703	0.703	0.703	0.649
	Testing accuracy 2	0.777	0.777	0.777	0.777	0.777	0.795
Average	Training accuracy	0.768	0.689	0.768	0.766	0.745	0.606
	Testing accuracy 1	0.719	0.719	0.719	0.718	0.701	0.558
Average	Training accuracy	0.732	0.732	0.732	0.731	0.707	0.586
	Testing accuracy 2	0.732	0.732	0.732	0.731	0.707	0.586

rate of adding L2 regularization is 0.3% and 0.6% higher than that of adding L1 regularization in the Letter and Pima datasets respectively; The prediction accuracy rate of adding L1 regularization is 1.3% and 0.8% higher than that of adding L2 regularization in the Breast and Monk2 datasets respectively. On other datasets, the average prediction accuracy of the two regularization models is the same. Therefore, the two regularization models have their own advantages, but the difference in impact on these datasets is small.

The experimental results in Tables 19 compared with Table 21 and Table 20 compared with Table 22 show that PSFI has a greater impact on the prediction accuracy of the logistic regression model. In the L1 regularized logistic regression model, the prediction accuracy of PSFI is increased by 9.7%, 2.6%, 6.5%, 0.5%, 32.3% in 7 datasets including Poker, Letter, Mushroom, Pima, Monk1-3, etc., 34.1%, and 13%; in the Heart and Breast datasets, they dropped 1.4% and 3.1%, respectively. In the L2 regularized logistic regression model, the prediction accuracy of PSFI is increased by 9.7%, 4.5%, 6.5%, 0.2%, 32.3% in 7 datasets including Poker, Letter, Mushroom, Pima,

Monk1-3, etc., 35.3%, and 13.4%; in the Heart and Breast datasets, they dropped by 1.5% and 1.3%, respectively. When $\lambda=e^{-5}$, the prediction accuracy of the logistic regression model based on PSFI and L1 regularization in the Mushroom dataset reaches 100%. Regularization uses the L2 norm. When the value of λ is e^{-5} , the prediction accuracy of the Mushroom dataset reaches 100%, and when the value of λ is 0.1, the prediction accuracy of the Monk2 dataset reaches 100%. Before PSFI was added, the prediction accuracy of the logistic regression model on the Mushroom and Monk2 datasets was only 93.5% and 66.1%. Experiments show that PSFI improves the prediction accuracy of the model on multiple datasets, and the improvement is large. This shows that PSFI is effective in feature combination, and the combined features it considers improves the classification accuracy of the model.

Table 22 shows that on the Breast dataset, as the value of λ increases, although the accuracy of the training dataset decreases, the accuracy of the testing dataset increases. This shows that combining features may increase the complexity of the model. When the

Table 20
Logistic regression algorithm based on L2 regularization

Datasets	Index	λ					
		0	1e-05	1e-04	1e-03	0.01	0.1
Poker	Training accuracy	0.502	0.502	0.502	0.502	0.502	0.502
	Testing accuracy 1	0.501	0.501	0.501	0.501	0.501	0.501
Letter	Testing accuracy 2	0.501	0.501	0.501	0.501	0.501	0.501
	Training accuracy	0.773	0.772	0.767	0.735	0.676	0.531
Mushroom	Testing accuracy 1	0.748	0.754	0.748	0.717	0.662	0.519
	Testing accuracy 2	0.741	0.741	0.737	0.733	0.646	0.501
Heart	Training accuracy	0.942	0.942	0.942	0.942	0.938	0.914
	Testing accuracy 1	0.934	0.934	0.934	0.934	0.929	0.903
Pima	Testing accuracy 2	0.935	0.935	0.935	0.935	0.930	0.905
	Training accuracy	0.904	0.904	0.904	0.904	0.904	0.897
Breast	Testing accuracy 1	0.785	0.785	0.785	0.785	0.785	0.815
	Testing accuracy 2	0.796	0.796	0.796	0.796	0.796	0.776
Monk1	Training accuracy	0.704	0.704	0.704	0.704	0.701	0.662
	Testing accuracy 1	0.685	0.685	0.685	0.685	0.685	0.678
Monk2	Testing accuracy 2	0.684	0.684	0.684	0.684	0.699	0.668
	Training accuracy	0.882	0.882	0.882	0.882	0.873	0.875
Monk3	Testing accuracy 1	0.877	0.877	0.873	0.873	0.867	0.833
	Testing accuracy 2	0.855	0.855	0.855	0.855	0.863	0.831
Average	Training accuracy	0.750	0.750	0.750	0.750	0.750	0.768
	Testing accuracy 1	0.615	0.615	0.615	0.615	0.615	0.615
Average	Testing accuracy 2	0.638	0.638	0.638	0.638	0.638	0.612
	Training accuracy	0.659	0.659	0.659	0.659	0.662	0.662
Average	Testing accuracy 1	0.632	0.632	0.632	0.632	0.632	0.640
	Testing accuracy 2	0.661	0.661	0.661	0.661	0.661	0.661
Average	Training accuracy	0.798	0.798	0.798	0.798	0.801	0.801
	Testing accuracy 1	0.703	0.703	0.703	0.703	0.703	0.694
Average	Testing accuracy 2	0.777	0.777	0.777	0.777	0.777	0.777
	Training accuracy	0.768	0.768	0.767	0.766	0.756	0.734
Average	Testing accuracy 1	0.719	0.720	0.719	0.718	0.708	0.688
	Testing accuracy 2	0.732	0.732	0.731	0.731	0.722	0.692

λ value is too small, the model will be overfitted. When the λ value is set to the maximum value of 0.1, the model has the highest prediction accuracy. We added the test when $\lambda = 1$. The accuracy of the model in testing dataset 1 and testing dataset 2 is 88.7% and 84.7%, respectively. The difference of the accuracy with the logistic regression model is less than 0.5%. From Tables 20 and 22, we also find that the prediction accuracy of the L2 regularized logistic regression model based on PSFI is higher than that of the L1 regularization model for Letter, Monk2 and Monk3. The performance for other datasets is comparable. This proves that the logistic regression model based on PSFI with L2 regularization has better generalization ability than that with L1 regularization.

Figure 9 shows the histogram of the accuracy of PSFI and the logistic regression model in testing dataset 1 in 9 datasets such as Poker when $\lambda = 0.01$ (0.01 is the default parameter of the logistic regression model). PSFI has a large accuracy improvement on the 6 datasets of Monk1-3, Poker, Letter and Mushroom. There is a small amount of accuracy

improvement on the Pima dataset, and the accuracy rate on the Heart and Breast datasets is slightly reduced. In general, PSFI has greatly improved the accuracy of the model in multiple datasets. Figure 9 shows the average accuracy histogram of PSFI and the logistic regression model in the nine datasets. PSFI has great accuracy for the five groups of λ values. In general, PSFI greatly improves the accuracy of the model when the value of λ is different.

Figure 10 shows the average accuracy histogram of the PSFI algorithm and the logistic regression model in nine data sets. The PSFI algorithm has great accuracy for the five groups of λ values. In general, the PSFI algorithm greatly improves the accuracy of the model when the value of λ is different.

4.2.3. Application of PSFI in unbalanced big data classification

Balanced data means that there is little difference in the number of samples of each category in the dataset. Unbalanced data means that the sample distribution of the category is uneven. Some categories

Table 22
PSFI impact on L2 logistic regression algorithm

Datasets	Index	λ					
		0	1e-05	1e-04	1e-03	0.01	0.1
Poker	Training accuracy	0.603	0.603	0.602	0.600	0.595	0.553
	Testing accuracy 1	0.598	0.597	0.596	0.595	0.590	0.550
	Testing accuracy 2	0.598	0.598	0.597	0.596	0.591	0.551
Letter	Training accuracy	1.000	1.000	1.000	1.000	0.998	0.961
	Testing accuracy 1	0.737	0.785	0.786	0.798	0.797	0.773
	Testing accuracy 2	0.742	0.776	0.782	0.785	0.788	0.761
Mushroom	Training accuracy	1.000	1.000	1.000	1.000	0.995	0.993
	Testing accuracy 1	1.000	1.000	1.000	0.999	0.995	0.990
	Testing accuracy 2	0.999	1.000	1.000	0.998	0.993	0.991
Heart	Training accuracy	1.000	1.000	1.000	0.994	0.974	0.942
	Testing accuracy 1	0.785	0.769	0.754	0.769	0.785	0.800
	Testing accuracy 2	0.776	0.755	0.735	0.776	0.755	0.735
Pima	Training accuracy	0.748	0.748	0.748	0.752	0.759	0.755
	Testing accuracy 1	0.671	0.671	0.671	0.671	0.678	0.651
	Testing accuracy 2	0.701	0.701	0.701	0.706	0.701	0.690
Breast	Training accuracy	1.000	1.000	1.000	1.000	0.993	0.986
	Testing accuracy 1	0.833	0.847	0.840	0.847	0.860	0.873
	Testing accuracy 2	0.806	0.806	0.806	0.823	0.831	0.831
Monk1	Training accuracy	0.967	0.967	0.967	0.964	0.964	0.964
	Testing accuracy 1	0.933	0.933	0.933	0.952	0.952	0.933
	Testing accuracy 2	0.940	0.940	0.940	0.948	0.948	0.905
Monk2	Training accuracy	1.000	1.000	1.000	1.000	1.000	1.000
	Testing accuracy 1	0.984	0.984	0.984	0.984	0.984	1.000
	Testing accuracy 2	1.000	1.000	1.000	1.000	1.000	1.000
Monk3	Training accuracy	0.900	0.900	0.900	0.897	0.891	0.873
	Testing accuracy 1	0.838	0.838	0.838	0.865	0.883	0.856
	Testing accuracy 2	0.777	0.795	0.795	0.857	0.866	0.875
Average	Training accuracy	0.913	0.913	0.913	0.911	0.907	0.891
	Testing accuracy 1	0.819	0.824	0.822	0.831	0.836	0.825
	Testing accuracy 2	0.815	0.819	0.817	0.832	0.830	0.815

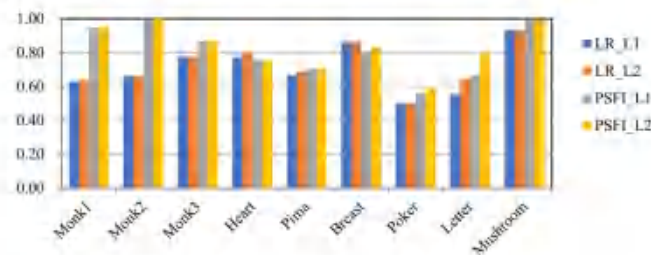


Fig. 9. Comparison of accuracy for each dataset between PSFI and logistic regression model.

category 0 to form 6 new datasets (T1-T6) with the data of category 1 respectively, where the ratio of the number of samples of category 0 to that of category 1 is approximately 1:1. We use T1-T5 as the training dataset and T6 as the validation set. Xgboost is adopted to build a classifier on the T1-T5 training dataset, and then use PSFI and T6 to build the fusion model of the Xgboost classifier. The model is shown in Fig. 11.

The prediction results of the Xgboost classifier in the T1-T5 for 300 iterations are shown in Table 24. In the experimental results, the Xgboost classifier has a higher accuracy and recall rate for each training dataset. Among them, the T5 dataset has the best classifier in which recall rate is 99.79% and F_1 Score is 0.9636. The Xgboost fusion model based on PSFI has improved accuracy and F_1 Score, reaching 99.85% and 0.9946 respectively. Especially, F_1 -score

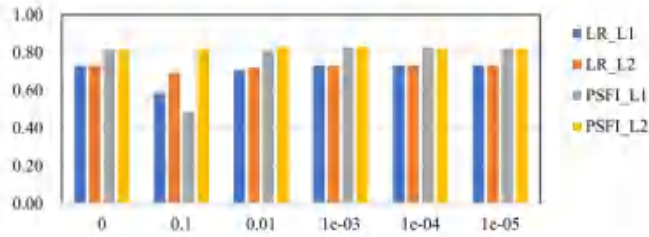


Fig. 10. Comparison of average accuracy for all datasets at each λ between PSFI and logistic regression model.

Table 23
Data distribution of Poker

	Classes	Samples
Training dataset(TR)	Class 0	126764
	Class 1	14500
Testing dataset(TE)	Class 0	295734
	Class 1	33122

Table 24
Experimental results of Xgboost for 300 iterations compared with PSFI

Training dataset	Testing dataset	Precision rate	Recall rate	F ₁ Score
T1	T6	0.9905	1.0000	0.9952
	TE	0.9101	0.9979	0.9520
T2	T6	0.9891	1.0000	0.9945
	TE	0.9112	0.9976	0.9525
T3	T6	0.9919	1.0000	0.9959
	TE	0.9229	0.9979	0.9590
T4	T6	0.9892	1.0000	0.9945
	TE	0.9102	0.9972	0.9517
T5	T6	0.9924	1.0000	0.9962
	TE	0.9316	0.9979	0.9636
T1-T5 fusion	T6	0.9999	1.0000	0.9999
	TE	0.9985	0.9907	0.9946

is higher than any other Xgboost classifiers. This shows that PSFI can be used as a fusion tool for excellent classifiers.

To further illustrate the problem, we continue to compare the Xgboost classification model for 200 iterations and its fusion model, as shown in Table 25. Experiments show that PSFI has a good effect on the fusion of the Xgboost model for 200 iterations, and the F₁ Score is higher 0.064 than the best situation of Xgboost model. In the training dataset, the recall rate of Xgboost is improved, but the false positive rate is also higher. PSFI fusion model reduces the error and greatly improves the accuracy, so the F₁ Score is higher.

Leading to the higher F₁ Score may be because T6 and the testing dataset have similar data distributions.

Therefore, we supplemented T6 to train the Xgboost model and compare with the fusion model. The experimental results are shown in Table 26. The F₁ Scores of the Xgboost model for 200 iterations and 300 iterations on T6 are 0.9112 and 0.9565, respectively, while the F₁ Scores of PSFI-based Xgboost fusion model for 200 iterations and 300 iterations are 0.9831 and 0.9946 respectively, the fusion model is better than

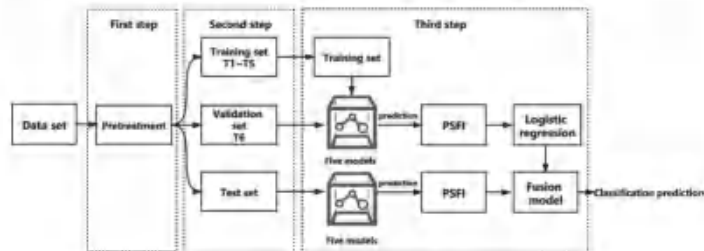


Fig. 11. Xgboost Fusion model based on PSFI for unbalanced data.

Table 25
Experimental results of Xgboost for 200 iterations compared with PSFI

Training dataset	Testing dataset	Precision rate	Recall rate	F ₁ Score
T1	T6	0.9774	1.0000	0.9885
	TE	0.8166	0.9114	0.8956
T2	T6	0.9762	1.0000	0.9879
	TE	0.8225	0.9922	0.8994
T3	T6	0.9803	1.0000	0.9901
	TE	0.8341	0.9936	0.9069
T4	T6	0.9759	1.0000	0.9878
	TE	0.8134	0.9919	0.8932
T5	T6	0.9816	1.0000	0.9907
	TE	0.8542	0.9936	0.9187
T1-T5 fusion	T6	0.9997	1.0000	0.9999
	TE	0.9965	0.9701	0.9831

Table 26
Comparison of Xgboost single model and fusion model with different iterations

No.	Model	Dataset	Precision	Recall	F ₁ Score
1	Xgboost 200 iterations	T6	0.9999	1.0000	1.0000
		TE	0.8414	0.9935	0.9112
2	Fusion model 200 iterations	T6	0.9997	1.0000	0.9999
		TE	0.9965	0.9701	0.9831
3	Xgboost 300 iterations	T6	1.0000	1.0000	1.0000
		TE	0.9184	0.9978	0.9563
4	Fusion model 300 iterations	T6	0.9999	1.0000	0.9999
		TE	0.9985	0.9907	0.9946

Xgboost. The F₁ Score of a single model increased by 7.19% and 3.81% respectively. The training accuracy of the Xgboost model reaches the training accuracy of the fusion model, and the F₁ Score on the training dataset is higher than that of the fusion model. It is not as good as the fusion model on the testing dataset. It proves that the fusion model can comprehensively consider the role of each classifier and prevent overfitting to a certain extent.

5. Conclusions and future works

This article introduced the application of fuzzy integral in classification problems and proposed an optimized PSFI combining with the Spark framework. PSFI adopted parallel computing and sparse storage used into solving fuzzy integral for the first time. It improved the calculation efficiency of fuzzy integral and enabled fuzzy integral to be extended to higher-dimensional big data.

- (1) Aiming at reducing the time and space complexity of fuzzy integral, parallel computing

and sparse storage on Spark was introduced. Meanwhile, PSFI was implemented based on the Cython language. It was verified that PSFI has a greater improvement in running time and storage. It is suitable for distributed environments very well. PSFI saved a lot of memory and disk space, and can be used to process higher-dimensional features. Therefore, PSFI has higher efficiency.

- (2) This article implemented multiple classification experiments to verify the effectiveness and practicability of PSFI. The comparison of classification accuracy with MNI verified that PSFI has better prediction performance. The results compared with the logistic regression model showed the superior of PSFI in processing those interactive features. In the application on unbalanced big data classification, PSFI can ensemble the existing excellent models to achieve a fusion model with higher accuracy. Therefore, PSFI is an important classifier and fusion tool in data mining.

In summary, PSFI is an optimized parallel algorithm, which has a high operating efficiency and can be used for classification problems. At the same time, PSFI is also a Python class library, which can be combined with data mining models such as Spark and Scikit-learn to realize the combination and optimization of data mining models.

Although PSFI algorithm has achieved some results, it still has some shortcomings. The time complexity level of PSFI algorithm remains unchanged, which is exponential. The running time will still increase with the increase of data dimension. In addition, if the feature dimension continues to increase, then the necessary memory overhead is still large. At present, there is no very efficient method to solve super large sparse linear equations. Therefore, feature extraction or feature sampling, establishing multiple fuzzy integral classifiers, and integrating multiple classifiers by using fuzzy integral may play a role in classification with a large number of features, which may reduce the running time and space costs of the model.

Acknowledgments

This work was supported in part by the Guangzhou Key Laboratory of Smart Agriculture (201902010081), the Science and Technology

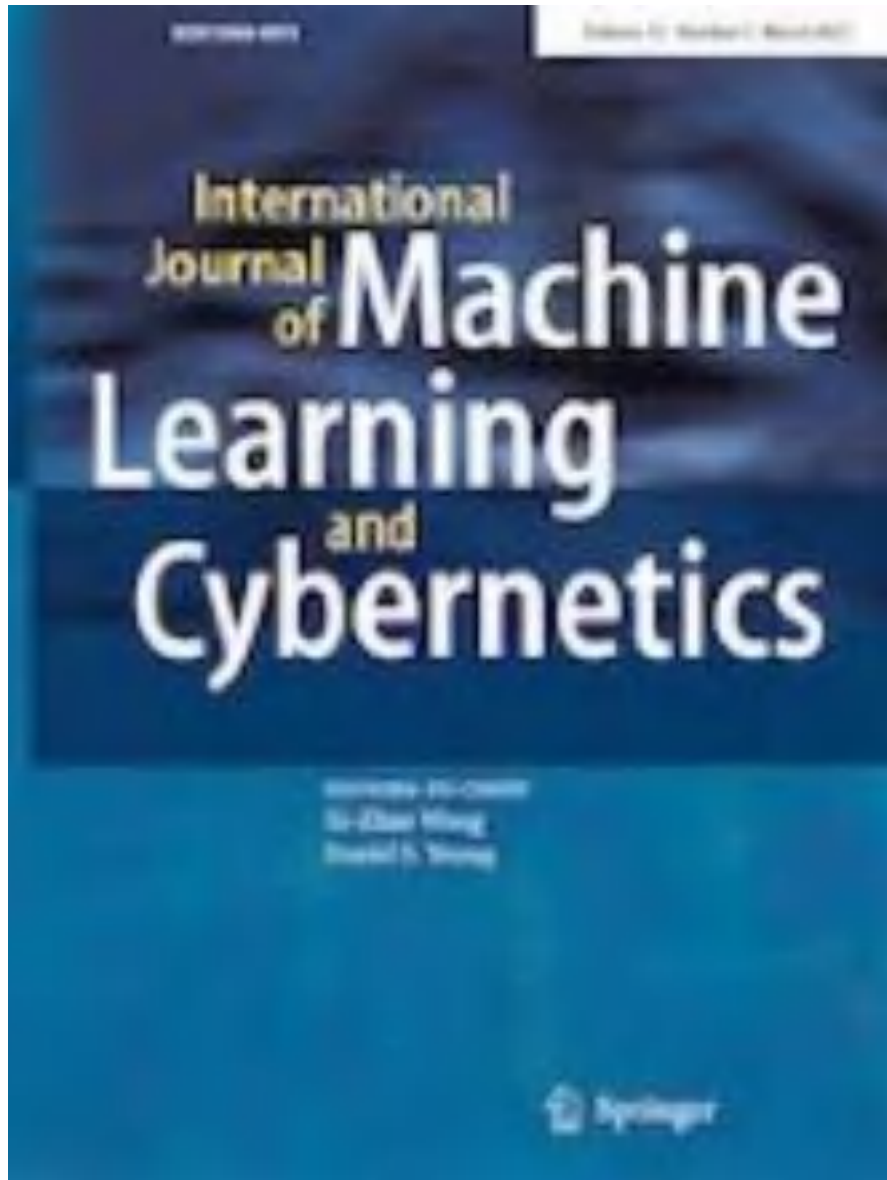
Planning Project of Guangdong Province (No.: 2017A040406023) and the Science and Technology Planning Project of Guangzhou City (No.: 201804010353).

References

- [1] M. Sugeno, Theory of fuzzy integrals and its applications, Doctoral Thesis, Tokyo Institute of Technology (1974).
- [2] T. Murafushi, M. Sugeno and M. Machida, Non monotonic fuzzy measures and the Choquet integral, *Fuzzy Sets and Systems* **64** (1994), 73–86.
- [3] K. Ishii and M. Sugeno, A model of human evaluation process using fuzzy measure, *International Journal of Man-Machine Studies* **22** (1985), 19–38.
- [4] T. Onizawa, M. Sugeno, Y. Nishiwaki, et al., Fuzzy measure analysis of public attitude towards the use of nuclear energy, *Fuzzy Sets and Systems* **20**(3) (1986), 259–289.
- [5] K. Tanaka and M. Sugeno, A study on subjective evaluation of color printing images, *Int J Approximate Reasoning* **5** (1991), 213–222.
- [6] K. Inoue and T. Aizawa, A study on the industrial design evaluation based upon non-additive measures, in: *The 7th Fuzzy System Symp.* (1991).
- [7] T. Washio, H. Takahashi and M. Kitamura, A method for supporting decision making on plant operation based on human reliability analysis by fuzzy integral, *The 2nd Int. Conf. on Fuzzy Logic and Neural Networks*, 1992[C].
- [8] M. Grabiec, The application of fuzzy integrals in multi-criteria decision making, *European Journal of Operational Research* **89**(3), (1996), 445–456.
- [9] Z. Wang, K.S. Leung, M.L. Wong, et al., A new type of nonlinear integrals and the computational algorithm[J], *Fuzzy Sets and Systems* **112**(2) (2000), 223–231.
- [10] K. Wu, Z. Wang, P.A. Heng, et al., Using generalized Choquet integral in projection pursuit based classification, in: *Ijca World Congress and Nafps International Conference*, 2001, Icaim, 2001.
- [11] K. Xu, Z. Wang, P.A. Heng, et al., Classification by nonlinear integral projections, *Fuzzy Transactions on Fuzzy Systems* **11**(2) (2003), 187–201.
- [12] Z. Wang and H.F. Guo, A new genetic algorithm for nonlinear multiregressions based on generalized Choquet integrals, in: *The IEEE International Conference on Fuzzy Systems* (2003).
- [13] J.F. Wang, K.H. Lee and K.S. Leung, L₁-norm Regularization Based Nonlinear Integrals, *Springer Berlin Heidelberg* (2009).
- [14] P.D. Gader, M.A. Mohamoud and J.M. Keller, Fusion of handwritten word classifiers, *Pattern Recognition Letters* **17**(6) (1996), 577–584.
- [15] J. Zhu, J. Wang and W. Hu, Combination of QSVM classifiers with fuzzy integral for large scale classification, *Journal of Intelligent & Fuzzy Systems* **28**(5) (2015), 2257–2268.
- [16] P. Chen and D. Zhang, Constructing support vector machines ensemble classification method for imbalanced datasets based on fuzzy integral, in: *Proceedings of the 27th International Conference on Modern Advances in Applied Intelligence*, **8481** (2014).
- [17] J. Zhu, S. Zhang, M. Zhang, et al., Fuzzy integral-based GLM ensemble for imbalanced big data classification, *Soft Computing* **2018** 1–13.
- [18] M. Grabiec, The application of fuzzy integrals in multi-criteria decision making, *European Journal of Operational Research* **89**(3) (1996), 445–456.
- [19] Z. Wang, K. Leung and J. Wang, A genetic algorithm for determining nonadditive set functions in information fusion, *Fuzzy Sets and Systems* **102**(3) (1999), 463–469.
- [20] Z. Wang, K.S. Leung and G.J. Klir, Applying fuzzy measures and nonlinear integrals in data mining, *Fuzzy Sets and Systems* **156**(3) (2005), 371–380.
- [21] Q. He, Q. Tan, X. Ma, et al., The High-Activity Parallel Implementation of Data Preprocessing Based on MapReduce, in: *International Conference on Rough Set & Knowledge Technology*, Springer-Verlag, (2010).
- [22] M. Grbić, A three-phase mapreduce-based algorithm for searching biomedical document databases, *IEEE - International Journal of Electrical Engineering and Computer* **3**(1) (2019).
- [23] J. Chen, K. Li, Z. Tang, et al., A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment[J], *IEEE Transactions on Parallel & Distributed Systems* **28**(4) (2017), 919–933.
- [24] Q. Wang and H. Chen, Optimization of parallel random forest algorithm based on distance weight[J], *Journal of Intelligent and Fuzzy Systems* **39**(6) (2020), 1–13.
- [25] M. Kumar and S.K. Rath, Classification of microarray using MapReduce based proximal support vector machine classifier, *Knowledge-Based Systems* **89** (2015), 584–602.
- [26] H. Zhu, P. Li, P. Zhang, et al., A High-Performance Parallel Ranking SVM with OpenCL on Multi-core and Many-core Platforms, *International Journal of Grid and High Performance Computing* **11**(1) (2019), 17–28.
- [27] M. Zaharia, M. Chowdhury, M.J. Franklin, et al., Spark: Cluster computing with working sets, *Hot Cloud* **10** 10–10 (2010), 95.
- [28] J. Qin, Y. Hu, P. Xu, et al., Inferring gene regulatory networks by integrating ChIP-seq/ChIP and transcriptomic data via LASSO-type regularization methods, *Methods* **67**(3) (2014), 294–303.
- [29] J. Deng, Z. Ou, Y. Zhu, et al., Towards efficient and scalable data mining using spark, in: *International Conference on Information and Communications Technologies (ICT 2014)*, (2014).
- [30] C. Jin, R. Liu, Z. Chen, et al., A Scalable Hierarchical Clustering Algorithm Using Spark, in: *IEEE First International Conference on Big Data Computing Service and Applications*, (2015). DOI: 10.1109/BigDataService.2015.67.
- [31] D. Chen, C. Shen, J. Feng, et al., An efficient parallel link similarity join for massive multidimensional data using spark, *International Journal of Database Theory and Application* **8**(3) (2015), 57–68.
- [32] A.K. Kolliopoulos, E. Ylapanis, F. Tekiner, et al., A Parallel Distributed Weka Framework for Big Data Mining Using Spark, in: *IEEE International Congress on Big Data* **2015** 9–16.
- [33] T. Murafushi and M. Sugeno, Fuzzy measures and fuzzy integrals, *Fuzzy Measures and Integrals: Theory and Applications* **2000** 3–41.
- [34] M. Assefi, E. Behrooosh, G. Liu and A.P. Taffi, Big Data Machine Learning using Apache Spark MLlib, *IEEE Big Data 2017*, (2017).

- [35] A. Asuncion and D.J. Newman, UCI Machine Learning Repository <http://www.ics.uci.edu/ml/MLRepository.html>. (2007).
- [36] J. Che, C. Shi, Y. Yu, et al., A synthetical performance evaluation of OpenVZ, Xen and KVM, in: *5th IEEE Asia-Pacific Services Computing Conference, APSCC 2010, 6–10 December 2010, Hangzhou, China*, (2010).
- [37] S. Behnel, R. Bradshaw, C. Citro, et al., Cython: The Best of Both Worlds, *Computing in Science & Engineering* **13**(2) (2011), 31–39.
- [38] J. Wang, K. Leung, K. Lee, et al., Multiple nonlinear integral for classification, *Journal of Intelligent & Fuzzy Systems* **28**(4) (2015), 1635–1645.
- [39] T. Chen and C. Guestrin, XGBoost: A Scalable Tree Boosting System, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data* (2016), 785–794. <https://doi.org/10.1145/2939672.2939785>.

2.3. Fuzzy measure with regularization for gene selection and cancer prediction



Original Article | Published: 20 April 2021

Fuzzy measure with regularization for gene selection and cancer prediction

JinFeng Wang , ZhenChen He, ShuaiJiu Huang, Hao Chen, Wen Zhong Wang & Fahad Pourghasemi*International Journal of Machine Learning and Cybernetics* 12, 2403–2405 (2021) | [Cite this article](#)
240 Accesses | 8 Citations | 1 Altmetric | [Metrics](#)

Abstract

Dealing with high-dimensional gene expression data is a challenging issue, and it is crucial to select multiple informative subsets of genes for cancer classification. In this regard, many statistical and machine learning methods with regulations have been developed. However, these methods neglected the epistasis, i.e., some genes may cover or affect other genes. In this article, we propose a fuzzy measure with regularization, which adopts L_1 and $L_{1/2}$ norms for sparse solutions, known as FMR, to describe the interaction between genes. Regularization with L_1 and $L_{1/2}$ can obtain a series of sparse solutions which help solving fuzzy measure quicker than traditional methods, such as Genetic Algorithm. FMR obtains a subset of genes corresponding to the fewest nonzero fuzzy measure values, and consequently, selects the important gene(s) according to the frequency of appearance in the selected gene subsets. Besides, three base classifiers, including SVM, KNN and DEN, are employed as underlying models to verify the effectiveness of the selected subset(s) of genes. Experimental results

Download PDF **Sections** [Figures](#) [References](#)[Abstract](#)[Introduction](#)[Preliminaries](#)[Models based on regularizations](#)[Experimental studies](#)[Conclusions](#)[Notes](#)[References](#)[Acknowledgements](#)[Funding](#)[Author information](#)[Additional information](#)[Rights and permissions](#)[Reprints and ads](#)



Fuzzy measure with regularization for gene selection and cancer prediction

JinFeng Wang¹ · ZhenYu He¹ · ShuaiHui Huang¹ · Hao Chen¹ · WenZhong Wang² · Farhad Pourpanah³

Received: 26 August 2020 / Accepted: 20 March 2021 / Published online: 20 April 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Dealing with high-dimensional gene expression data is a challenging issue, and it is crucial to select multiple informative subsets of genes for cancer classification. In this regard, many statistical and machine learning methods with regularizations have been developed. However, these methods neglected the epistasis, i.e., some genes may cover or affect other genes. In this article, we propose a fuzzy measure with regularization, which adopts L_1 and $L_{1/2}$ norms for sparse solutions, known as FMR, to describe the interaction between genes. Regularization with L_1 and $L_{1/2}$ can obtain a series of sparse solutions which help solving fuzzy measure quicker than traditional methods, such as Genetic Algorithm. FMR obtains a subset of genes corresponding to the fewest nonzero fuzzy measure values, and consequently, selects the important gene(s) according to the frequency of appearance in the selected gene subsets. Besides, three base classifiers, including SVM, KNN and DBN, are employed as underlying models to verify the effectiveness of the selected subset(s) of genes. Experimental results indicate that the selected genes by FMR are consistent with several clinical studies. In addition, it can produce comparable results in terms of accuracy as compared with other methods reported in the literature. The codes used in this article are freely available at: <https://github.com/wangphoenix/ICMLC>.

Keywords Gene selection · Fuzzy measure · Regularization · Cancer classification

1 Introduction

Cancer is one of the most deadly diseases worldwide in which early detection can dramatically improve the patient's survival. With recent advancements in microarray technology, gene expression profiling has provided considerable evidence to predict different types of cancer. Recently, many machine learning techniques such as Naïve Bayes (NB), decision tree (DT), neural networks (NN), k -nearest neighbor (KNN), and support vector machine (SVM) [1–5] have been applied to solve cancer classification problems.

Due to the high cost of obtaining the gene expression data, they usually contain a few high-dimensional samples, in which the number of features exceeds the number of samples. Thus, extracting useful information from such data becomes difficult and optimization equations based on genes cannot be solved efficiently [6–9]. It is crucial to select a small subset of genes that contains as much information as the high-dimensional genomic data. In this regard, gene selection becomes a key pre-processing step for cancer classification using genomic data [10].

Over the years, many gene selection techniques have been proposed [11–15]. Most of these methods suppose that there is no relationship between genes. However, the epistasis shows that some genes may cover or affect other genes, and selecting genes individually may affect the disease diagnosis [16]. Therefore, selecting a group of correlated genes is desirable than selecting individual ones. In bioinformatics and computational biology, genes can be grouped according to a specific structure, such as the biological pathway.

Gene selection methods were proposed based on classical feature selection algorithms. Feature selection include three kind of ways, i.e. wrapped Many methods, such as

✉ JinFeng Wang
wangjinfeng@scau.edu.cn

¹ College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China

² College of Economics and Management, South China Agricultural University, Guangzhou 510642, China

³ College of Mathematics and Statistics, Guangdong Key Lab. of Intelligent Information Processing, Shenzhen University, Shenzhen 518060, China

smoothly clipped absolute deviation (SCAD) [17], adaptive least absolute selection and shrinkage operator (LASSO) [18], Elastic net [19], Stage wise LASSO[20] and Dantzig selector [21], with L_1 regularization have been developed for data mining. Some of them were applied for gene selection with logistic regression [22–26]. However, L_1 regularization cannot perform effectively when there is a high correlation between genes. It may obtain inconsistent genes in some situations [27] and lead to some bias when estimating the parameters in the logistic regression [28]. Comparatively, $L_{1/2}$ regularization is an effective technique that can produce stable results and increase the predictive accuracy of the model. Nonetheless, $L_{1/2}$ regularization with the least square estimation cannot address the parsimony of the model and interpretability of the coefficient values.

Apart from co-relationship in statistics, there is an interaction of genes to describe the relationship between their appearing values, in which the Lebesgue-like integral with respect to a nonadditive set function cannot deal due to the uncertainty. To overcome this problem, the fuzzy measure with a nonadditive set function has been proposed [29]. Many classification models based on fuzzy measure were proposed [30, 31]. The fuzzy measure is an effective method that can describe the importance of gene groups [32, 33]. For a given gene set $G = \{g_1, \dots, g_n\}$, the related fuzzy measure is a function: $\mu: \{A_j \rightarrow [0, 1]\}$ where A_j is a subset of power set of G . There will be 2^n fuzzy measures. Since the genes' size is very large, the fuzzy measure's size will be extremely large. Traditionally, evolutionary computation based methods, such as Genetic Algorithm (GA) [34], was used to optimize the values of fuzzy measure and achieved high accuracy for cancer prediction. However, it has a slow convergence speed and time-consuming search. So a new method needs to be found for solving fuzzy measure with sparse values and high speed. Regularization with L_1 and $L_{1/2}$ can obtain a series of sparse solutions which help solving fuzzy measure quicker than those traditional methods. An optimization model is to be constructed about fuzzy measure. Regularization with L_1 and $L_{1/2}$ was adopted for solving the fuzzy measure. The non-zero values in the sparse fuzzy measure correspond to important genes.

In this article, we propose a group of new methods based on fuzzy measure using L_1 and $L_{1/2}$ regularizations, known as FMR- L_1 and FMR- $L_{1/2}$, for gene selection and cancer classification. Three base classifiers including KNN, SVM, and DBN are used as underlying models to evaluate the performance of the selected subset of genes by FMR- L_1 and FMR- $L_{1/2}$. The experimental outcome indicates that FMR- L_1 and FMR- $L_{1/2}$ are able to select the most informative genes and produce comparable results as compared with other methods reported in the literature.

The rest of this article is organized as follows: Section 2 presents the structure of the fuzzy measure. The details of

the proposed FMR- L_1 and FMR- $L_{1/2}$ models are described in Sect. 3. Section 4 provides the experimental results and discussion. Finally, concluding remarks and suggestions for future studies are given in Sect. 5.

2 Preliminaries

In this section, we firstly present the mechanism of fuzzy measure, and then discuss how it can be used to describe the interaction between genes. For a given data set $\{(G_i, y_i)\}_{i=1}^l$ with l samples, where $G_i = (g_{i1}, \dots, g_{in}) \in \mathbb{R}^n$ and y_i indicate the gene vector and disease class of the i th sample, respectively, the genes' expression data is numerical and can be described by an n -dimensional vector $(f(x_1), f(y_1), \dots, f(x_n))$.

2.1 Fuzzy measure

When the gene set G is finite, the power set of G is denoted by $P(G)$. \mathcal{F} represents a σ -algebra of subsets of $P(G)$.

Definition 1 Let (G, \mathcal{F}) be a measurable gene data space. Set function $\mu: P(G) \rightarrow (-\infty, \infty)$ is called as fuzzy measure if it is monotone non-decreasingly, that is, $A \subseteq B$ implies $\mu(A) \leq \mu(B)$ for any $A, B \in \mathcal{F}$.

Due to the nonadditivity of the fuzzy measure, the global influence of genes to the disease is not just the sum of their individual contributions. A signed fuzzy measure, which is a generalization of fuzzy measure, has been defined in [35].

Definition 2 A set function $\mu: P(G) \rightarrow (-\infty, \infty)$ is called a signed (non-monotonic) fuzzy measure if $\mu(\emptyset) = 0$.

For fuzzy measure $\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_{2^n-1})$, how to find the genes corresponding to each value will affect the computing efficiency. For simplicity, binary decoding is adopted. As such, the subscript j of each fuzzy measure μ_j is expressed as a sequence of binary digits $k_1 k_2 \dots k_n$. k_j is used to denote $\mu(G_j)$, where $G_j = \{g_i | k_i = 1\}$, $j = k_1 k_2 \dots k_n$. Suppose that the gene number $n = 3$, the maximum of j is $2^3 - 1 = 7$. For example, assume that the solution is $(0, 0, \mu_1, \dots, \mu_7, \dots, 0)$, the process of obtaining important genes is shown in Fig. 1. μ_3 corresponds to μ_{011} , so gene group includes g_1 and g_2 .

The following example explains that a Signed Fuzzy Measure can be used for describing the interaction among the contributions from the combinations of genes towards a certain disease. Worker efficiency corresponds to Fuzzy Measure.

Example 1 Let $X = \{x_1, x_2, x_3\}$ be the set of three workers. They are hired for manufacturing a certain type of products.

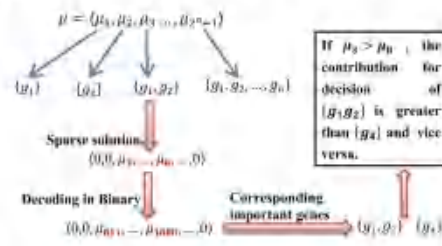


Fig. 1 Process of obtaining the important genes

Their individual and joint efficiency $\mu: P(X) \rightarrow (0, \infty]$ is given as follows:

$$\mu(E) = \begin{cases} 0 & \text{if } E = \emptyset \\ 5 & \text{if } E = \{x_1\} \\ 3 & \text{if } E = \{x_2\} \\ 10 & \text{if } E = \{x_1, x_2\} \\ 4 & \text{if } E = \{x_3\} \\ 4 & \text{if } E = \{x_1, x_2\} \\ 6 & \text{if } E = \{x_2, x_3\} \\ 9 & \text{if } E = X. \end{cases}$$

Set function μ is a fuzzy measure that describes the interaction among the contributions from individual workers towards the target, the total number of toys manufactured by these workers. In this example, $\mu(\{x_1, x_2\}) > \mu(\{x_1\}) + \mu(\{x_2\})$ shows that workers x_1 and x_2 cooperate well, therefore, the interaction between their contributions is mutually promoting. While $\mu(\{x_1, x_3\}) < \mu(\{x_1\}) + \mu(\{x_3\})$, even $\mu(\{x_1, x_2\}) < \mu(\{x_1\})$, shows that workers x_1 and x_2 cooperate very badly and the interaction between their contributions is mutually inhibitive. This set function is not monotonic.

2.2 Solving fuzzy measure with regularization method

For solving fuzzy measure, a kind of classification tool-fuzzy integral with respect to fuzzy measure μ is adopted. The generalized fuzzy integral of a function f with respect to fuzzy measure, μ , is defined by:

$$\int f d\mu = \int_{-\infty}^{\mu} [\mu(F_a) - \mu(G)] da + \int_{\mu}^{\infty} \mu(F_a) da, \quad (1)$$

where $F_a = \{g | f(g) \geq a\}$ for any $a \in (-\infty, \infty)$ is called the a -cutoff. Indeed, fuzzy integral is a linear-based operator

to deal with nonlinear problems. For speeding up computation, we need to have aid of the transformation of fuzzy integral.

To compute the fuzzy integral, the values of f , i.e. $f(g_1), f(g_2), \dots, f(g_n)$, are usually sorted in a non-decreasing order, so that $f(g'_1) \leq f(g'_2) \leq \dots \leq f(g'_n)$, where $\{g'_1, g'_2, \dots, g'_n\}$ is a certain permutation of $\{g_1, g_2, \dots, g_n\}$. Thus, the value of fuzzy integral can be obtained by:

$$\int f d\mu = \sum_{i=1}^n [f(g'_i) - f(g'_{i-1})] \mu(\{g'_i, g'_{i-1}, \dots, g'_n\}), \quad (2)$$

where $f(g'_0) = 0$.

To simplify, we adopt a new scheme to calculate the value of fuzzy integral by the inner product of two $(2^n - 1)$ -dimension vectors as:

$$Y = \int f d\mu = \sum_{j=1}^{2^n-1} z_j \mu_j$$

where

$$z_j = \begin{cases} \min_{i: \{ \frac{i}{2} \} \in I^j} f(g_i) - \max_{i: \{ \frac{i}{2} \} \in I^j} f(g_i), & \text{if } \nu > 0, \nu \in \{2^k - 1\} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

for $j=1, 2, \dots, 2^n - 1$ with a convention that the maximum on the empty set is zero [27]. Here, $\text{frac}(\frac{i}{2})$ denotes the fractional part of $\frac{i}{2}$. In formula (3), if we express j in the binary form $k_1 k_2 \dots k_n$, then $\{i | \text{frac}(\frac{i}{2}) \in [\frac{1}{2}, 1)\} = \{i | k_1 = 1\}$ and $\{i | \text{frac}(\frac{i}{2}) \in [0, \frac{1}{2})\} = \{i | k_1 = 0\}$. For understanding previous formula, we give an example for explanation.

Example 2 Suppose that $n=3$, $f(g_1) = 0.1$, $f(g_2) = 0.8$, $f(g_3) = 0.6$, the computation process of coefficient $z_j, j = 1, 2, \dots, 7$ is presented as Table 1.

This calculation scheme can easily find the coefficient matrix of a linear system with unknown variables μ . It has been applied to many applications, such as regression and classification [36–38]. In these practical applications, values of the signed fuzzy measure are usually considered as unknown parameters which are to be estimated using the training data sets. Using an algebraic method, such as the least square method, make this scheme convenient to estimate μ and reduce the complexity. In order to use the regularization method to determine μ , we need to construct a $l \times (2^n - 1)$ augmented matrix, i.e., $Z = [z_{ij}]$, for l cases.

According to (3).

Table 1 Illustration for computing coefficients

j	$\{k_j, l_j\}$	$\{l(k_j = 1)\}$	min	$\{l(k_j = 0)\}$	max	τ_j
1	{00}	$i=1$	$\min(f(g_1)) = 0.1$	$i=2,3$	$\max(f(g_2), f(g_3)) = 0.8$	0
2	{01}	$i=2$	$\min(f(g_2)) = 0.8$	$i=1,3$	$\max(f(g_1), f(g_3)) = 0.6$	0.2
3	{01}	$i=1,2$	$\min(f(g_1), f(g_2)) = 0.1$	$i=3$	$\max(f(g_3)) = 0.6$	0
4	{10}	$i=3$	$\min(f(g_3)) = 0.6$	$i=1,2$	$\max(f(g_1), f(g_2)) = 0.8$	0
5	{10}	$i=1,3$	$\min(f(g_1), f(g_3)) = 0.1$	$i=2$	$\max(f(g_2)) = 0.6$	0
6	{11}	$i=2,3$	$\min(f(g_2), f(g_3)) = 0.6$	$i=1$	$\max(f(g_1)) = 0.1$	0.5
7	{11}	$i=1,2,3$	$\min(f(g_1), f(g_2), f(g_3)) = 0.1$	$i=0$	0	0.1

$$\tau_{mi} = \begin{cases} \min_{k=1}^i (f(g_{mi})) - \max_{k=0}^i (f(g_{mi})), & \text{if } > 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where $j = 1, \dots, 2^n - 1, m = 1, 2, \dots, l, j_i$ indicates the value of the i -th element in binary expression of j -th case.

The transformed system of linear equations can be written as follows:

$$\begin{cases} z_1 \mu_1 + z_2 \mu_2 + \dots + z_{1(2^n-1)} \mu_{(2^n-1)} = y_1 \\ z_2 \mu_1 + z_2 \mu_2 + \dots + z_{2(2^n-1)} \mu_{(2^n-1)} = y_2 \\ \dots \dots \dots \\ z_n \mu_1 + z_n \mu_2 + \dots + z_{n(2^n-1)} \mu_{(2^n-1)} = y_n \end{cases} \quad (5)$$

For $n \gg 1$, solving (5) is a NP problem. In this article, the regularization method is introduced to solve this problem. The regularization method for solving the fuzzy measure is as follows:

$$L(\lambda, \mu) = \operatorname{argmin} \left[\frac{1}{l} \sum_{i=1}^l \|y_i - Z_i \mu\|^2 + \lambda \|\mu\|_k \right] \quad (6)$$

s.t. $Z\mu = Y$

where $\frac{1}{l} \sum_{i=1}^l \|y_i - Z_i \mu\|^2$ is loss function and λ is the penalty parameter. When $k=0$, it is called as the L_0 regularizer [39, 40], which is the earliest regularization method applied for feature extraction. Constrained by the number of coefficients, the L_0 regularizer yields the sparsest solutions. However, it suffers from the problem of combinatorial optimization, which is difficult to be solved. When $k=1$, it is known as LASSO [41], which just needs to solve a quadratic programming problem. LASSO is sparser than the L_0 regularization. Donoho had proved that the solutions of the L_1 regularizer are equivalent to those of the L_0 regularizer for the sparsity problem under some conditions [42–44]. Therefore, the L_1 regularizer can avoid the NP optimization problem. When $k=2$, it is the ridge regression which has the solutions with the properties of being smooth. Nonetheless, L_2 cannot possess the sparse property. It has been proved that regularization

with L_q has more sparseness when $0 < q < 1$ and $L_{1/2}$ is totally representative of L_0 [45].

3 Models based on regularizations

As described in Sect. 2, the L_1 regularization contains most properties of the L_2 regularization, and obtains sparse solutions that can be easily interpreted [46]. Based on the above account, we firstly adopt L_1 -norm to solve fuzzy measure in (6) (Sect. 2.1). Then, we introduce $L_{1/2}$ to further improve the performance of the model (Sect. 2.2).

3.1 Fuzzy measure with L_1

Determining the fuzzy measure is the key procedure in constructing fuzzy integral-based models. Fuzzy measure represents the importance of genes and their interaction degree. The aim is to find a solution of Fuzzy measure with the fewest nonzero values that contains the most informative genes and gene combinations. Using L_1 -norm regularization, we can minimize the following function to reduce the number of nonzero values in fuzzy measure:

$$L(\lambda, \mu) = \operatorname{argmin} \left[\frac{1}{l} \sum_{i=1}^l \|y_i - Z_i \mu\|^2 + \lambda \|\mu\|_1 \right] \quad (7)$$

s.t. $Z\mu = Y$

where λ is the penalty term to control the condensation compress degree of the fuzzy measure. In Ref. [47], the LASSO model based on Gauss–Seidel method was proposed. The main advantages of the Gauss–Seidel approach include its: (i) simplicity, and (ii) low iteration cost. Thus, we applied this kind of LASSO to solve fuzzy measure. As an example, assume $\{0, 0, 0.6, 0.9, 0, 0, -0.3\}$ is an obtained solution by fuzzy measure, which means $u_i = u(\lambda_j) = u(\{g_1, g_2, g_3\}) = 0.6, u_i = u(\lambda_k) = u(\{g_2, g_3\}) = 0.9$, and $\mu_7 = \mu(\lambda_7) = \mu(\{g_1, g_2, g_3\}) = -0.3$. In this example, g_3 is the most important gene because μ_4 is with the highest

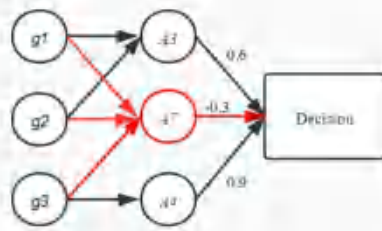


Fig. 2 The relationship of all genes with respect to decision

fuzzy measure. The second gene group is $\{g_1, g_2\}$. The positive value indicates that genes g_1 and g_2 cooperate very well and, therefore, the interaction between their contributions is mutually promoting. The fuzzy measure of the universal set $\{g_1, g_2, g_3\}$ is negative, which means if g_3 joins into $\{g_1, g_2\}$, the affection of $\{g_1, g_2\}$ will be restrained. This finding has important meaning for disease analysis and pharmaceuticals. Figure 2 shows the relationship of the genes.

3.2 Fuzzy measure with $L_{1/2}$

The $L_{1/2}$ regularization is more stable and performs similar to L_1 in sparsity and gene selection. Therefore, we introduced it into FMR model. Refs [45] and [17] proposed algorithms to transform nonconvex problem into regularization problem with L_1 and L_2 , respectively. However, both algorithms are computationally expensive. To alleviate this issue, half threshold iterative algorithm [17], which is based on existence of gradient components of $\|\mu\|_{1/2}^{1/2}$, was proposed. It operates as follow:

$$h(x) = \begin{cases} f_d(x), & |x| > \lambda^* \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where f_d is called as defining function.

If there exists $H(g) = \{h_1(g), h_2(g), \dots, h_n(g)\}^T$ and each $h_i(g)$ is determined by g_i , then $H(g)$ is called as threshold operator. If L_q regularization can be represented by h , then the solution of L_q regularization can be expressed as $g^{t+1} = H(Bg^t)$, in which t indicates the iteration number and B is an affine operator which can be deduced by the following formula (4).

As a result, if half threshold iterative is adopted to solve $L_{1/2}$ regularization, then threshold h and defining function f_d need to be constructed. The threshold operator of $L_{1/2}$ regularization is formulated as follows:

$$H(g, \lambda) = \{h(g_1, \lambda), h(g_2, \lambda), \dots, h(g_n, \lambda)\}^T, \quad (9)$$

where

$$h(g, \lambda) = \begin{cases} k(g, \lambda), & |g| < \lambda \\ 0, & \text{otherwise} \end{cases}$$

$$k(g, \lambda) = \frac{2}{3}g \left(1 + \cos \left(\frac{2}{3}\pi - \frac{2}{3}f(g, \lambda) \right) \right),$$

$$f(g, \lambda) = \arccos \left(\frac{\sqrt{2}}{2} \left(\frac{\lambda}{|g|} \right)^3 \right).$$

This algorithm will converge to sparse solution after T iterations. The convergence of threshold operator has been proved in [17].

3.3 Gene selection model

Theoretically, a brute force search can obtain the best gene group by trying all situations, but it is computationally expensive. Thus, we propose wrapper-based gene selection algorithms based on fuzzy integral combined with L_1 or $L_{1/2}$ regularizations, which uses the fuzzy measure to define the correlation between genes. The proposed algorithm can be divided into two parts. The first part employs the L_1 or $L_{1/2}$ regularizations to remove the most redundant genes, i.e., genes with low correlation are removed, while the second part sorts all subsets in descending order according to the fuzzy measure's values. The top one subset corresponds to those genes with the most important genes as the follows:

$$G_{top} = \{g_j | \max(\mu(A_j)) \text{ and } j = 1\}, \quad (10)$$

$$i = 1, 2, \dots, n; j = 1, 2, \dots, 2^n - 1$$

Then a classifier can be used to evaluate the effectiveness of the selected genes. If the predictive accuracy reaches to an optimal value, the subset with top G_{top} will be considered as the most important one, else the classifier will continue training. The best situation is that the accuracy reaches to an acceptable level as well as gene subsets number converges to one. In this algorithm, the non-zero fuzzy measures are considered. The optimized model is solved by L_1 or $L_{1/2}$ regularization, so the most values of fuzzy measures are zero. Gene selection can remove the less informative genes successively according to their correlation and avoid the drawback of brute searching. This algorithm adopts regularization method instead of heuristics. The computing complexity of the new algorithm is kept being consistent with $L_{1/2}$ computation. It will not change with the number

Table 2 Details of gene expression datasets examined

Dataset	Samples	Control	Normal	Genes
DLBCL	77	19	56	7129
Prostate	102	30	52	12,625
Colon	62	22	40	2000
GLI-85	85	59	26	22,283

of features. It is related with the machine computing power. The detailed algorithm is summarized in Algorithm 1.

Algorithm 1 The Pseudocode of Gene selection based on FMR-L₁

Input: Data set $D = \{(G_1, y), (G_2, y), \dots, (G_n, y)\}$, in which $G_i = (g_{i1}, g_{i2}, \dots, g_{im})$.
Output: Gene group $A = \{A_1, A_2, \dots, A_n\}$.

- Adopt regularization to remove redundant genes and obtain the preliminary gene group $A' = \{A'_1, A'_2, \dots, A'_n\}$;
- Construct the optimization model according to A' using fuzzy integral. (Regularization method is used for solving the model for sparse solution $\mu = \{\mu_1, \mu_2, \mu_3, \dots, \mu_{|A'|}\}$);
- Sort genes in increasing order under $A' = \{A'_1, A'_2, \dots, A'_n\}$;
- Trim a classifies according to μ' and obtain the accuracy c^0 ;
- While $1 - \mu'_{c^0}$ is not input:
 - Remove A'_i from A' , then the accuracy is c^1 after calling classifier;
 - If $c^1 \geq c^0$, then $c^0 = c^1$ and go to 4;
 - If A'_i is placed back and A'_{i+1} will be try again;
- A' is assigned to A .

In summary, there are six steps to build a model based on fuzzy integral for gene selection and cancer classification, as follow:

- Step 1: Divide data set into training set and testing set.
- Step 2: Compute the coefficient matrix $Z = [z_{ij}]$ (5) using training data.
- Step 3: Optimize (6) using improved LASSO, fuzzy measure can be determined as sparse solvent.
- Step 4: Select important genes according to the sparse values of fuzzy measure.
- Step 5: Compute the fuzzy integral for each test sample using the coefficient matrix and fuzzy measure obtained during training.
- Step 6: Classify test samples according to fuzzy integral values and update performance indicator.

4 Experimental studies

In this section, four binary benchmark microarray data sets are used to evaluate the effectiveness of the proposed FMR models. These data sets include DLBCL (Lymphatic tumor) [46], Prostate (prostatic cancer) [47], Colon (Colon cancer) [48] and GLI-85 (Glioma) [49]. The details of the data sets are listed in Table 2. Among them, the processed Colon data

set is publicly available,¹ while the rest are required to be pre-processed. All experiments are conducted using MATLAB 2016b with Intel Core i7 and 8G memory.

Two performance indicators, i.e., *accuracy* and *recall*, are used for performance comparison.

4.1 Data pre-processing

Data pre-processing plays an important role in data mining. It decreases the systematic error that may lead to data distortion in experiments. In Bioconductor, several pre-processing techniques are provided such as MAS5 (Micro-Array Suite 5.0) [50], RMA (Robust Multichip Average) [51], and GC-RMA (GeneChip RMA) [52]. Several tools can be used to select appropriate pre-processing techniques. In this study, the distributions of intensity chart and the MA chart are employed to select the best pre-processing technique. The distribution of intensity chart presents data distribution and helps to check normalization. The horizontal and vertical coordinates are intensity and density, respectively. While the MA chart shows the relationship between differential expression and intensity of gene, in which $M = \log_2 \frac{R}{G}$, $A = \frac{1}{2} \log_2 R * G$. In a single-channel chip, R is a sample and G is a control sample. For a finite number of genes, the performance is good if $M \rightarrow 0$.

Figure 3 shows the distribution of all chip data of the DLBCL data set before and after pre-processing with MAS5, RMA, and GC-RMA. Before pre-processing (Fig. 3a), the distribution of data is not consistent; some of the samples do not coincide with a normal distribution. While pre-processing (Fig. 3c–d) affected the distribution of the data, GC-RMA performed better than other methods because the distribution of all chip data is almost coincident. In addition, Fig. 4 shows the MA chart of the DLBCL data set. As can be seen, M tends to be 0 for GC-RMA, thus it is selected to pre-process the DLBCL data set.

Similarly, from Figs. 5 and 6 can be found that GC-RMA is a more appropriate pre-processing technique for the Prostate data set. While, both GC-RMA and RMA techniques produced similar results for the GLI-85 data set (Figs. 7, 8). Specifically, GCRMA performs slightly better than RMA, but it deviates the normalization. Instead, RMA is able to maintain normal distribution. Therefore, RMA is selected as a pre-processing technique for the GLI-85 data set.

4.2 Results and discussions

In this section, we firstly, present the results of the FMR with L₁-norm (FMR-L₁) for the DLBCL and Prostate data sets to show the effectiveness of the selected genes by

¹ <http://genomics.wpi.edu/genom/cdu/bioinfo/>.

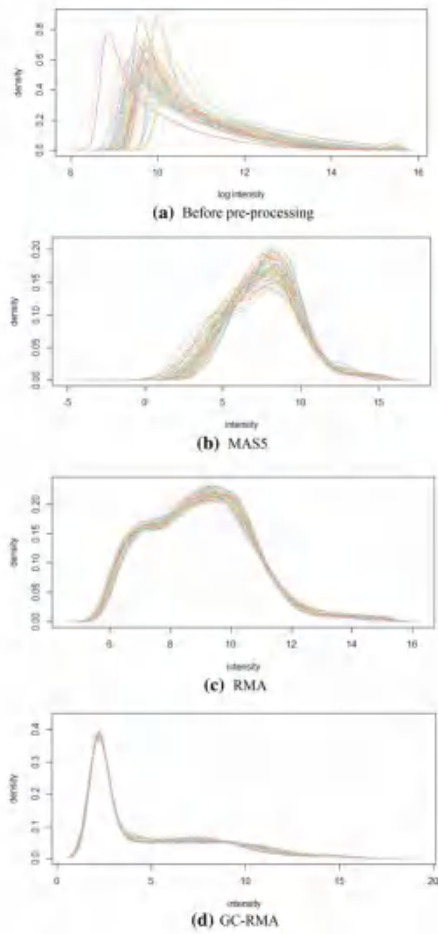


Fig. 3 Distribution chart of intense for the DLBCL data set before and after pre-processing

fuzzy measure. Then, the proposed FMR with $L_{1/2}$ -norm is compared with Relief, t test, pure $L_{1/2}$ and other gene selection methods reported in the literature. In addition, KNN, SVM, and DBN are used to evaluate the effectiveness of the selected subset of genes in tackling classification problems. Due to the small number of samples, five-fold cross-validation is used and repeated 100 times.

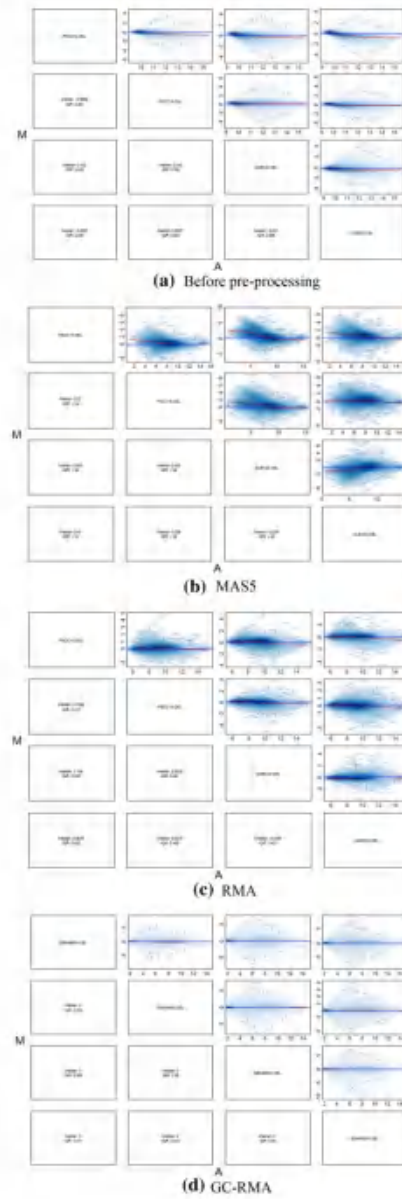


Fig. 4 MA chart for the DLBCL data set before and after pre-processing

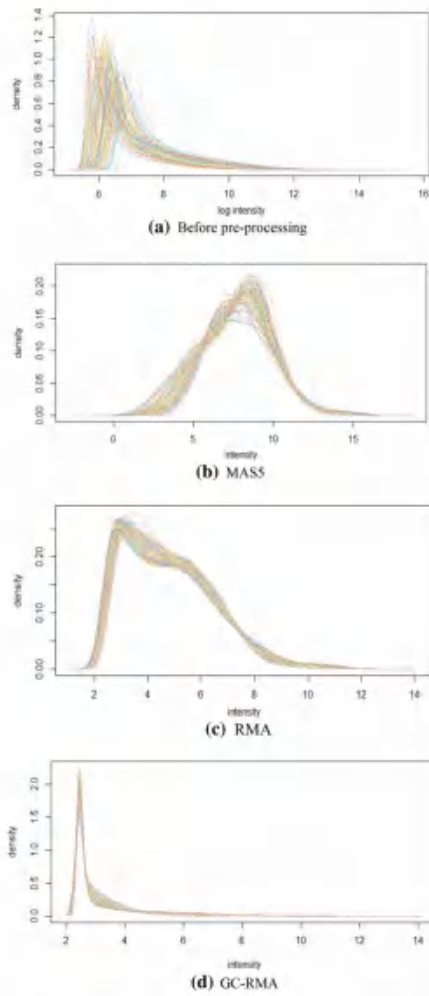


Fig. 5 Distribution chart of intense for the Prostate data set before and after preprocessing

4.2.1 Validating the performance of regularization

Solving fuzzy measure with regularization is introduced for quicker convergence and less time-consuming than traditional version. For validating this point, a comparison experiment was implemented on DLBCL just between

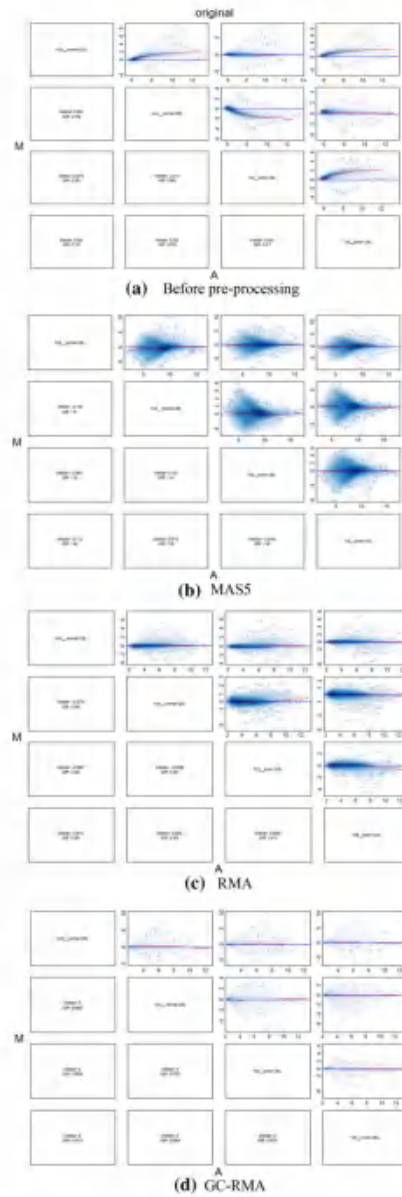


Fig. 6 MA chart for the Prostate data set before and after pre-processing

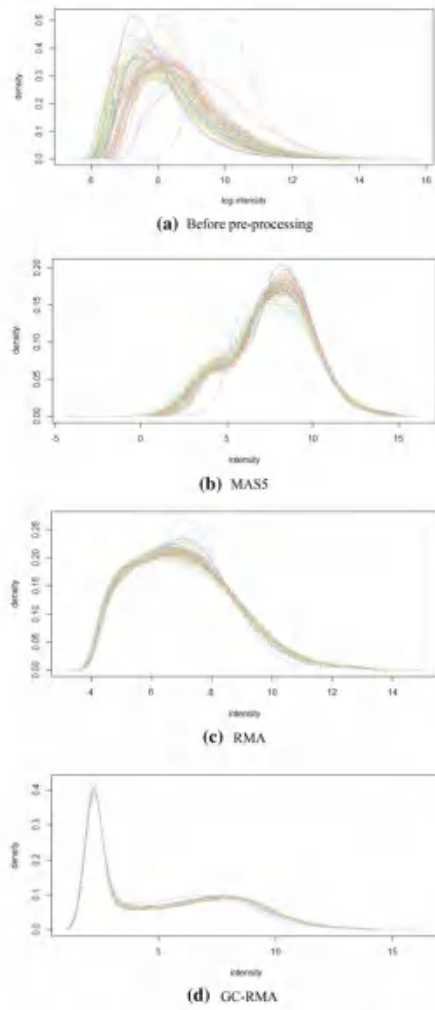


Fig. 7 Distribution chart of intense for the GLI-85 data set before and after pre-processing

solving fuzzy measure with regularization and GA. When the number of features is small, two kinds of method have little difference. But the number of fuzzy measure increases exponentially with feature growth. The method with GA will be slower greatly than that with regularization.

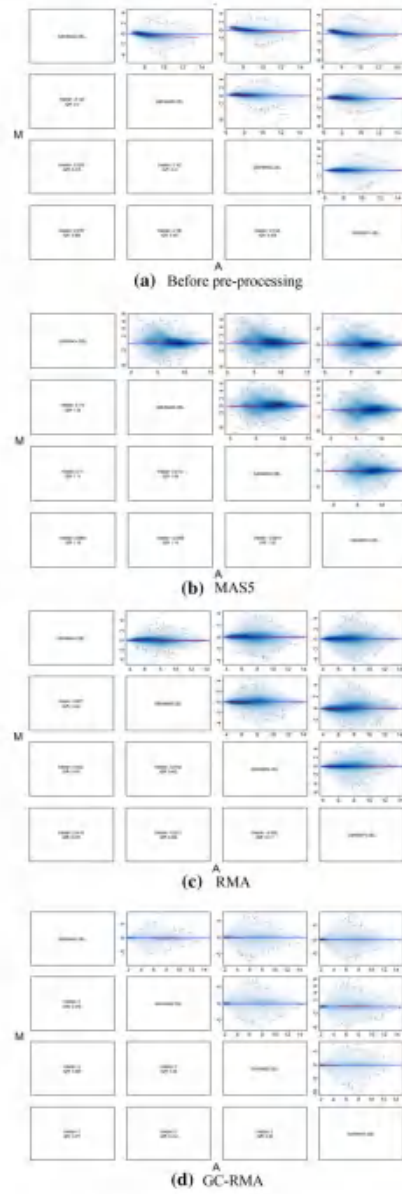


Fig. 8 MA chart for the GLI-85 data set before and after pre-processing

Table 3 Comparison results between with and without regularization on DLBCL

Index	With regularization			Without regularization		
	3 features	6 features	10 features	3 features	6 features	10 features
Training accuracy	0.753	0.922	0.964	0.971	1.000	0.997
Training recall	1.000	0.911	0.965	0.907	1.000	1.000
Testing accuracy	0.754	0.911	0.961	0.911	0.860	0.922
Testing recall	1.000	0.904	0.947	0.750	0.650	0.850
Runtime(s)	0.109	0.155	1.842	68.61	428.88	8.51e+03

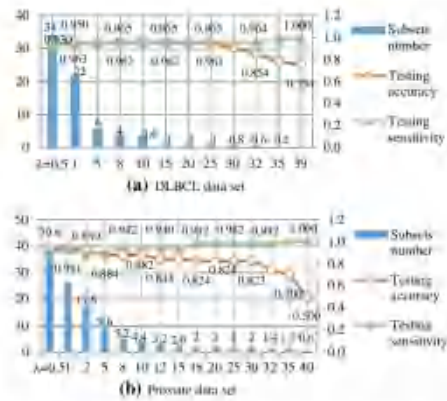


Fig. 9 The tendency of accuracy, and the number selected subsets of genes with different λ settings for **a** DLBCL and **b** Prostate data sets

In validation experiments, we select 3 features and 10 features in DLBCL respectively for testing the runtime and classification accuracy. The results are listed in Table 3. It can be seen that the classification performance with two methods is roughly the same. But the runtime of traditional method is more than that with regularization, especially for 10 features. So regularization is introduced to solve fuzzy measure in this article, which can improve the performance of models from efficiency. When the amount of features is increased greatly, FML can still obtain the feature subset and better performance quickly. The performance will be validated in following sections.

4.2.2 Testing FMRL on DLBCL and Prostate

We selected the top 10 genes of DLBCL and Prostate data sets. Table 4 shows the annotations of the selected top 10 genes for both data sets. Note that it cannot be proved that all the selected genes are the most important ones for classification. FMRL₁ with different λ setting is applied. The aim is to find an optimal λ that balances the trade-off between the

Table 4 The annotations of selected top 10 genes of DLBCL and Prostate data sets

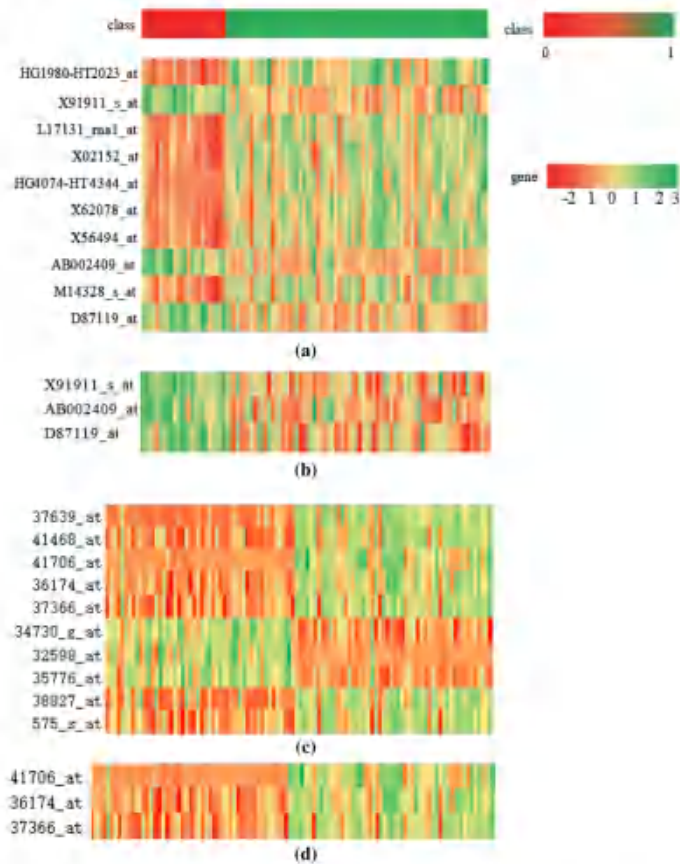
Rank	DLBCL		Prostate	
	Probe ID	Gene symbols	Probe ID	Gene symbols
1	HG1980- HT2023_at	TUBB2A	37639_at	HPN
2	X91911_s_at	GLIPR1	41468_at	TBGC1
3	L17131_m1_at	HMOA1	41706_at	AMACR
4	X02152_at	LDHL	56174_at	MARCKSL1
5	HG4074- HT4344_at	FEN1	37366_at	RDLIM5
6	X62078_at	GM2A	34730_s_at	TRO
7	X56494_at	PKM	32598_at	NELL2
8	AB002409_at	CCL21	35776_at	IFSN1
9	M14328_s_at	ENO1	38827_at	AGR2
10	D87119_at	TRIB2	575_s_at	EPCAM

predictive accuracy and number of gene subsets. Figure 9 shows the changing tendency of accuracy and the number of selected gene subsets. For both of datasets, accuracy and the number of selected gene subsets decreased by increasing λ . For DLBCL (Fig. 9a), the testing accuracy remained around 0.96 when λ varied from 5 to 26. In addition, the number of selected gene subsets is converged to 1 when λ varied from 15 to 25. The corresponding gene subset is $A_{642} = \{g_2, g_8, g_{10}\}$. The symbols of selected genes are GLIPR1, CCL21, and TRIB2, which are highlighted in Table 4.

For Prostate (Fig. 9b), although the number of selected subsets of genes did not converge to 1, it remained 2 when λ varied from 18 to 30. The genes are separated into two parts, i.e. $A_{328} = \{g_6, g_7, g_8\}$ and $A_{290} = \{g_1, g_2, g_3, g_4, g_5, g_9, g_{10}\}$. Both subsets have the same importance for the cancer classification.

In cancer prediction, it is very important to correctly predict all positive samples, i.e., higher sensitivity rate, rather than achieving the balanced sensitivity and specificity rates. For both data sets, the sensitivity can reach to 1 when λ is set to a large value (Fig. 9). These results prove the effectiveness of gene selection with fuzzy measures based on regularization.

Fig. 10 The heat map of a top 10 genes for DLBCL, **b** selected genes for DLBCL, **c** top 10 genes for Prostate, and **d** selected genes for Prostate



The heat map is a powerful visualization method for analyzing the distribution of data. It can be used to compare the new distribution with the original one. Figure 10 shows the heat maps of the top 10 genes and 3 selected genes by FMR for DLBCL and Prostate data sets. The top color bar represents the classes, including Follicular Lymphoma (FL)/normal (red) and DLBCLs/Tumors (green). Color transition in heat map from red to green corresponds to the shift from low to high normalized values of gene expression data. Figure 10a, b represent the heat maps of the top 10 genes selected by p -value and 3 selected genes by FMR for all samples of DLBCL, respectively. As can be seen, the FMR can divide most samples into two groups. In addition, the heat map of top 10 genes selected by p -value and 3 selected

genes by FMR for all samples of Prostate are shown in Fig. 10c, d, respectively.

Figure 11 plots the normalized value and the mean of each gene for DLBCL and Prostate to ensure that the selected genes are informative. The red lines indicate FL/normal cases and the green lines indicate DLBCL/Tumor cases. The straight lines represent the mean value of each case. It can be seen that the difference between mean values of samples from different classes of DLBCL for GLIPR1, CCL21 and TRIB2 genes are 1.699, 1.416 and 1.407, respectively. Consequently, the difference between means values of samples from different classes of the prostate for TRO, NELL2 and ITSN1 are 1.273, 1.273 and 1.231, respectively. This roughly keeps the model stable.

Fig. 11 The scatter of all samples according to genes: **a** distribution of gene GLIPR1 is with mean=1.158 for FL and mean=-0.379 for DLBCL; **b** distribution of gene CCL21 is with mean=1.066 for FL and mean=-0.349 for DLBCL; **c** distribution of gene TRIB2 is with mean=1.060 for FL and mean=-0.347 for DLBCL; **d** distribution of gene TRO is with mean=0.649 for normal and mean=-0.624 for tumor; **e** distribution of gene NEFL2 is with mean=0.649 for normal and mean=-0.624 for tumor; **f** distribution of gene ITSN1 is with mean=0.628 for normal and mean=-0.603 for tumor

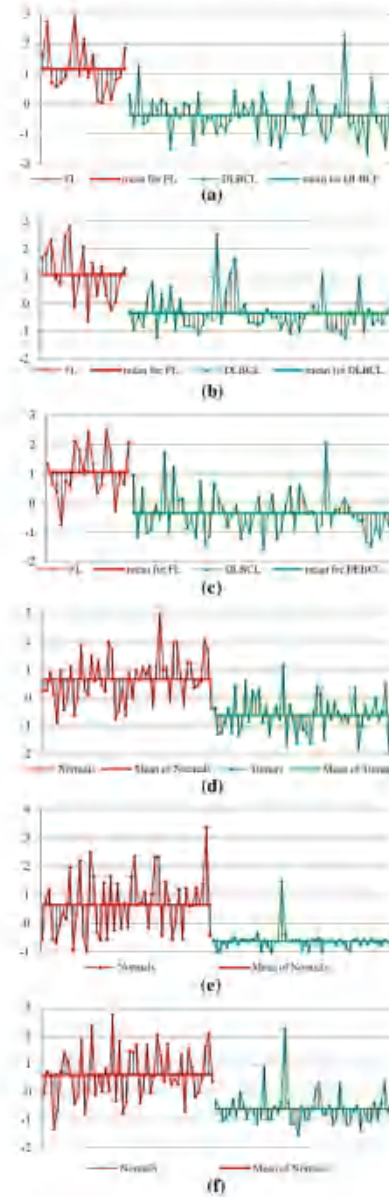
4.2.3 Testing FMR with $L_{1/2}$

In this section, the performance of the proposed FMR- $L_{1/2}$ gene selection model is compared with Relief, t -test and the conventional $L_{1/2}$ regularization. The KNN, SVM, and DBN are adopted as base classifiers, and the classification accuracy is used as a performance indicator. For just comparing the performance of gene selection of several methods, the base classifiers are the classical ones which locate in Matlab and use optimal parameters' set. The selected genes obtained by $L_{1/2}$ and FMR- $L_{1/2}$ for DLBCL, Prostate, Colon and GLI-85 data sets are presented in Table 3, respectively. FMR- $L_{1/2}$ is able to select fewer genes than conventional $L_{1/2}$. Meanwhile, it can be found that most of the selected genes by FMR- $L_{1/2}$ are covered by $L_{1/2}$, which proves that FMR- $L_{1/2}$ can provide a more condensed gene set for classification.

Figures 12, 13 and 14 show the accuracy rates of SVM, KNN, and DBN with different gene selection methods, respectively. The LOOCV (leave-one-out-cross-validation) [53] method was used and repeated 50 times for means and variances. Except for Prostate with SVM classifier, FMR- $L_{1/2}$ outperforms or performs similar to other methods.

In addition, Tables 6, 7, 8 and 9 compares the performance of FMR- $L_{1/2}$ using SVM as a base classifier with other methods reported in the literature in terms of classification accuracy and the number of selected features. GS2 with SVM [19] proposed a novel gene scoring functions to design the stable gene selection method GS2 for gene selection and using SVM for classification. Logistic $L_{1/2}$ [20] investigated a sparse logistic regression with the $L_{1/2}$ penalty for gene selection in cancer classification problems. IC-SVM [18] proposed a hybrid gene selection method, i.e. information gain-support vector machine (IG-SVM). In PCMI [21], a new feature selection method was proposed in which the component extraction is based on the correlation between feature space and class coding space. CEMFE [25] proposed a cross-entropy based multi-filter ensemble method for microarray data classification which used multiple filters to select top genes.

FMR- $L_{1/2}$ with SVM outperforms 3 out of 4 data sets (i.e., DLBCL, COLON and GLI-85), and ranks as the second-best method for DLBCL. However, it is not able to produce comparable results for Prostate data set.



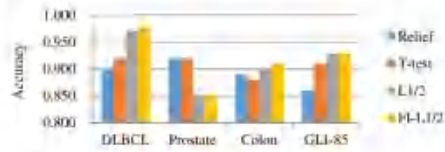


Fig. 12 Classification accuracy with SVM based on different gene selection methods

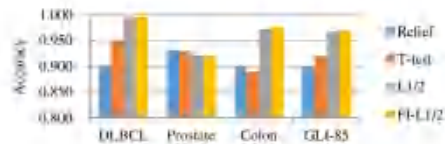


Fig. 13 Classification accuracy with KNN based on different gene selection methods

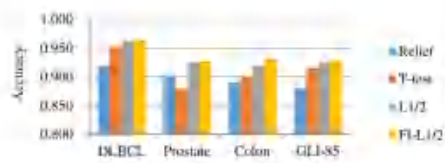


Fig. 14 Classification accuracy with DBN based on different gene selection methods

4.2.4 Informative genes analysis

Since microarray data are high dimensional and gene selection methods use different mechanisms, the selected genes by a method may be different from other methods. As can be seen from Table 4, the selected genes by FMR-L_{1/2} are consistent with those selected by other methods. The top 5 selected genes by FMR-L_{1/2} for DLBCL, Prostate, Colon and GLI-85 are discussed as follows.

The top five selected genes for DLBCL are CCL21, MCM7, TRIB2, GPR18 and HLA-DQB1. CCL21 (C-C Motif Chemokine Ligand 21) is a Protein Coding gene. Diseases associated with CCL21 include Postero-lateral Myocardial Infarction and Skin Carcinoma In Situ. Some research showed that Chemokine CCL21 and its receptor CCR7 play important role in the selective and specific metastasis of cancer cells. It may become a new target of anti-metastasis drugs [54]. MCM7 (Minichromosome maintenance complex component 7), plays an essential role in initiating DNA replication. Ref. [55] found that MCM7 high expression level

Table 5 The selected genes for each dataset by L_{1/2} and FMR-L_{1/2}

Datasets	Gene symbols(L _{1/2})	Gene symbols(FMR-L _{1/2})
DLBCL	CCL21	TTK
	MCM7	CREB1
	TRIB2	UBE2C
	IGH	MT2A
	GPR18	NME1
	HLA-DQB1	CKS2
	PRKCB	MYBL1
	CR2	
		CRS2
		MYBL1
Prostate	KLK3	HPN
	SATB2	NPY
	ITSN2	AGR2
	AQR	KLK3
	KLK2	ZNF148
	HIST1H2BM	RFG1
	TSPAN1	KLK2
	PDLIM5	PDLIM5
		HPN
		NPY
Colon	Hsa.3353	Hsa.5392
	Hsa.1832	Hsa.24582
	Hsa.1380	Hsa.2928
	Hsa.41159	Hsa.1454
	Hsa.627	Hsa.2337
	Hsa.2794	Hsa.2291
	Hsa.8192	Hsa.4333
	Hsa.41096	
		Hsa.3353
		Hsa.2291
GLI-85	ANXA1	DPYD
	EIF5A	SCAMP1
	RRM2	PEG3
	RP84Y1	SLC
	LTF	SULF1
	LDLR	ADCY1
	ADM	DIO2
	STMN2	
		ANXA1
		ADCY1

is associated with worse overall survival of HCC patients. TRIB2 (tribbles homologue 2) suppression, which is presented via direct interaction with AKT, is a key signaling protein in cell proliferation, survival and metabolism pathways. It reveals TRIB2 function as a regulatory component of the PI3K network, activates AKT in cancer cells and a novel regulatory mechanism underlying drug resistance [56]. GPR18 (G-protein-coupled receptor 18) mRNA is found to be highly expressed in several human T-Cell lymphotropic virus-transformed cell lines and primary human peripheral lymphocyte subsets, with very high expression identified in phytohaemagglutinin-activated CD4+ T-cells [57]. GPR18 is the most abundantly overexpressed orphan GPCR in all melanoma metastases, which is constitutively active and inhibits apoptosis, indicating an important role for GPR18 in

Table 6 Comparison results on DLBCL

Methods	No. selected features	Accuracy (%)	Remarks
GSI2, SVM [19]	55	96.10	LOOCV, mean of 100 runs
Logistic $L_{1/2}$, KNN [20]	10	94.80	LOOCV, mean of 50 runs
FMR- $L_{1/2}$, SVM	10	99.60	LOOCV, mean of 50 runs

The bold values represents the highest accuracy

Table 7 Comparison results on Prostate

Methods	No. selected features	Accuracy (%)	Remarks
IC-SVM [18]	3	96.08	unknown
Logistic $L_{1/2}$, KNN [20]	10	95.10	LOOCV, mean of 50 runs
PCMI [21]	11	95.10	LOOCV
GSI2, SVM [19]	47	95.10	LOOCV, mean of 100 runs
FMR- $L_{1/2}$, SVM	15	92.70	LOOCV, mean of 50 runs

The bold values represents the highest accuracy

Table 8 Comparison results on Colon

Methods	No. selected features	Accuracy (%)	Remarks
CEMFE, KNN [25]	14	96.77	Unknown
logistic $L_{1/2}$, KNN [20]	10	95.10	LOOCV, mean of 50 runs
CFS-IBPSO, NB [26]	4	94.90	LOOCV
IC-SVM [18]	3	90.00	Unknown
FMR- $L_{1/2}$, SVM	10	97.60	LOOCV, mean of 50 runs

The bold values represents the highest accuracy

Table 9 Comparison results on GLI-85

Methods	No. selected features	Accuracy (%)	Remarks
PCMI, SVM [21]	14	92.94	LOOCV
FMR- $L_{1/2}$, SVM	9	96.90	LOOCV, mean of 50 runs

The bold values represents the highest accuracy

tumor cell survival. GPR18 may be considered as a potential anticancer target in metastatic melanoma [58]. Meanwhile, it is indicated that GPR18 expression on PMNs is associated with increased severity and poorer outcome of sepsis [59].

The top five selected genes for Prostate are KLK3, KLK2, TSPAN1, PDLIM5 and HPN. KLK3 (Kallikrein Related Peptidase 3) is a protein-coding gene in the species *Homo sapiens*. It is found that rs1058205 polymorphism of KLK3 is a risk factor for prostate cancer development [60]. KLK2 (kallikrein-related peptidase 2) is a Protein Coding gene. It is one of the clinically validated urinary prostate cancer RNA biomarkers [61]. TSPAN1 (Tetraspanin 1) gene encodes the protein which is a member of the transmembrane 4 superfamily. TSPAN1 is an androgen-driven contributor to cell survival and motility in prostate cancer [62]. PDLIM5 (PDZ

And LIM Domain 5) encodes a member of a family of proteins that possess a 100-amino acid PDZ domain at the N-terminus and one to three LIM domains at the C-terminus. PDLIM5 is one of the biomarkers, which is used to predict prostate biopsy [63]. HPN (Hepsin) encodes a type II transmembrane serine protease that may be involved in diverse cellular functions, including blood coagulation and the maintenance of cell morphology. The HPN is one of the most frequently and prominently overexpressed genes in prostate cancer [64].

The top five ESTs (expressed sequence tags) of colon are Hsa.3353, Hsa.24582, Hsa.627, Hsa.2794, and Hsa.41098. Because some genes have not published in colon, ESTs are used as the labels in this dataset. The first one is Hsa.3353 corresponds to IFI6 (interferon alpha inducible protein 6) which is Interferon-inducible protein 6. It has been reported that IFI6 can inhibit apoptosis via mitochondrial-dependent pathway in dengue virus 2 infected vascular endothelial cells [65]. Ref. [66] revealed that IFI6 is an ER-localized integral membrane effector which has little effect on other mammalian RNA viruses, including the related Flaviviridae family member hepatitis C virus. The annotation for Hsa.24582 is vascular endothelial growth factor (VEGF). Researchers investigated that vascular endothelial growth factor (VEGF) and hypoxia-inducible factor 1alpha (HIF-1alpha)

expressions are prognostic significant markers of angiogenesis in colon cancer [67]. Hsa.627 corresponds to gene name CXCL8 (C-X-C motif chemokine ligand 8). In 2010, it has been proved that CXCL8 can be selected as important diagnostic markers for colon cancer and key targets for therapeutic intervention of the disease [68]. Recently, researchers found that CXCL3 and CXCL8 were identified as the key genes for diagnosis and prognosis of colon adenocarcinoma by gene expression and methylation profiles [69]. Hsa.2794 is heat shock protein 27 which can enhance the tumorigenicity of immunogenic rat colon carcinoma cell clones. This has been verified in 1998 [70]. Furthermore, some experiments indicated that HSP27 is closely connected with 5-FU resistance in colon cancer [71]. The last EST Hsa.41098 corresponds to a kind of Human Mullerian inhibiting substance gene. It was found that Hsa.41098 can inhibit growth of a human ovarian cancer in nude mice [72]. But for now, there is no medical evidence that MIS is related to COLON.

Finally, the top five selected genes for GLI-85 are ANXA1, EIF5A, FOXD1-AS1, ADM and STMN2. ANXA1 (Annexin A1) encodes a membrane-localized protein that binds phospholipids. Elevated expression of ANXA1 is observed in reactive astrocytes in peri-infarct regions [73]. EIF5A (Eukaryotic Translation Initiation Factor 5A) gene encodes the Eukaryotic translation initiation factor 5A-1 which is the only known protein to contain the unusual amino acid hypusine. FOXD1-AS1, a miR339/342 target, affected biological processes via protein EIF5A; thus, it might be considered as a new therapeutic target for glioblastoma. About FOXD1-AS1, its official full name is FOXD1 antisense RNA 1. A new research revealed that FOXD1-AS1 can be used as a potential oncogenic biomarker in glioma [74]. ADM (Adrenomedullin) is a Protein Coding gene. Diseases associated with ADM include Pheochromocytoma and Malignant Hypertension. Gene Ontology (GO) annotations related to this gene include signaling receptor binding and adrenomedullin receptor binding [75]. Diseases associated with ADM include Pheochromocytoma and Malignant Hypertension. STMN2 is a Protein Coding gene too. Diseases associated with STMN2 include Goldberg-Sprentzen Syndrome and Kuru. Recent research validated that STMN2 is the one of key markers in the developing intervertebral disc [76].

5 Conclusions

In this paper, gene selection methods based on a fuzzy measure with L_1 and $L_{0/2}$ regularizations, i.e., FMR- L_1 and FMR- $L_{0/2}$ for cancer classification have been proposed. Both methods are able to describe the interaction between genes; meanwhile, they can provide a sparse solution of fuzzy measure. Four microarray gene expression data were

employed to verify the effectiveness of FMR- L_1 and FMR- $L_{0/2}$. Before applying FMR- L_1 and FMR- $L_{0/2}$, for each data set an optimal pre-processing technique was selected to decrease the systematic error. Three base classifiers were used to compare the effectiveness of the selected subset of genes with other methods reported in the literature. In addition, the selected genes by FMR- $L_{0/2}$ were found to be more consistent with other clinical researches. In the future, we will continue to mine the effect of regularization in fuzzy measure and feature selection.

Acknowledgements We would also like to thank Prof. Xi-Zhao Wang in Shenzhen University for his support and revision.

Funding Funding was provided by the Science and Technology Planning Project of Guangdong Province of China (Grant no. 2017A040400023).

References

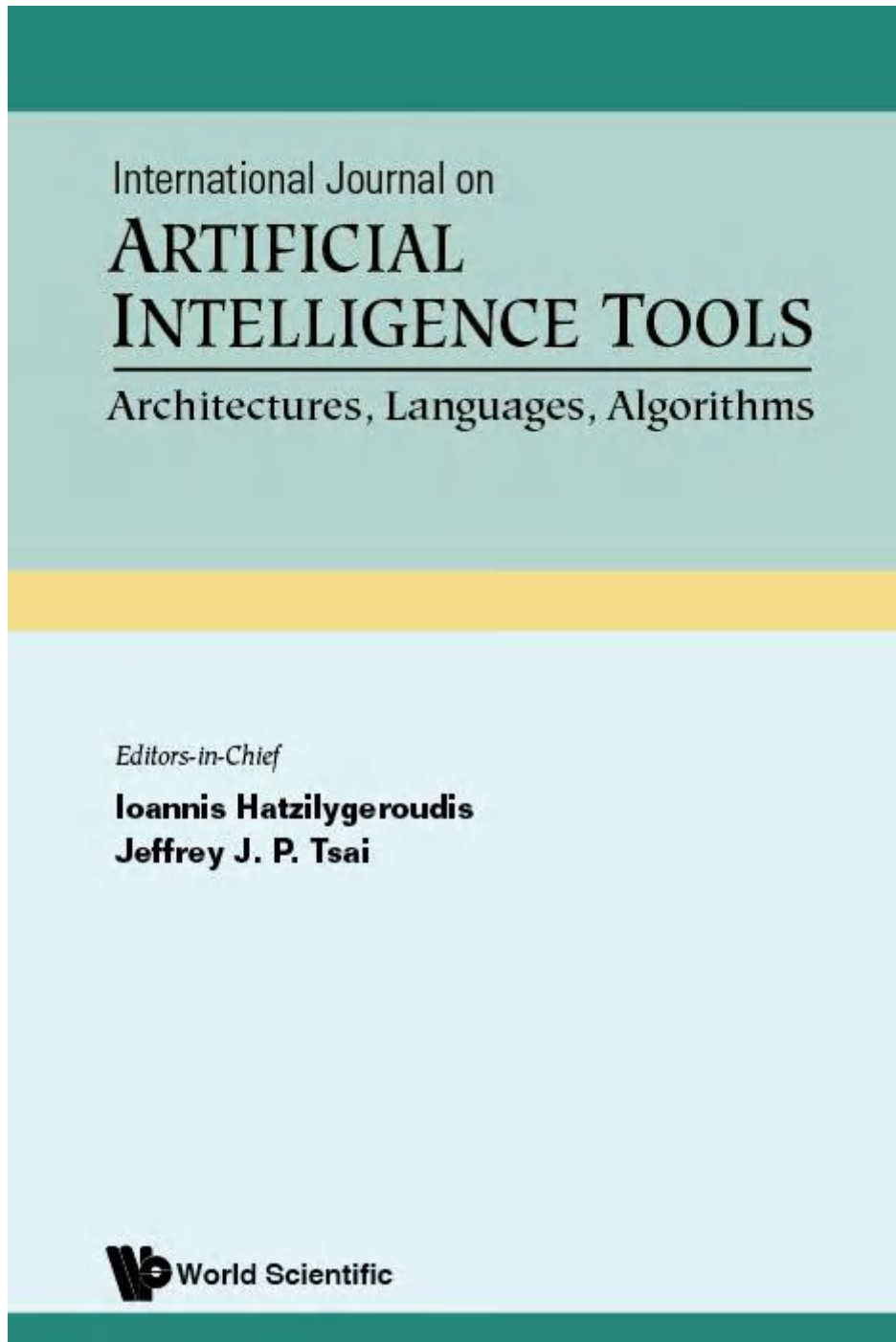
1. Gayathri BM, Sunanthi CP, Santhanam T (2013) Breast cancer diagnosis using machine learning algorithm—e survey. *Int J Distrib Parallel Syst* 4(3):105–112
2. Kharya S (2012) Using data mining techniques for diagnosis and prognosis of cancer disease. *Int J Comput Sci Eng Informat* 3(2):55–66
3. Krasuta K, Esarchos TP, Esarchos KP, Karamanidis MY, Iliadis DI (2013) Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* 13:6–17
4. Shajilhan SS, Shambi S, Manojithra V (2017) Applications of data mining techniques to model breast cancer data. *Int J Biomed Technol Adv Eng* 3(1):362–369
5. Srivastava SS, Sami A, Aharwal RP (2013) An *in silico* view on data mining approach in breast cancer data. *Int J Adv Comput Res* 5(4):256–262
6. Alonso-González CJ, Moró-Sánchez QJ, Simón-Juárez A, Villar-Aranda R (2012) Microarray gene expression classification with few genes: criteria in core-base attribute selection and classification methods. *Expert Syst Appl* 39(7):7170–7280
7. Cai Y, Zheng CH, Yang J, Shu W (2013) Sparse maximum margin discriminant analysis for feature extraction and gene selection on gene expression data. *Comput Biol Med* 43:993–994
8. Kalra J (2014) Classification methods for high-dimensional genetic data. *Biochem Biomed Eng* 34:10–18
9. Piao Y, Piao M, Park K, Ryu KH (2012) An ensemble correlation-based gene selection algorithm for cancer classification with gene expression data. *Bioinformatics* 28:3306–3315
10. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Collier H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286(539):531–537
11. Crawley GC, Tibshirani RL (2006) Gene selection in cancer classification using sparse logistic regression with Bayesian regularization. *Bioinformatics* 22(19):2349–2355
12. Zhou LT, Cao YH, Lv LL et al (2017) Feature selection and classification of urinary mRNA microarray data by iterative random forest to diagnose renal fibrosis: a two-stage study. *Sci Rep* 7:59832

13. Zhan G, Wu Y (2016) Feature subset selection for cancer classification using weight local sparsity. *Sci Rep* 6:34759. <https://doi.org/10.1038/srep34759>
14. Huang HH, Liu XY, Liang Y (2016) Feature selection and cancer classification via sparse logistic regression with the hybrid $L_{1/2}$ regularization. *PLoS ONE* 11(5):e0149675
15. Jayasuriya L, Krishna Anand S (2016) Feature selection for microarray data using WGCNA based fuzzy forest to map reduce paradigm. *Indian J Sci Technol*. <http://dx.doi.org/10.17485/ijst/2016/04480107971>
16. Algamal ZY, Lee MH (2015) Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. *Expert Syst Appl* 42(23):9326–9332
17. Xu Z, Chung K, Xu F et al (2012) $L_{1/2}$ regularization: a thresholding representation theory and a fast solver. *IEEE Trans Neural Netw* 23(7):1013–1027
18. Gao L, Ye M, Lu X et al (2017) Hybrid method based on information gain and support vector machine for gene selection in cancer classification. *Genomics Proteomics Bioinform* 15(6):389–395
19. Yang KJ, Cai Z, Li J et al (2006) A stable gene selection in microarray data analysis. *BMC Bioinform* 7(1):228–228
20. Liang Y, Liu C, Lam X et al (2013) Sparse logistic regression with a $L_{1/2}$ penalty for gene selection in cancer classification. *BMC Bioinform* 14(1):198–198
21. Yuan M, Yang Z, Ji G et al (2019) Partial maximum correlation information: a new feature selection method for microarray data classification. *Neurocomputing* 323:251–247. <https://doi.org/10.1016/j.neucom.2018.08.094>
22. Li C, Li H (2008) Network-constructed regularization and variable selection for analysis of genomic data. *Bioinformatics* 24(9):1175–1182
23. Smith V, Forde S, Jordan MI et al (2015) $L_{1/2}$ -regularized distributed optimization: a communication-efficient primal-dual framework. <https://arxiv.org/pdf/1512.04011v4.pdf>
24. Yuan GX, Hu CH, Liu CJ (2012) An improved GLMNET for $L_{1/2}$ -regularized logistic regression. *J Mach Learn Res* 13:1999–2030
25. Sun Y, Lu C, Li X (2018) The cross-entropy based multi-filter ensemble method for gene selection. *Genes* 9(5):258
26. Jain I, Jain VK, Jain R (2018) Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification. *Appl Soft Comput* 62:203–215
27. Zou H (2006) The adaptive LASSO and its oracle properties. *J Am Stat Assoc* (Taylor & Francis) 101:1418–1429
28. Meinshausen N, Yu B (2009) LASSO-type recovery of sparse representations for high-dimensional data. *Ann Stat* 37(2):246–270
29. Wang Z (1983) Asymptotic structural characteristics of fuzzy measure and their applications. *Fuzzy Sets Syst* 16(3):277–280
30. Chou R, Guo S, Wang X et al (2019) Fusion of multi-RSMOTE with fuzzy integral to classify bug reports with an imbalanced distribution. *IEEE Trans Fuzzy Syst* 27:2406–2420
31. Zhu J, Zhou X, Zhang S et al (2019) Ensemble RBF-based classifier using fuzzy integral for big data classification. *Int J Mach Learn Cybern* 10:3327–3337
32. Grabovich M (2000) The symmetrical Sugeno integral. *Fuzzy Sets Syst* 139:473–490
33. Wang Z, Guo HF (2003) A new genetic algorithm for nonlinear multiregressions based on generalized fuzzy integrals. *IEEE Int Conf Fuzzy Syst* 2:819–821
34. Mizukoshi T, Sugeno M, Mochida M (1994) Non monotonic fuzzy measures and the fuzzy integral. *Fuzzy Sets Syst* 64:73–88
35. Wang Z (2001) A new genetic algorithm for nonlinear multiregressions based on generalized Choquet integrals. In: Proc. 12th IEEE Intern. Conf. fuzzy systems, vol 2, pp 819–821
36. Wang W, Wang ZY, Kfir G (1998) Genetic algorithm for determining fuzzy measures from data. *J Intell Fuzzy Syst* 6(17):183
37. Leong KS, Wong ML, Lam W, Wang Z, Xu K (2002) Learning nonlinear multiregression networks based on evolutionary computation. *IEEE Trans Syst Man Cybern Part B* 32(5):630–644
38. Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov RN, Caki F (eds) *Second international symposium on information theory*. Budapest, pp 267–281
39. Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464
40. Tibshirani R (1996) Regression shrinkage and selection via the LASSO. *J R Stat Soc B* 58:267–288
41. Donoho DL, Huo X (2001) Uncertainty principles and ideal atomic decomposition. *IEEE Trans Inf Theory* 47:2845–2862
42. Donoho DL, Elad E (2003) Maximal sparsity representation via H^1 minimization. *Proc Natl Acad Sci* 100:2197–2202
43. Chen S, Donoho DL, Saunders M (2001) Atomic decomposition by basis pursuit. *SIAM Rev* 43:129–159
44. Xu ZH, Hu Z, Yao W et al (2010) $L_{1/2}$ regularization. *Sci China Inf Sci* 53(6):1159–1169
45. Stipp MA, Ross KN, Tamayo P et al (2002) Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 8(1):68–74
46. Singh D, Febo PG, Ross K et al (2002) Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 1(2):203–209
47. Daubechies I, Devore R, Fornasier M (2010) Iteratively reweighted least squares minimization for sparse recovery. *Commun Pure Appl Math* 63(1):1–38
48. Alizadeh A, Brannan N, Hastie T et al (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci USA* 96(12):6745–6750
49. Freije WA, Edmundo Castro-Vargas F, Fang Z et al (2008) Gene expression profiling of gliomas strongly predicts survival. *Clin Res* 66(16):6503–6510
50. Affymetrix (2001) Microarray suite user's guide version 5.0. Affymetrix Inc., Santa Clara
51. Bolstad BM, Irizarry RA, Anders M et al (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19(2):183–193
52. Wu Z, Irizarry RA, Gentleman R et al (2004) A model based background adjustment for oligonucleotide expression arrays. *J Am Stat Assoc* 99(468):909–917
53. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the fourteenth international joint conference on artificial intelligence*, Montreal, QC, Canada, pp 1137–1143
54. Lin Y, Sharma S, John MS (2013) CCL21 Cancer immunotherapy. *Cancers* 6:1098–1110
55. Qu K, Wang Z, Fan H et al (2017) MCM7 promotes cancer progression through cyclin D1-dependent signaling and serves as a prognostic marker for patients with hepatocellular carcinoma. *Cell Death Dis* 8(2):e2603. <https://doi.org/10.1038/s41422-016-3571-5>
56. Hill R, Madarista PA, Ferreira B et al (2017) TR102 confers resistance to anti-cancer therapy by activating the serine/threonine protein kinase AKT. *Nat Commun* 8:14667
57. Kato M, Hasegawa H, Inoue A, Murakami M, Yashiki T, Oka K, Yoshikawa M (2006) Identification of N-acetylhomocysteine as the endogenous ligand for orphan G-protein-coupled receptor GPR118. *Biochem Biophys Res Commun* 347(5):827–832
58. Finlay DB, Joseph WR, Guimsey NL, Glass MC (2016) GPR118 undergoes a high degree of constitutive trafficking but is unresponsive to N-acetylhomocysteine. *PLoS One*. <https://doi.org/10.7717/peerj.1835>

59. Zhang L, Qiu C, Yang L et al (2019) GPR18 expression on PMNs as biomarker for outcome in patient with sepsis. *Life Sci* 217:49–56
60. Ding WH, Ren KW, Yue C et al (2008) Association between three genetic variants in kallikrein 3 and prostate cancer risk. *Biosci Rep*. <https://doi.org/10.1042/BSR20181151>
61. Wang J, Koo KM, Wang Y et al (2018) 'Mix-to-Go' silver colloidal strategy for prostate cancer molecular profiling and risk prediction. *Anal Chem* 90:12698–12705
62. Munkley J, McClurg UL, Livermore KE et al (2017) The cancer-associated cell migration protein TSPAN1 is under control of androgens and its upregulation increases prostate cancer cell migration. *Sci Rep* 7:5249. <https://doi.org/10.1038/s41598-017-05489-5>
63. Albitar M, Ma W, Land L et al (2016) Predicting prostate biopsy results using a panel of plasma and urine biomarkers combined in a scoring system. *J Cancer* 7(3):297–303
64. Willbold R, Wirth K, Martini T, Holger S, Wittig R (2019) Excess hepsinproteolytic activity limits oncogenic signaling and induces ER stress and autophagy in prostate cancer cells. *Cell Death Dis*. <https://doi.org/10.1038/s41419-019-1830-8>
65. Qi Y, Li Y, Zhang Y, Zhang L, Wang Z (2015) IF16 inhibits apoptosis via mitochondrial-dependent pathway in dengue virus 2 infected vascular endothelial cells. *PLoS ONE* 10(8):e0132743
66. Blake RR, Ohlson MB, Edlson JL et al (2018) A CRISPR screen identifies IF16 as an ER-resident interferon effector that blocks flavivirus replication. *Nat Microbiol* 3:1214–1223
67. Choi YY, Cho HD, Park DG, Kim SY, Baek MD (2008) Expression of hypoxia-inducible factor-1 α and vascular endothelial growth factor in colon cancer: relationship to the prognosis and tumor markers. *Ann Coloproctol* 24(5):337
68. Mia HJ, Qi XG (2010) Role of cxc18/cxcr1 in the metastasis of human colon cancer. *World Chin J Digestol* 18(22):2379
69. Zhuo Q, Jiang C, Gao Q, Zhang Y, Wang G, Chen X et al (2020) Gene expression and methylation profiles identified cxc13 and cxc18 as key genes for diagnosis and prognosis of colon adenocarcinoma. *J Cell Physiol* 235:4902–4912
70. Garrido A, Fromentin A, Bonnotte B, Favre N, Moutet M, Arrigo AP et al (1998) Heat shock protein 27 enhances the tumorigenicity of immunogenic rat colon carcinoma cell clones. *Can Res* 58(23):5495–5499
71. Tsuruta, (2008) Heat shock protein 27, a novel regulator of 5-fluorouracil resistance in colon cancer. *Oncol Rep* 20(5):1165–1172. https://doi.org/10.3892/or_00000125
72. Donahoe PK, Fuller AF, Scully RE, Gory SR, Budzik GP (1981) Mullerian inhibiting substance inhibits growth of a human ovarian cancer in nude mice. *Ann Surg* 194(4):472–480
73. Masahiro S, Hалеomi H, Hiroyuki H, Sasaki SO, Masaki T, Tesuro A et al (2019) Upregulation of Annexin A1 in reactive astrocytes and its subtle induction in microglia at the boundaries of human brain infarcts. *J NeuropatholExpNeurol* 78(10):961–970. <https://doi.org/10.1093/jnen/olz079>
74. Gao YF, Liu JY, Mao XY et al (2020) LncRNA FOXD1-AS1 acts as a potential oncogenic biomarker in glioma. *CNS Neurosci Ther* 26:66–75. <https://doi.org/10.1111/cns.13152>
75. Kitamura K, Sakata J, Kangawa K, Kojima M, Matsuo H, Ito T (1993) Cloning and characterization of cDNA encoding a precursor for human adrenomedullin. *BiochemBiophys Res Commun* 194(2):720–725
76. Rodrigues-Pinto R, Ward L, Humphreys M et al (2018) Human notochordal cell transcriptome unveils potential regulators of cell function in the developing intervertebral disc. *Sci Rep* 8(1):12866. <https://doi.org/10.1038/s41598-018-31172-4>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations

2.4. Driver Fatigue Detection Using Improved Deep Learning and Personalized Framework



Influence of Building Recognition of High-point Monitoring Image by the Optimized Faster R-CNN on Urban Planning

Haiyuan Tang and Uge Peng

2250013

<https://doi.org/10.1142/S0218213022500130>

Preview Abstract

Abstract | PDF/EPUB

Regular Papers

No Access

Prediction Method of Human Group Emotion Perception Tendency Based on a Machine Learning Model

Yang Wang, Shaolin Li, Shuchun Li and Fan Zhu

2250014

<https://doi.org/10.1142/S0218213022500142>

Preview Abstract

Abstract | PDF/EPUB

Regular Papers

No Access

Driver Fatigue Detection Using Improved Deep Learning and Personalized Framework

Jinfeng Wang, Shuaihui Huang, Junyang Liu, Dong Huang and Wenzhong Wang

2250024

<https://doi.org/10.1142/S0218213022500245>

International Journal on Artificial Intelligence Tools / Vol. 31, No. 02, 2250024 (2022) | Regular Paper

Driver Fatigue Detection Using Improved Deep Learning and Personalized Framework

Jinfeng Wang, Shuaihui Huang, Junyang Liu, Dong Huang and Wenzhong Wang

<https://doi.org/10.1142/S0218213022500245> | Cited by: 0

Preview

PDF/EPUB

Thats Share

Abstract

In transportation, driver's state directly affects traffic safety. Therefore, an accurate driver's fatigue detection is crucial for ensuring driving safety. Real-time and accurate technology is needed for driver fatigue detection. To address this problem, this article proposes a fatigue detection method based on an improved deep learning and personalized framework. First, clustering is applied to face size, and cluster numbers are used to determine the number of detection layers. Then, the size of anchor boxes is set according to the face size. In the proposed framework, the number of convolutional networks is set according to the principle that the receptive field should match the face size in the predicted feature map. Finally, a variety of fatigue features are learned by minimizing the loss function. In addition, a personalized face fatigue detection method is put forward for building a fatigue detection framework to judge the driver's fatigue status more reasonably. The experimental results show

Figure | Exhibition | Related | Details



Vol. 31, No. 02

Metrics

Downloaded 3 times

History

Received: 4 August 2021
Accepted: 25 December 2021
Published: 31 March 2022

Driver Fatigue Detection Using Improved Deep Learning and Personalized Framework

Jinfeng Wang, Shuaihui Huang, Hao Chen, Zhishen Zheng, Jianbin Ou, Fajian Jiang

College of Mathematics and Informatics, South China Agricultural University, Guangzhou, Guangdong, 510642, China.

Junyang Liu

the Industrial and Commercial Bank of China, Zhuhai, Guangdong, China.

Wenzhong Wang*

College of Economics and Management, South China Agricultural University, Guangzhou, Guangdong, 510642, China.

Abstract

In transportation, drivers' state directly affects traffic safety. Therefore, an accurate driver's fatigue detection is crucial for ensuring driving safety. Real-time and accurate technology is needed for driver fatigue detection. To address this problem, this article proposes a fatigue detection method based on an improved deep learning and personalized framework. First, clustering is applied to face size, and cluster numbers are used to determine the number of detection layers. Then, the size of anchor boxes is set according to the face size. In the proposed framework, the number of convolutional networks is set according to the principle that the receptive field should match the face size in the predicted feature map. Finally, a variety of fatigue features are learned by minimizing the loss function. In addition, a personalized face fatigue detection method is put forward for building a fatigue detection framework to judge the driver's fatigue status more reasonably. The experimental results show that the proposed

*Corresponding author: Jinfeng Wang; Wenzhong Wang
Email addresses: wangjinfeng@scau.edu.cn (Jinfeng Wang),
wangwenzhong@scau.edu.cn (Wenzhong Wang*)

method based on an improved clustering method and local receptive field can improve the detection speed of driver's fatigue while maintaining high detection accuracy. The proposed method can reach 125 fps by using GPU GeForce GTX TITAN, which satisfies the real-time requirement. In addition, the personalized framework can achieve high detection accuracy while keeping acceptable speed. The proposed model can accurately and timely detect driver fatigue, which can help to avoid accidents.

Keywords: Fatigue detection, face recognition, deep learning, convolutional network clustering, personalized framework

1. Introduction

Driving fatigue is one of the main factors causing traffic accidents [1]-[2]. However, in addition to the rule restriction reminder, there has been no direct and effective auxiliary method for driving fatigue detection from the technical point of view. According to the literature on fatigue identification, the early methods mainly depended on hardware equipment [3]-[4]. They required to collect all parameters of objects by different sensors and judged the driver state statement based on mathematical statistics and prior knowledge. Fatigue detection can be performed in many ways, for instance, by testing physiological parameters of drivers or based on the electrocardiogram, electroencephalogram, operation records, or vehicle running parameters. The mentioned data can help to judge the degree of driver fatigue to a certain degree. Although these data can be measured and collected easily by professional devices, it is very hard to collect these data in a real scene.

With the proposal of the PERCLOS theory, the fatigue detection problem has been transformed into the eye's detection problem [5]. In this way, the fatigue detection accuracy has been improved significantly. In past years, researchers focused on eyes' fatigue detection, and many different methods have been proposed [6]. However, with the rapid development of image recognition and computer vision technology, driver fatigue detection through video moni-

toring has become one of the research hotspots [7]. A fatigue detection network named the LittleFace has been proposed to locate a face and classify it into two states: small-yaw-angle state denoted as "normal" and large-yaw-angle state denoted as "distract" [8]. This network can run in real-time on edge computing devices with high accuracy, but it focuses on the angle while ignoring the face expression. However, the micro-expression of a person can present fatigue directly. Hence, face recognition is the key technology for fatigue detection.

At present, face recognition technology can discriminate the identity of people in images. Face alignment technology can locate a series of key points as facial features [9]-[11]. Also, face comparison can measure the similarity between two faces. After a long period of research, some classical algorithms for fatigue detection have been applied to many fields. Deep learning is a representative of learning techniques that can simulate the neural structure of the human brain. The most commonly used deep learning method is a deep neural network. This article combined deep learning with a personalized framework to detect and analyze driving fatigue. The proposed method can extract plenty of features that have not been considered in the previous related research. The proposed method improves the robustness and real-time performance of driving fatigue significantly. The remainder of this paper is organized as follows. Section 2 introduces principles of face detection and recognition. Section 3 proposes a combined fatigue detection model based on improved deep learning with clustering and a personalized framework. Section 4 presents the experimental analysis of the proposed method. Finally, Section 5 concludes the paper.

2. Face detection and recognition

The primary task of driver's fatigue detection is to detect a face from an image, crop the face, and recognize the driver's state. Therefore, how to detect face quickly and accurately from an image is crucial for fatigue detection. Many methods for face recognition have been recently proposed, including Haar features [11]-[13], skin color segmentation [14], and deep learning [15]-[16]. The

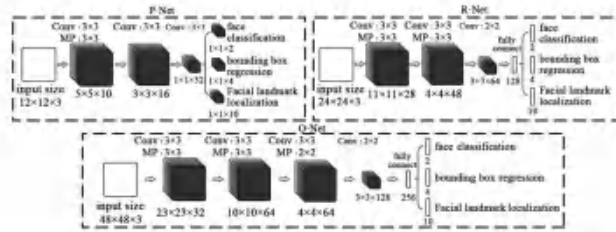


Figure 1: MTCNN network structure.

input information of the face detection is an image, and the output information is a coordinate sequence of a face bounding box (zero face bounding box, one face bounding box, or multiple face bounding boxes). In general, a coordinate frame is a square facing upwards. However, certain face detection methods output a rectangle facing upwards, or a rectangle with a rotation direction.

In recent times, face detection algorithms based on deep learning have become the mainstream detection methods. One of the representative methods is a multi-task cascaded convolutional network (MTCNN) [17]. Compared with the general detection methods based on region-based convolutional neural network (R-CNN) [18], the multi-task convolutional neural network (MTCNN) is more focused on the special task of face detection and has faster speed and higher detection accuracy. The MTCNN represents a multi-task face detection framework, which simultaneously performs face recognition and face feature point detection [19]. The MTCNN uses three convolutional neural networks connected in cascade to implement a framework, which is from coarse to fine, as shown in Fig. 1.

In the field of face recognition and analysis, the detection of face key points is a crucial step. It denotes a prerequisite for other tasks related to detection and recognition, including automatic face recognition, expression analysis, 3D face reconstruction, and 3D animation. In the field of fatigue detection, the face



Figure 2: Face key points data detected by the ASM labeled by number sequence

70 angle can be corrected after obtaining the key points of a face for the purpose
 of better recognition of the fatigue state. From traditional algorithms to deep
 learning-based methods, the research goal has always been to detect the face
 key points accurately and efficiently. The active shape model (ASM) is a classic
 face key point detection algorithm proposed by Cootes [20]. The ASM abstracts
 75 the target object using the shape model based on the point distribution model
 (PDM), as shown in Fig. 2

The main advantage of the ASM algorithm is that it is simple and straight-
 forward. Its architecture is clear and unambiguous. Also, the ASM is easy to
 understand and apply with a strong constraint on the contour shape. However,
 80 its key point positioning method is similar to the exhaustive search, which lim-
 its its operational efficiency to a certain extent. In 2013, Sun et al. applied
 CNN to face key point detection for the first time and proposed a cascaded
 CNN-DCNN, which represents a cascade regression method [21]. By designing
 a cascaded convolutional neural network with three levels, not only the problem
 85 of local optimization caused by improper initialization is solved, but also a more
 accurate key point detection is achieved with the help of CNN's powerful fea-
 ture extraction capability. As shown in Fig. 3, DCNN consists of three levels.
 Level 1 is composed of three convolutional neural networks; Level 2 and level

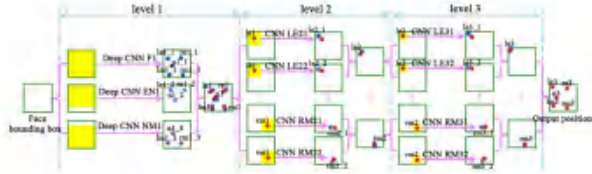


Figure 3: DCNN network structure

3 are composed of 10 convolutional neural networks. Although the efficiency
of DCNN is not too high, it has pioneered the use of neural networks for face
key points detection. In 2016, MTCNN was proposed [22] for not only face
detection but also face key points detection. Moreover, the detection speed has
been greatly improved by using MTCNN.

Face recognition is the core area of face detection technology. For an input
face feature vector, by comparison with the existing vectors in a database, a fea-
ture vector with the smallest distance from the input feature vector is found. If
this distance is less than the predefined threshold, the feature is kept; otherwise,
it is concluded that there is no similar feature in the database.

The FaceNet is a representative face recognition network [23]. The difference
between the FaceNet and the other related algorithms is that it can directly
learn the mapping from face to the feature vector. Then, it can judge whether
two people are the same person according to the Euclidean distance between
the face and feature vector. The FaceNet uses the optimization method of
random gradient descent in the training process. It has also tried other network
architectures, such as GoogleNet. The main purpose is to compare the impacts
of different basic networks on the accuracy of face recognition. The core of the
GoogleNet is the inception structure. Compared with conventional convolution,
the network size of the GoogleNet is unchanged, but the feature extraction
ability is enhanced.

Convolution kernels with different sizes represent different sizes of receptive

fields. Also, the integrated output represents the fusion of features with different scales. This approach effectively improves the feature extraction capability of the network and enhances the recognition accuracy and robustness.

The technology introduced in this section represents the foundation for fatigue detection based on deep learning. The face recognition method can be extended to the personalized fatigue detection model for improving accuracy and real-time performance. The personalized fatigue detection method based on various face-related technologies is described in the next section.

3. Fatigue detection method based on deep learning

This section introduces two methods for face fatigue detection. The first method represents a model based on one-stage face fatigue detection combined with cluster detection, and the second method is a personalized face fatigue detection method.

The end-to-end face fatigue detection process is very simple, as shown in Fig. 4. It considers the fatigue detection problem as a target detection problem. The target detection network processing can be performed using a classical deep learning algorithm, such as Fast-RCNN [23], single-shot multi-box detector (SSD) [25], or You Only Look Once (YOLO) [26]. This section discusses how to select a proper detection model and improve the model structure to achieve better detection performance. Personalized facial fatigue detection can be added to the end-to-end model to further refine each step of fatigue detection, as shown in Fig. 5. Although it is not comparable to end-to-end fatigue detection regarding real-time performance, it is more comprehensible, and it can improve the details in each step.

3.1. End-to-end facial fatigue detection

The end-to-end method can merge face detection and fatigue detection in the same network. The end-to-end face fatigue detection includes one-stage and two-stage face fatigue detection algorithms, which are explained in detail in the following.

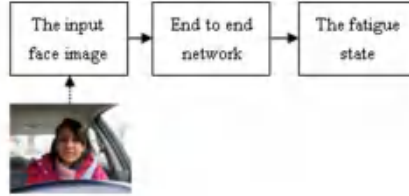


Figure 4: End-to-end fatigue identification.

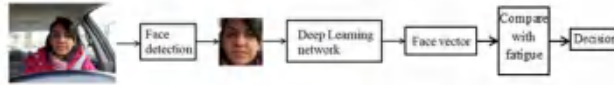


Figure 5: Face fatigue detection by steps with personalized features.

140 *3.1.1. Two-stage and one-stage face fatigue detection algorithms*

The two-stage face fatigue detection algorithm consists of two parts, i.e., stages. First, a candidate box (region proposal) is generated, and then the candidate regions are classified. The R-CNN series based on region proposal is the typical representative of these algorithms, such as R-CNN [18] and Faster-RCNN [27].

150 The process of two-stage face fatigue detection is shown in Fig. 6. The input image is processed by the convolution and pooling layers in turn to obtain the feature map. Then, the candidate regions are generated by the region proposal network (RPN) and divided into two parts, background and target. To obtain a preliminary prediction of the target location, the region of interest (ROI) pooling is used to adjust the candidate area from the corresponding position of the feature map to a uniform size. Then, the feature vector is obtained by a fully connected (FC) layer. Finally, the candidate target category and position are determined by two branches of classification and regression.

155 Compared with the two-stage algorithm, the one-stage algorithm does not require the RPN. It can directly generate the target category probability and

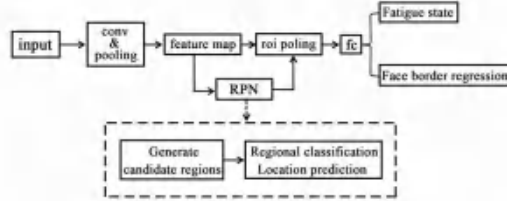


Figure 6: The block diagram of the two-stage face fatigue detection algorithm

Table 1: Comparison of detection performance between one-stage method and two-stage method

Algorithm	Accuracy(%)	Time(ms)
Two-stage face fatigue detection	94.6	203.84
One-stage face fatigue detection	92.9	54.88

position coordinates. The final detection result is obtained after one detection.

The comparison of the YOLO [23] and Faster-RCNN [28], as representatives of the one-stage algorithms and two-stage, respectively, is presented in Table 1.

As shown in Table 1, the recognition accuracies of the two models are similar, but the testing time of the one-stage fatigue detection algorithm is three times shorter than that of the two-stage fatigue detection algorithm. Since driver fatigue detection requires high real-time performance, the proposed model is based on a one-stage framework to meet the real-time requirement.

3.1.2. One-stage method based on cluster detection

The CNNs have achieved good results in image classification because they can extract rich features using multiple convolutional layers. Compared with the performances of a multi-layer perceptron network with weight sharing, the number of parameters and the calculation cost of CNN are greatly reduced. In addition, the convolutional layer uses a sliding window to scan the input layer without conducting the vectorization. In this way, the geometric features of the

input layer are retained, so the resulting feature map is more meaningful.

In a fatigue detection scenario, in addition to high accuracy of fatigue detection, fast detection speed is also required. Namely, only real-time detection can play an early warning role. Most of the existing convolutional neural networks improve detection accuracy by deepening the number or breadth of the network. However, as the network structure becomes more complex, the network computation time also increases. Therefore, this article makes improvements based on the one-stage algorithm.

(1) *Feature extraction network based on receptive field.* A feature extraction network is designed based on the receptive field, which can effectively extract facial features and reduce the number of network layers, as shown in Fig. 7.

In Fig. 7 the calculation part in which the feature map is unchanged is called a block. Down sampling is performed between every two blocks. The purpose of down sampling is to reduce the number of network parameters for the purpose of compressing the network. Instead of the traditional pooling method, this article uses convolutional layers for down sampling, because convolutional layers can strengthen the learning ability of the network. At the same time, the receptive field is used to design the feature network for optimization. The receptive field and step size of the final feature map of the network are respectively calculated by:

$$s' = s_0 + s_1 \quad (1)$$

$$k' = s_0 * (k_1 - 1) + k_0 \quad (2)$$

in which s_0 is the original step size, which is initialized as one, k_0 denotes the size of the convolutional kernel, s_1 is the step size of convolution, s' is the step size of each feature compared to the original image, k_1 is the receptive field of the previous image, and k' is the final receptive field.

After five blocks of feature extraction network, the size of the feature map becomes $19 \times 19 \times 512$. The receptive field of the feature map after the first convolution is 3×3 . Since the receptive field of a low-level convolutional layer is

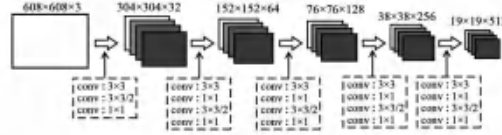


Figure 7: Feature extraction network(In the figure, conv 3×3 represents a 3×3 convolutional layer with a step size of one, conv $3 \times 3 / 2$ represents a 3×3 convolutional layer with a step size of two, and the size of the feature map after this layer is half of the original; and conv 1×1 represents a 1×1 convolutional layer with a step size of one.)

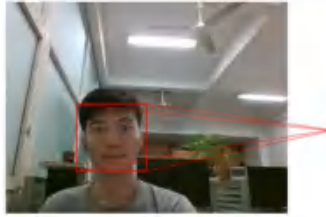


Figure 8: The receptive field in the original image

small, the network can learn only low-level features, such as lines. After each 3×3 convolution, the receptive field of the feature map is increased. The feature extraction network performs 3×3 convolutions 10 times, and the receptive field of the final feature map is 145×145 with a total step size 32. The network structure should be adapted to the recognized objects. Fatigue detection needs to integrate face characteristics so that the final receptive field of the network matches the face size, as shown in Fig. 8

(2) *Cluster detection framework.* In addition to designing feature extraction networks combined with the receptive fields, a cluster detection framework (detection with cluster-DWC) is introduced to improve the detection speed further. Cluster detection refers to the Chinese Whisper (CW) cluster analysis of faces before setting the prior frame [28]. After obtaining the category number of

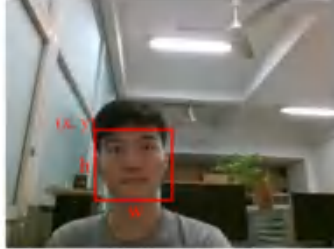


Figure 9: Annotated samples

the face size, the detection layer is selected according to the detection object. Finally, the size of the prior frame is set.

The CW clustering can search for the number of categories of face samples with unknown classification number and perform fast clustering. In the initialization phase, the CW algorithm constructs an undirected graph where each node represents a category and calculates the similarity between different nodes. When the similarity exceeds the threshold value, an association edge is formed between two nodes, and the weight value is set to the similarity. In the iterative phase, node a is randomly selected. The node with the largest weight among the adjacent nodes is selected and denoted as b . Then, a is classified into the same category as b . All nodes are traversed, and the iteration is repeated until the maximum number of iterations is reached.

As shown in Fig. 8, the label information contains the upper left corner coordinates (x, y) , and width w and length h of the face frame. The width and length are grouped into a two-dimensional vector (w, h) for clustering. In order to reduce randomness, the CW clustering on (w, h) is performed 50 times. The clustering results are all one, which means the training data have similar face sizes. The prediction process needs only one feature map.

This article adopts the idea of YOLO detection and combines it with the cluster detection framework to improve the real-time performance of fatigue

detection. The YOLO detection does not include the region proposal, and it directly obtains the coordinates of the detection frame, the probability of containing the detection object, and the probability of object category. The two tasks, i.e., detection and classification, are completed by the same network. The final channels number C of the predicted feature map can be computed as follows,

$$C = B * (loc + conf + cls) \quad (3)$$

where loc denotes the coordinate of the detection frame, $conf$ is the location reliability, cls represents the channel class, and B denotes the number of detection frames for each point in the feature map. In practice, the value of loc in equation 3 is commonly set to four, which means that four channels are used to predict the position of the detection frame. The predicted values of four channels include the upper left corner coordinates, and the length and width of the detection frame. The value of $conf$ is set one, which means that one channel is used to predict the location reliability. This channel is used to predict the probability of containing a face in the detection frame. The value of cls is set as two, which indicates that there are two possible classes to be predicted, and in face fatigue detection, these classes are fatigue and non-fatigue. Further, the number of detection frames B is predicted by each point in the feature map. The final one-stage face fatigue detection structure is shown Fig. 10.

Different loss functions are adopted for different features of the bounding box, and face positioning and fatigue detection are performed simultaneously. The overall network loss is calculated by:

$$loss = \lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] +$$

$$\lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{ij}^{obj} \left[(w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2 \right] +$$

$$\sum_{i=0}^{s^2} \sum_{j=0}^B 1_{ij}^{noobj} BCE(C_i, \hat{C}_i) +$$

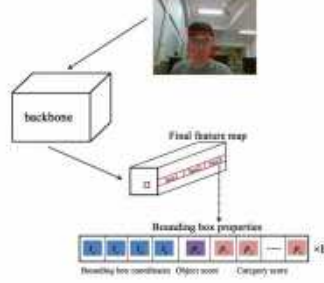


Figure 10: The one-stage face fatigue detection structure, where loc equals four, corresponding to four coordinates t_x , t_y , t_w and t_h ; $conf$ equals one due to object score p_0 ; cls corresponds to C in the category score

$$\lambda_{nobj} \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{ij}^{nobj} BCE(C_i, \hat{C}_i) + \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{ij}^{obj} \sum_{c \in classes} BCE(\hat{p}_{ij}(c), p_{ij}(c)) \quad (4)$$

where S^2 represents the number of grids in the final feature map; 1_{ij}^{obj} equals one
 260 if the object is present in the i th grid; BCE represents the binary cross-entropy
 loss function; C_i denotes the confidence of the network output and estimated
 values; λ_{coord} and λ_{nobj} are the weights of the loss function in different states,
 i.e., face coordination state and confidence loss state; $p_{ij}(c)$ denotes the mark
 265 of classes and its estimated value; lastly, \hat{x}_i , \hat{y}_i , \hat{w}_i , \hat{h}_i , \hat{C}_i and $\hat{p}_{ij}(c)$ correspond
 to the estimated values of elements in the prediction frame.

In equation 4, the first and second terms represent the coordinate error in
 face positioning, the third and fourth terms represent the loss of object confi-
 dence, and the last term denotes the classification error of fatigue detection.
 The pixels in the final feature map are called cells, and each cell predicts B
 270 bounding boxes. Also, each loss function constitutes the final loss function.

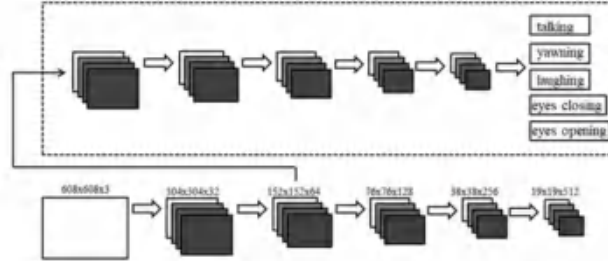


Figure 11: The network structure with the auxiliary network. The auxiliary network is marked by a dotted box. The auxiliary network can judge face features, such as talking, yawning, and laughing, and it is used only in the training process.

3.2. Personalized facial fatigue detection

Compared to the end-to-end face fatigue detection method, each step of personalized fatigue detection is more clear and controllable, and each step can be optimized. In the proposed face fatigue identification method, an MTCNN is used for face detection, and an auxiliary network is used for fatigue detection.

A multi-tasking network with an auxiliary network is designed, and it is displayed as Fig. 11 where the auxiliary network is in the dotted box, and the backbone network is outside the dotted box. The auxiliary network is used to judge features of the human face, and it is used only during the network training process. The final output of the backbone network is the face feature vector.

A triple loss function is used to train the network, as shown in Fig. 12. As mentioned above, the network output is the face feature vector, which is a more intuitive and reasonable method than extracting the intermediate result as a face feature vector in the classification network.

Both positive and negative images need to be input during training. Initially, an image to be tested cannot correspond to the correct fatigue state very well. The training purpose is to reduce the distance between the image to be tested and the correct image and to increase the distance between the sample to be



Figure 12: The training process of triple

tested and the wrong images as much as possible.

280 Unlike the end-to-end face fatigue detection, the personalized face fatigue detection requires three feature vectors in the testing phase, which are the face feature vector to be tested, the fatigue state vector of a driver, and the non-fatigue state vector of the same driver. The triple loss function for network training can be defined as follows:

$$\|f(x_i^o) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^o) - f(x_i^f)\|_2^2 \quad (5)$$

285 $\forall (f(x_i^o), f(x_i^f), f(x_i^p)) \in T$

$$L = \sum_i^N [\|f(x_i^o) - f(x_i^p)\|_2^2 - [\|f(x_i^o) - f(x_i^f)\|_2^2 + \alpha]_+ \quad (6)$$

Equation 5 gives the target of fatigue identification. The distance between the predicted sample and the real class needs to be larger than the distance between the testing sample and the wrong sample, which is equal to α , where α represents a hyper-parameter that was preset to 0.1 in the experiments. The key to the triple loss function is the choice of triples. In equation 5 and 6 x_i^o represents the training sample, and x_i^f represents the same state of the same driver (fatigue or non-fatigue) so that $\operatorname{argmax}_{x_i^p} \|f(x_i^o) - f(x_i^p)\|_2^2$; x_i^p represents different states of the same person so that $\operatorname{argmax}_{x_i^f} \|f(x_i^o) - f(x_i^f)\|_2^2$. Equation 6 expresses the form of the final triple loss function.

290

305 4. Experimental results and analysis

In this article, the one-stage face fatigue detection and personalized fatigue detection based on the DWC are proposed. The effectiveness of the proposed algorithm is verified by experiments regarding two indicators, fatigue detection accuracy and fatigue detection time.

310 4.1. DWC effectiveness verification

The one-stage face fatigue detection based on the DWC used the images of both fatigue and normal states of a face in training. The training data consisted of two parts, online collection and self-built. The self-built part included a total of 800 images of various fatigue and non-fatigue states obtained from 315 the video stream. In addition, 200 images that met the input requirements of the experiments were collected from the Internet. The experimental platform included the Intel Core i5-4460@3.20GHz processor, RAM16GB memory, 12GB video memory, and GeForce GTX TITAN graphics card. The input image resolution was 640×480 . The image size was unified to 608×608 before 320 inputting to the network. The initial learning rate was set to 0.001, and the maximum number of iterations was set to 10000. Before each iteration, a new training image was generated by changing the saturation, exposure, and hue of the image. In order to allow the algorithm to learn a variety of fatigue expressions, the training set contained a variety of behaviors that characterize 325 fatigue, including the normal state of wearing glasses, not wearing glasses, and facial expressions such as yawning and napping.

The feature extraction network was trained using VOC2007 [29] and VOC2012 [30] datasets, and then the entire detection network was training using the arranged training set consisting of 1000 images. As shown in Table 2 the 330 pre-training caused severe overfitting when the training data were insufficient. Therefore, transfer learning was adopted to accelerate network convergence. The loss curve after transfer learning is presented in Fig. 13 where it can be seen that the initial loss was very low. Namely, the transfer learning not only

Table 2: Comparison of Transfer Learning with and without Pre-training

VOC	Pro-training	Training accuracy(%)	Test accuracy(%)
On		98.0	13.5
Off		94.5	94.0

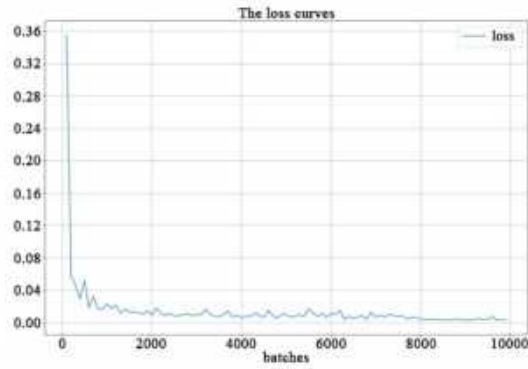


Figure 13: The training process of triple

strengthened the generalization ability of the network but also accelerated the convergence of the network.

In training, both positive and negative samples were fed to the network input. In order to meet the real-time requirement of fatigue detection and reduce network redundancy, a cluster detection framework (DWC) was developed. The feature extraction network part and target detection part were optimized. The intersection over union (IOU) curve obtained by the one-stage face fatigue detection algorithm is presented in Fig. 14. The IOU curve represents the positioning accuracy of the algorithm. The detection accuracy of the detection layer was very high after clustering, and the IOU value was basically above 0.8. After

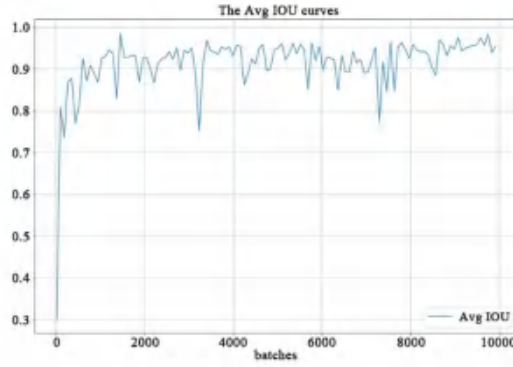


Figure 14: The IOU curve obtained by the one-stage face fatigue detection algorithm.

clustering the face width and height (w, h) in the training sample into one class, the average values of w and h were obtained. The average value of w was 138.8 pixels, and the average value of h was 143.3 pixels. The minimum value of w in the dataset was 119 pixels, and the minimum value of h was 119 pixels.

To verify the effectiveness of the clustering detection framework, YOLOv3 [31] was compared with the model-YOLOv3* that includes clustering. The YOLOv3 originally had three target detection layers, and YOLOv3* represents an improved end-to-end network based on YOLOv3. In YOLOv3*, only one large target detection layer is retained, one layer with low efficiency is removed, and other layers are simplified. Due to these structural changes, the network processing time is reduced by 17% while maintaining an unchanged accuracy level. The results are shown in Table 3.

Although the proposed algorithm has only one detection layer, the detection ability is maintained at the same level. The reason is that this detection layer is obtained through cluster analysis, and the preset frame size of the detection

Table 3: The effect of cluster detection of different algorithms

Algorithm	AvgIOU	Time(ms)
YOLOv3	0.881	55.93
YOLOv3*	0.886	23.86

Table 4: Comparison of the proposed model and the traditional algorithms

Algorithm	Accuracy(%)	Time(ms)
Faster-RCNN	97.0	207.36
MTCNN+CNN	98.5	37.32
DWC	96.2	8.10

layer is set reasonably. Thus, the effectiveness of the cluster detection framework
 180 is verified.

The two-stage and one-stage methods were compared to verify the performance of the proposed one-stage face fatigue detection algorithm. The Faster-RCNN and MTCNN+CNN were used as representatives of the two-stage detection algorithms, while the YOLOv3 and the proposed algorithm based on the
 185 DWC were used as representatives of the one-stage detection algorithms.

As shown in Table 4, the detection accuracy of the two-stage methods was higher than that of the one-stage methods. The Faster-RCNN algorithm included the RPN and CNN. In the first stage, the Faster-RCNN sent the data to the RPN to obtain the ROI that might contain the face. In the second stage,
 190 the ROI was sent to the CNN to obtain the result. The MTCNN could quickly detect the driver's face. In addition to detecting human faces, it could also detect eyes and mouths. The MTCNN+CNN detected and cropped faces by using the MTCNN in the first stage, and then identified the cropped faces by using the CNN in the second stage. This method could separate faces from the
 195 background, and it obtained the highest recognition accuracy. The YOLOv3 contained 53 convolution layers, which constituted the feature extraction network Darknet53, a large target detection layer, a middle target detection layer,

Table 5: Comparison of different fatigue identification algorithms on the self-built test dataset

Algorithm	Accuracy(%)	Time(ms)
fatigue detection based on ASM	92.57	45.03
MTCNN+HOG+randomferns	94.1	14.20
MTCNN+MultiHPOG+SVM	94.4	48.67
MTCNN+MSP-Net	95.3	11.90
DWC-based one-stage algorithm	95.3	11.90

and a small target detection layer. The detection speed of the YOLOv3 was three times faster than that of the Faster-RCNN. The proposed algorithm used an optimization network with the receptive field in the feature extraction stage and CW clustering in the detection stage to reduce redundancy. The detection accuracy of the proposed model was the same as that of the YOLOv3 and slightly lower than those of the two-stage algorithms. The detection speed of the proposed algorithm was 4.5 times faster than that of the YOLOv3. The reason was that the receptive field in the proposed algorithm was more reasonable, and the detection layer obtained by clustering was more efficient.

4.2. Comparison of one-stage method and classical methods

The proposed one-stage algorithm based on DWC was compared with several existing algorithms in detecting fatigue based on facial features on the self-built dataset, CASIA-FACEV5 dataset [32], and YawDD dataset [33]. The number of test images was 200. The examples of images of the three datasets are displayed in Fig. 13. The fatigue detection algorithm based on the ASM combined with Haar cascade features [34] achieved successful recognition of eye fatigue state. The MTCNN + HOG + random ferns [35], MTCNN + MultiHPOG + SVM [36], and MTCNN + MSP-Net [37] used the MTCNN to detect eyes and mouth and then judged the fatigue state according to the closing of eyes and mouth. The proposed one-stage algorithm based on the DWC integrated all information on the face. The comparison results are shown in Tables 5-7

Table 6: Comparison of different fatigue identification algorithms on the CASIA-FACEV5 dataset

Algorithm	Accuracy(%)	Time(ms)
fatigue detection based on ASM	92.2	45.76
MTCNN+ HOG+ random ferns	95.2	14.63
MTCNN+ MultiHPOG+SVM	94.4	48.32
MTCNN+ MSP-Net	95.2	12.14
DWC-based one-stage algorithm	96.8	8.05

Table 7: Comparison of different fatigue identification algorithms on the YawDD dataset

Algorithm	Accuracy(%)	Time(ms)
fatigue detection based on ASM	91.5	46.09
MTCNN+HOG+random ferns	94.1	15.48
MTCNN+MultiHPOG +SVM	93.3	47.77
MTCNN+MSP-Net	94.7	12.71
DWC-based one-stage algorithm	96.2	8.25



Figure 15: Example of test data with and without fatigue. Fatigue states are with red labels, and normal ones are with green labels.

The comparison results on the three datasets show that the proposed algorithm has higher accuracy and efficiency than the other algorithms. Most data in the CASIA-FACEV5 dataset were non-fatigue samples. There were no fluctuations in the accuracy of any of the algorithms. In the YawnDD dataset, there were some samples representing speaking and yawning states, which decreased the detection accuracy due to the presence of mouth closure. This was because the fatigue detection algorithm focused only on these facial features. When the considered face features (eyes or mouth) were near to the fatigue threshold, the recognition accuracy decreased, but the proposed one-stage algorithm still maintained a relatively stable accuracy.

However, when all testing data were used to test the proposed one-stage method based on the DWC, the accuracy decreased significantly, as shown in Fig. 16.

This was because more different characters appeared as the number of testing samples increased. Drivers showed different forms when they were tired, so the fatigue detection algorithm could not be performed well using the same judging

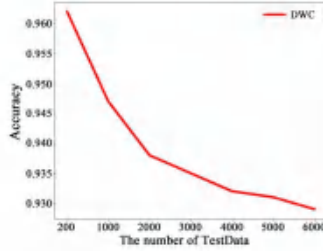


Figure 16: Example of test data with and without fatigue. Fatigue states are with red labels, and normal ones are with green labels.

Table 8: Auxiliary Networks

Auxiliary networks	Accuracy(%)	Time(ms)
On	94.3	35.12
Off	96.9	35.66

415 criteria. To solve this problem, a personalized fatigue detection algorithm was developed. The realization of personalized fatigue detection was mainly achieved by changing the training parameters, using the triple loss function, and adding the auxiliary network.

The results of the ablation experiment are given in [8](#). In the training phase, 420 the test accuracy of the model decreased by 2.6% after removing the auxiliary network. The auxiliary network was used to determine whether a person in an image opened eyes, talked, or yawned. When there was no auxiliary network, it was more difficult for the triplet to converge, which resulted in a decrease in the detection accuracy. Since the auxiliary network did not participate in the calculation during the testing phase, the detection time stayed unchanged. As 425 shown in [9](#) the test accuracy of the network with the triple loss function was higher than that of the network with the cross-entropy loss function.

The personalized fatigue detection expressed the driver’s face by using a

Table 9: Comparison of Loss Functions

Loss function	Accuracy(%)	Time(ms)
Cross entropy loss function	93.5	35.24
Triple loss function	96.9	35.66

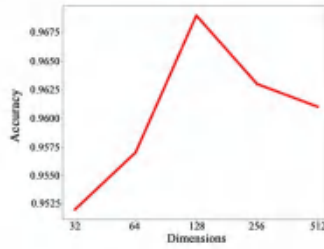


Figure 17: Influence of the face vector dimension on the detection accuracy

vector. Compared with the fatigue vector, it could judge whether the object
 430 was in the fatigue state. The experimental analysis of the vector dimension
 indicated that 128-dimensional vectors represented the human faces the best,
 as shown in Figure 18.

4.3. Comparison of two proposed models

Finally, the two proposed fatigue detection algorithms, i.e., the one-stage
 435 face fatigue detection algorithm based on the DWC and the personalized fatigue
 detection algorithm, were compared. The training dataset was the same as that
 used in the previous experiments. The results are shown in [10](#). Although the
 detection time of the end-to-end fatigue detection algorithm was shorter, the
 stepwise fatigue detection could also meet the real-time requirements, and the
 440 accuracy was higher.

The presented experimental results show that the proposed one-stage face
 fatigue detection algorithm based on the DWC can identify fatigue rapidly.
 Meanwhile, the proposed personalized fatigue detection algorithm can identify

Table 10: Comparison of the end-to-end fatigue identification algorithm and the personalized fatigue identification algorithm.

Algorithm	Accuracy(%)	Time(ms)
DWC-based one-stage recognition	92.9	8.38
Personalized fatigue detection algorithm	96.9	35.66

the fatigue of the face with different characteristics, thus achieving the personalization property.

5. Summary

This article proposes a driver fatigue detection method based on the one-stage target detection algorithm using optimized clustering and the receptive field. Also, an efficient feature extraction network is designed, and the detection network is simplified. The experimental results show that the proposed detection method improves detection efficiency while ensuring high detection accuracy. Moreover, a personalized fatigue detection algorithm is proposed. During the training phase, the triple loss function is used to make the network learn differences between the fatigue and non-fatigue states of the same person. At the same time, the auxiliary network is introduced to refine the facial features further. The experimental results show that the detection speed of the proposed personalized fatigue detection algorithm is not as fast as that of the proposed one-stage face fatigue detection algorithm, but it can meet the real-time requirements while improving the recognition accuracy and providing a more reasonable design.

In future work, more information from more views will be used to judge the fatigue degree of a driver. Furthermore, a warning system will be designed to keep a driver safe based on the fatigue detection result.

Acknowledgment

485 This work was supported in part by the Guangzhou Key Laboratory of Smart
Agriculture (201902010081), the Science and Technology Planning Project of
Guangdong Province (No.: 2017A040406023) and the Science and Technology
Planning Project of Guangzhou City (No.: 201804010353). We thank LetPub
(www.letpub.com) for its linguistic assistance during the preparation of this
490 manuscript. We would also like to thank Ms. Linqi He in South China Agricul-
tural University for her support in improving the quality of figures.

References

References

- [1] Jap B.T., Lal S., Fischer P., Bekiaris E., Using EEG spectral components
495 to assess algorithms for detecting fatigue, *Expert Systems with Applications*,
36(2), 2352-2359(2009).
- [2] Pylkkonen M., Sihvola M., Hyvarinen H.K., Puttonen S., Hublin C., Sallinen
M., Sleepiness, sleep, and use of sleepiness countermeasures in shift-working
long-haul truck drivers. *Accident Analysis and Prevention*, 80, 201-210(2015)
- 498 [3] Chowdhury R.S., Kavakli M.H., Sensor Applications and Physiological Fea-
tures in Drivers' Drowsiness Detection: A Review, *International Research
Training Program*, 18(8), 3055-3067(2018)
- [4] Chellappa R.E., Fatigue Detection Techniques: A Review. *International
Journal of Pure and Applied Mathematics*, 117(16), 503-510(2017)
- 499 [5] De la Torre, F., Rubio, C. J. G., Martínez, E., Subspace eyetracking for
driver warning, In *IEEE International Conference on Image Processing*, 3,
329-332(2003)
- [6] Mohsin H., Abdullah S. H., Pupil detection algorithm based on feature ex-
traction for eye gaze. *Proceedings of the 2017 6th International Conference on*

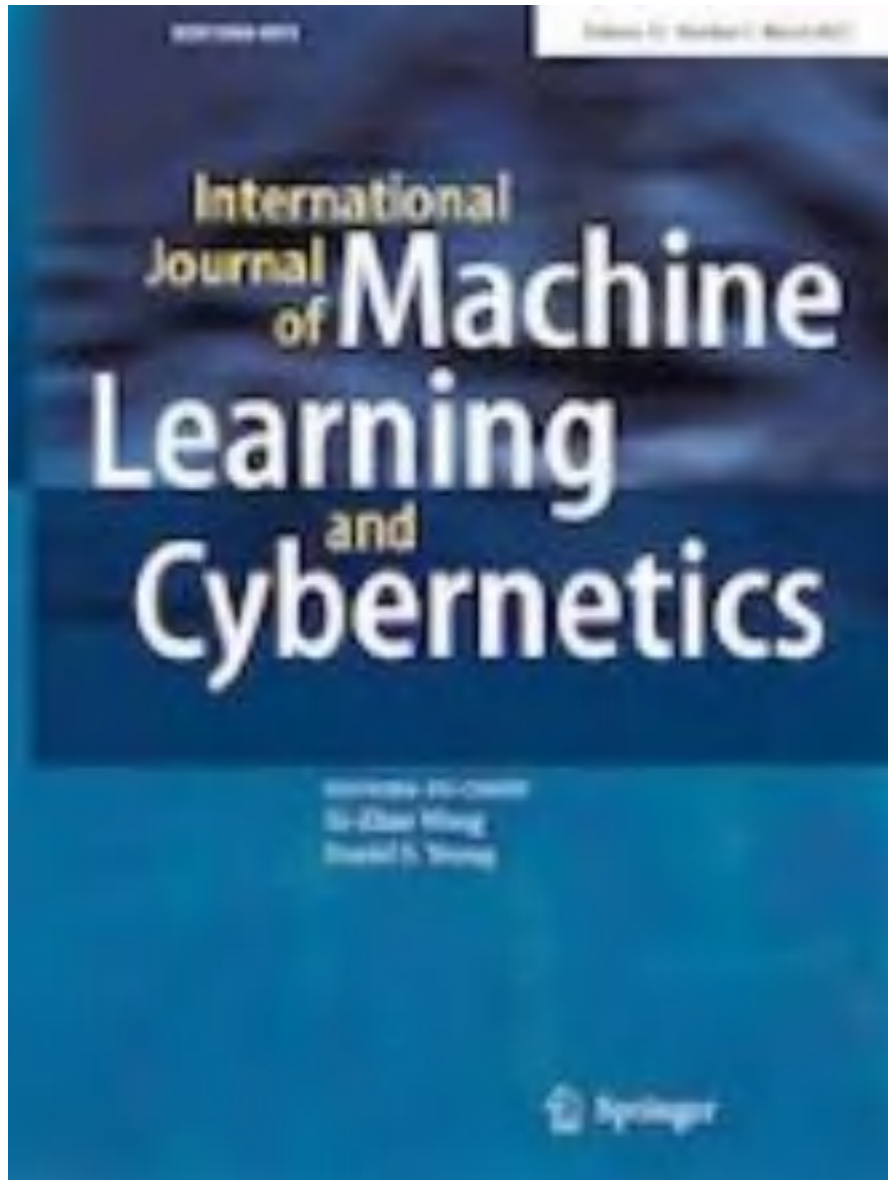
- 490 Information and Communication Technology and Accessibility(ICTA), IEEE Press (2017)
- [7] Nisha Yadav, Kakoli Banerjee, Vikram Bali, A Survey on Fatigue Detection of Workers Using Machine Learning, International Journal of E-Health and Medical Communications (IJEHMC), IGI Global,11(3),1-8(2020)
- 495 [8] Li X , Xia J , Cao L , et al, Driver fatigue detection based on convolutional neural network and face alignment for edge computing device. Proceedings of the Institution of Mechanical Engineers Part D Journal of Automobile Engineering, 2021:095440702199948.
- [9] Kai T. W. C., Jian L. H., Fatigue reliability analysis and life bench 500 test of buffer block in car damper. Proceedings of the IEEE International Conference on Engineering Technology and Innovation, IEEE Press. 10.1109/ICE.2018.8436252(2018).
- [10] Kashyap V. S., Drowsiness detection system using Matlab,International Journal of Advance Research in Science and Engineering, 6(5), 689–695(2017)
- 505 [11] Moghimi M M, Nayeri M, Pourahmadi M, et al, Moving Vehicle Detection Using AdaBoost and Haar-Like Feature in Surveillance Videos, International Journal of Imaging and Robotics, 18(1),94-106(2018)
- [12] Wang S , Wen G , Cai H, Research on face detection based on fast Haar feature, 2017 10th International Congress on Image and Signal Processing, 510 BioMedical Engineering and Informatics (CISP-BMEI), Shanghai, China, 1-6(2017)
- [13] Luo A, An F, Zhang X, et al., A Hardware-Efficient Recognition Accelerator Using Haar-Like Feature and SVM Classifier, IEEE Access, 7, 14472-14487(2019)
- 515 [14] Naji S A, Zainuddin R, Jalab H A, Skin segmentation based on multi pixel color clustering models. Digital Signal Processing, 22(6),933-940(2012)

- [15] Sun X, Wu P, Hoi S C H, Face Detection using Deep Learning: An Improved Faster RCNN Approach. *Neurocomputing*, 299,42-50(2018)
- [16] Zheng Y, Zhu C, Luu K, et al. Towards a deep learning framework for unconstrained face detection, 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), (2016)
- [17] Zhang K, Zhang Z, Li Z, et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters*, 23(10):1499-1503(2016)
- [18] Jiang H, Learned-Miller E, Face Detection with the Faster R-CNN, 650-657(2016).
- [19] Shi, Wenfeng , Li Jun, Yang Yuanjie, Face Fatigue Detection Method Based on MTCNN and Machine Vision. *International Conference on Applications and Techniques in Cyber Intelligence*, 233-240(2020)
- [20] Cootes T F, Taylor C J. Using grey-level models to improve active shape model search//*International Conference on Pattern Recognition*, 1, 63-67(1994).
- [21] Yi S, Wang X, Tang X. Deep Convolutional Network Cascade for Facial Point Detection//*2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Portland, OR, USA, 3476-3483 (2013)
- [22] Zhang K, Zhang Z, Li Z, et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters*, 23(10):1499-1503(2016)
- [23] Schroff Florian, Kalenichenko Dmitry, Philbin James, FaceNet: A Unified Embedding for Face Recognition and Clustering//*2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 815-823(2015).
- [24] Girshick R, Fast R-CNN, *Computer Science*(2015)

- [25] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector. European Conference on Computer Vision Springer, Cham,21-37(2016).
545
- [26] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016
- [27] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6),1137-1149(2017).
540
- [28] Biemann C. Chinese whispers: An efficient graph clustering algorithm and its application to natural language processing problems: Proceedings of the first workshop on graph based methods for natural language processing, 73-80(2006)
535
- [29] Everingham M, Zisserman A, Williams C K I, et al, The 2005 PASCAL visual object classes challenge//Quinonero-candela J, Dagan I, Magnini B, et al., Lecture Notes in Artificial Intelligence. 117-176(2006).
- [30] Everingham M, Van Gool L, Williams C K I, et al. The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 88,303-338(2010).
530
- [31] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement, arXiv e-prints, 2018.
- [32] CASIA-FaceV5, Chinese Academy of Sciences-Institute of Automation (CASIA), <http://biometrics.idealtest.org/>. Acessado em: 21 de janeiro de 2012.
525
- [33] Abtahi S, Omidyeganeh M, Shirmohammadi S, et al. YawDD: a yawning detection dataset. 2014. DOI :10.1145/2557642.2563678
- [34] Cuimei L, Zhiliang Q, Nan J, et al. Human face detection algorithm via Haar cascade classifier combined with three additional classifiers, 483-487(2017).
520

- [35] Dong Y, Yan Z, Yue J, et al. Comparison of random forest, random ferns and support vector machine for eye state classification. *Multimedia Tools and Applications*, 75(19):11763-11783(2015).
- 325 [36] Song F, Tan X, Liu X, et al. Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients. *Pattern Recognition*, 47(9):2825-2838(2014) .
- [37] Ji Y, Wang S, Zhao Y, et al. Fatigue State Detection Based on Multi-Index Fusion and State Recognition Network. *IEEE Access*,7:64136-64147(2019).

2.5. FF-GLAM-cs: a fusion framework based on GLAM with channel shuffle for speech emotion recognition





FF-GLAM-cs: a fusion framework based on GLAM with channel shuffle for speech emotion recognition

Jinfeng Wang¹ · Zhishen Zheng¹ · Yong Liang² · Jing Qin³ · Wenzhong Wang⁴

Received: 4 January 2023 / Accepted: 11 August 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

With the support of artificial intelligence, speech emotion recognition is integrated into people's daily lives in the form of smart speakers. Some high-accuracy models have poor applicability due to their large size. However, the accuracy of lightweight models is unsatisfactory. In this article, an integrated framework based on Global-Aware Multiscale with channel shuffle (FF-GLAM-cs) is proposed, fusing multiple lightweight models to ensure a small size and high accuracy. Channel shuffle is added to solve the computational duplication caused by multiscale convolution. In addition, a fuzzy integral fusion method is adopted that describes the interactions of classifiers ignored by traditional methods. The impact of different combinations of classifiers is analyzed. In experiments, the performance of the new model and the effect of fusion are verified by the model to four speech emotion datasets. The model is validated and analyzed with parameter and ablation testing, kernel model comparisons and fusion verification. The results show that FF-GLAM-cs is superior to state-of-the-art methods in terms of accuracy and efficiency. In particular, the fusion module presents excellent improvements in accuracy. The source code of this work is available at <https://github.com/bishen33/GLAM-cs-and-FI-for-SER.git>.

Keywords Speech emotion recognition · Multiscale features · Fuzzy integral · Deep learning

1 Introduction

The rapid development of artificial intelligence technology has led to significant advances in speech processing. The phrase "Hi, Siri" is the first impression that many users have with intelligent speech interaction. With a simple command, a machine can provide services such as music, home appliance control or user information lookup. Therefore, speech interference technology is becoming increasingly popular among users because it is the easiest and most convenient way for people to communicate. While many smart speakers are coming into thousands of homes and providing services for people, many problems have been exposed. Many users reflect that the smart speaker can only listen and execute commands but cannot understand the needs of emotions the way people do. However, emotion plays an important role in human-to-human communication.

Speech emotion recognition (SER) describes a computer's attempt to perceive the emotions contained in speech. It is the key to human-robot emotional interaction, enabling various intelligent devices to understand the user's intention to the maximum extent and improve the humanization of robots to better serve humans [1, 2].

Jinfeng Wang
wangjinfeng@scau.edu.cn

Zhishen Zheng
zheng_zhishen@163.com

Yong Liang
2371366625@163.com

Jing Qin
harry.qin@polyu.edu.hk

Wenzhong Wang
wangwenzhong@163.com

¹ College of Mathematics and Informatics, South China Agricultural University, Guangzhou, China

² Peng Cheng Laboratory, Shenzhen, China

³ School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China

⁴ College of Economics and Management, South China Agricultural University, Guangzhou, China

To make machines understand human emotions better, researchers have conducted much research on speech emotion recognition tasks and have made great progress. Since convolutional neural networks have shown excellent performance in picture classification, some researchers have introduced them to speech emotion recognition tasks [3–6]. These studies all transfer speech from the time domain to the frequency domain, extract spectrogram features, treat spectrograms as pictures, use convolutional neural networks for extracting sentiment features in speech, and finally predict the sentiment category. However, as research has advanced, some neural network models have abandoned the consideration of model size and practicality to pursue high accuracy.

Some simple-structured convolutional neural networks [7–9], require fewer computations to meet low-power consumption requirements. However, their accuracy is not satisfactory. The AACNN (Area Attention CNN) [7] and MHCNN (Multifocal Attention CNN) [9] models adopt the same convolutional structure, using two parallel convolutional layers to extract features separately from the time axis and frequency axis. After four consecutive convolutional layers, the attention layer manages the representation and sends the output to the fully connected layer for classification. Global-Aware Multiscale (GLAM) [8] was proposed by adding multiscale convolution to collect receptive fields of different sizes, capturing multiscale feature representations and further improving the feature extraction ability of the model.

However, GLAM is still encumbered by large computations and feature redundancy. Multiscale convolutions are adopted for capturing information from receptive fields of different sizes in computer vision. However, different receptive fields may extract redundant features from speech spectrograms with the same time-frequency unit. Therefore, the multiscale convolution module needs to be enhanced to reduce the computational complexity and boost the accuracy with a fusion mechanism.

Ensemble learning can yield a better and more comprehensive performance by combining multiple weak classification models. Traditionally, ensemble learning uses strategies such as arithmetic averaging and weighted averaging to aggregate the decision structures of multiple weak classification models. However, these strategies do not take into account the correlation of individual classification models. A classifier may perform well on its own. However, when its decisions are combined with those of other classifiers, the performance of the integrated framework may be degraded. Some outliers have a significant impact on the average, which results in much lower accuracy.

To solve this problem, a fusion approach based on a fuzzy integral is adopted to improve the accuracy of the integrated framework. The fuzzy integral takes into account the importance of the combinations of classifiers and uses

fuzzy measures to represent the impacts of subsets of the set of classifiers. Furthermore, the fusions of outlier classifiers should be independent of each other.

The contributions can be summarized as follows:

- Channel shuffle is added to a multiscale convolution module, which not only enhances the representation learning ability of the model but also reduces the computational complexity.
- A fuzzy integral is adopted as the fusion method for integrating multiple models, as it can consider the interactive impacts among multiple classifiers.
- The proposed framework achieves a 1% to 3% improvement in performance over state-of-the-art methods via testing on several benchmark datasets. The effect of the fusion module is also verified with comparisons with traditional methods.

The rest of this paper is organized as follows. Related works are introduced in Sect. 2. Section 3 presents the proposed framework. Section 4 presents the experimental results and discussions. Finally, the conclusions are summarized.

2 Related work

2.1 Speech emotion recognition

In recent years, many researchers have conducted research on deep learning techniques for speech emotion recognition. There are three major types of structures of deep learning applied in SER, i.e., convolutional neural networks (CNNs), recurrent neural networks (RNNs), and Transformer networks (TNs).

In some studies, researchers treated the spectrograms of speech as images and used CNNs to extract their features [3, 5, 6] because of their powerful feature extraction capabilities. Due to the time-series nature of speech data, some researchers have used RNNs to learn related features [3, 5, 10], which have achieved good results. Transformer has a strong performance in text processing, so some researchers have applied TNs to SER and obtained excellent results [11–13]. The success of TN cannot be achieved without an attention mechanism. In processing input information, an attention mechanism can be used to select only some key input information to improve the performance of the neural network.

However, it is also undesirable to pursue high-accuracy models while ignoring the number of model parameters. They cannot meet the real-time requirements of human-machine speech interaction in smart speakers. Some researchers have combined CNNs with attention mechanisms to achieve lightweight models with high accuracy [7,

8, [4], among which GLAM uses a multiscale convolutional module to extract features at different scales. However, the multiscale module of GLAM still has computational redundancy. In this article, the structure of GLAM will be further enhanced to reduce the complexity and boost the accuracy of the model.

2.2 Classifier fusion

The method of combining decisions from multiple weak classification models to obtain better accuracy is called integration learning. This approach can also be applied to combine the decisions of several lightweight models. Some researchers [15, 16] have attempted averaging and voting methods to fuse the decisions of multiple classifiers in speech emotion recognition and achieved good results. However, these fusion methods do not take into account the correlations between classifiers and ignore the effect of outliers. The fuzzy integral has good performance in describing the interaction of multiple factors due to fuzzy measures. Therefore, it was introduced to fusion models.

In fusion methods based on the fuzzy integral, the fuzzy measure expresses not only the importance of a single classifier but also the interaction between classifiers [17]. Fuzzy measures represent the importance of each element based on their relationship with the correct decision. In this article, the Choquet integral is used as a representative of the fuzzy integral. It is adopted as a fusion tool, which is defined as an integration function from X to R^+ with respect to μ as follows [18]:

$$(\sigma) \int \mu dH = \sum_{i=1}^n (f(x_i) - f(x_{i-1})) \mu(A_i) \quad (11)$$

where μ is a fuzzy measure on set X and x_i is an element of X .

The Choquet integral exploits the degree of uncertainty in the decision. The fuzzy measure determines the strength of each classifier and the strength of all possible combinations of classifiers. More specifically, $f(x)$ represents the decision score of the classifier, and μ is the weighted strength of the classifier or combination of classifiers. The fuzzy measure value of each classifier needs to be determined by learning in the model.

Due to the abovementioned advantages of the fusion method based on the fuzzy integral, many researchers have applied it in different fields in recent years, including action recognition [19], COVID-19 detection [20, 21], and cervical cancer detection [22]. In these studies, the researchers verified the performance of multiple deep learning models by adding fuzzy integral fusion. In this paper, we apply the Choquet integral to fuse lightweight models to improve the accuracy of the framework in speech emotion recognition.

3 Proposed method

In this section, a fusion framework based on Global-Aware Multiscale (GLAM) [8] with channel shuffle is proposed, as shown in Fig. 1. The first part of the framework performs audio preprocessing on raw speech data before they are input into the neural network. The purpose is to segment the speech data. Then, the features of MFCCs are extracted, and the Mixup method is used for data enhancement. The second part of the framework is feature extraction by the neural network. The learning ability of the model is enhanced by adopting channel shuffle to adjust the multiscale convolution module. In the final part, the fuzzy integral is adopted to fuse the results of multiple models to advance the accuracy of the framework.

3.1 Feature extraction and augmentation

Feature extraction plays a crucial role in the success of any machine learning model. Appropriate feature selection could lead to a better trained model, while inappropriately selected features would significantly hinder the training process. In the new framework, the Librosa audio library [23] is utilized for feature extraction. Mel-scaled spectrograms and MFCCs (mel-frequency cepstral coefficients) [24, 25] are widely used in the field of sound classification and SER. These features can mimic the pattern of intrinsic sound frequencies a human can perceive to a certain extent. In particular, MFCCs collectively comprise the mel-frequency spectrum, which is defined as the representation of the short-term power spectrum of sound. A visualization of some MFCCs is shown in Figure 2(a). In this work, MFCCs are used as the input to the deep learning model.

Speech emotion data are usually scarce and valuable. Data augmentation can improve the generalizability of the model by increasing the number of samples. Two data augmentation methods are adopted to enrich the generalizability of the deep learning model. The first splits a long audio segment into small, fixed-duration segments, which not only meets the requirement of fixed shape input for deep learning models but also increases the number of data samples. Specifically, the fragment length is set as S and the overlap length is set as O . For an audio segment with length L , a window of size O is moved to generate the next segment after intercepting a segment with length S . Visualization of the audio segmentation process can be seen in Fig. 2b.

The second augmentation method is a kind of Mixup method [26], which can increase the generalizability of the model by performing a simple linear transformation of the input data. Using two pairs of data, (x_1, y_1) and (x_2, y_2) as examples, a new constructed example can be generated:

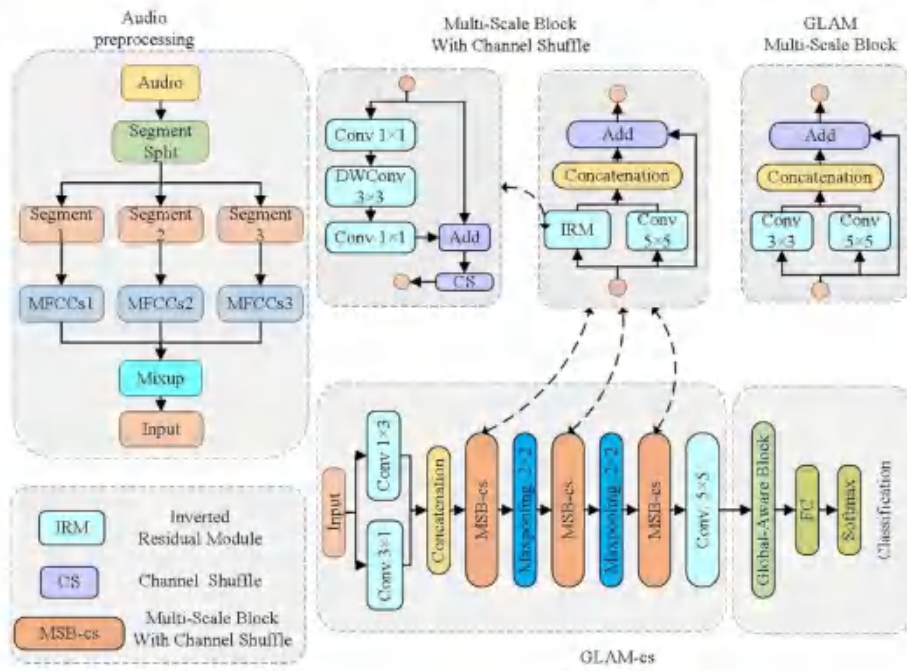


Fig. 1 Model architecture



Fig. 2 (a) Visualization of MFCCs from the EMODB dataset (b) Speech segmentation

$$\bar{x} = \lambda x_i + (1 - \lambda)x_j \quad (2)$$

$$\bar{y} = \lambda y_i + (1 - \lambda)y_j \quad (3)$$

where $(x_i; y_i)$ and $(x_j; y_j)$ are randomly drawn from training data and $\lambda \sim \beta(\alpha; \alpha)$ with $\alpha \in (0, \infty)$. The Mixup method

can impose implicit regularization constraints on the model at the data level.

3.2 Deep feature extraction and model training

The primary goal of deep feature extraction is to extract salient information from MFCCs and represent this information in a lower dimensional space. GLAM [8] adopts a multiscale

convolution block to extract the deep features of MFCCs, in which multiscale feature representations are captured by receiving fields with different sizes. However, convolution kernels of 3×3 and 5×5 may lead to feature redundancy and computational redundancy. Therefore, the 3×3 convolution layer needs to be improved to reduce the number of redundant calculations and boost the feature extraction ability. A new model, GLobal-Aware Multiscale with channel shuffle (GLAM-cs), is proposed. Specifically, an inverted residual module (IRM) with channel shuffle [27] is implemented, as shown in the right lower panel of Fig. 1. More specifically, the 3×3 convolution kernel is replaced with an IRM. The IRM contains two depthwise convolution kernels of sizes 1×1 and 3×3 (DWCov), a residual structure and a channel shuffle operator. The purpose of channel shuffle is to reorganize the channel information of different groups so that the model can capture the information between different groups. The structure of GLAM-cs is different from that of GLAM [8], AACNN [7] and MHCNN [9], which will be used for comparison in later experiments. The detailed settings are listed in Table 1.

Since the Mixup method is used for data augmentation, the loss function must be modified. The labels of two pairs of samples in SER are not transformed by linear weighting, but both labels are used simultaneously in a loss function, which is expressed as:

$$L = \lambda * L_{CE}(y_i, \tilde{y}) + (1 - \lambda) * L_{CE}(y_j, \tilde{y}) \quad (4)$$

where y represents the labels predicted by the model, (y_i, y_j) represents the labels of the pair of examples, L_{CE} is the cross-entropy of the loss function, and λ is the mixed weight.

3.3 Fuzzy integral fusion

In this section, an example will be introduced to specify the process of fuzzy integral fusion [17]. Suppose there are L classifiers for a classification task with C classes, and each classifier outputs a probability vector $d \in R^C$. All classifiers are defined as a set, denoted as $D = \{D_1, D_2, \dots, D_L\}$, and a fuzzy measure μ_c is defined on set D for each class C . Finally, the fuzzy integral fuses the probability vectors output by each classifier on sample x .

Example 1.

For example, consider three classifiers: A , B and C . There will be 8 (2^3) fuzzy measures for the three classifiers which represents the importance value for each element of the power set of D , i.e., $\{\emptyset, \{A\}, \{B\}, \{C\}, \{A, B\}, \{A, C\}, \{B, C\}, \{A, B, C\}\}$. An example of the measure μ for the set of $\{A, B, C\}$ is presented in Table 2. Let X be a subset of the set of classifiers; then, $\mu(X)$ is defined as the importance of all elements in X .

Table 1 Architecture of different model (k: kernel size, s: stride, BN: batch normalization)

Model	MHCNN [9]	AACNN [7]	GLAM [8]	GLAM-cs(Ours)
Stage1	conv1 k=(1,1) s=1 s=1 BN ReLU	conv1 k=(1,1) s=1 BN ReLU	conv1 k=(1,1) s=1 BN ReLU	conv1 k=(1,1) s=1 BN ReLU
Stage2	conv2 k=(2,2) s=1 BN ReLU	conv2 k=(2,2) s=1 BN ReLU	conv2 k=(2,2) s=1 BN ReLU	conv2 k=(2,2) s=1 BN ReLU
Stage3	conv3 k=(3,3) s=1 BN ReLU	conv3 k=(3,3) s=1 BN ReLU	conv3 k=(3,3) s=1 BN ReLU	conv3 k=(3,3) s=1 BN ReLU
Stage4	conv4 k=(3,3) s=1 BN ReLU	conv4 k=(3,3) s=1 BN ReLU	conv4 k=(3,3) s=1 BN ReLU	conv4 k=(3,3) s=1 BN ReLU
Stage5	conv5 k=(3,3) s=1 BN ReLU	conv5 k=(3,3) s=1 BN ReLU	conv5 k=(3,3) s=1 BN ReLU	conv5 k=(3,3) s=1 BN ReLU
Stage6	Multi-head Attention	AreaAttention	gMLP	gMLP
Stage7	Adaptive AvgPool	FC	FC	FC
			ResMultiConv2 MaxPool	ResMultiConv-cs2 MaxPool
			ResMultiConv3 MaxPool	ResMultiConv-cs3 MaxPool
			ResMultiConv8	ResMultiConv-cs4
			conv5 k=(5,5) s=1	conv5 k=(5,5) s=1

Table 2 Example of fuzzy measures for sets $\{A, B, C\}$

X	$\mu(X)$
$\{\}$	0
$\{A\}$	0.1234
$\{B\}$	0.234
$\{C\}$	0.3456
$\{A, B\}$	0.7890
$\{A, C\}$	0.8901
$\{B, C\}$	0.9012
$\{A, B, C\}$	1.000

The three classifiers make predictions on a sample x to yield three probability vectors. The decision profile is composed of three vectors and can be expressed as follows:

$$D = \begin{bmatrix} p_{A1} & p_{A2} & p_{A3} \\ p_{B1} & p_{B2} & p_{B3} \\ p_{C1} & p_{C2} & p_{C3} \end{bmatrix} \quad (5)$$

When computing the final fusion result for sample x , the i^{th} column of the decision profile is treated as a function f_i on D . Then, the Choquet integral of f_i with respect to the fuzzy measure μ_i is computed according to $e_i = \int f_i d\mu_i$. Finally, the class corresponding to the greatest possibility, $R = \text{argmax}(e_i)$, is confirmed as the classification of sample x . The fuzzy measure μ needs to be determined before computing the fuzzy integral. However, the calculation of fuzzy measures is an NP-hard problem [28]. In this work, the fuzzy measures are treated as hyperparameters, and the Optuna framework [29] is used to perform hyperparameter tuning on the validation data. First, the decision profiles of the three models in the validation set and testing set are calculated. Then, the fuzzy measure is calculated using the decision profile in the validation set. Meanwhile, the Choquet integral is used to make fusion decisions in the testing set. Finally, the results are evaluated using the labels of the testing set. This process is shown in Fig. 3. The pseudocode of the fusion module is shown in Algorithm 1.

Algorithm 1 Fuzzy integral fusion algorithm

```

Require: MFCS's datasets  $X$ , Models  $M1, M2, M3$ , Fuzzy Measure  $\mu$ , Number of iterations  $C$ 
1. Feed the input  $X$  into the Models and get the probability  $P$ 
2. for  $i=1$  to  $C$  do
3.    $P_i = M_i(X)$ 
4. end for
5. for  $i=1$  to  $C$  do
6.    $E_i = (P_i \times \mu_{A_i} \times \mu_{B_i})$ 
7.   Get fuzzy integral value  $e_i$  using Eq.(1)
8. end for
9.  $E = [e_1, e_2, \dots, e_C]$ 
10.  $R = \text{argmax}(E)$ 
11. Return  $R$ 
    
```

4 Experiments

In this section, several benchmark datasets are introduced to validate the new framework. Hyperparameter testing will be implemented to determine the settings for data augmentation. Ablation experiments will verify which convolutional layers can be replaced by IRMs. Then, the kernel-proposed model will be compared with state-of-the-art methods to validate the performance improvements. The fusion module based on the fuzzy integral will also be tested. Finally, the fusion framework will be compared to those in several recently published articles to prove the efficiency of the new model.

4.1 Experimental settings

Datasets. CaFE [30] is the Canadian French Emotional (CaFE) speech dataset with 936 utterances. CaFE [30] contains six different sentences pronounced by 12 actors of male or female sex. Six basic emotions plus one neutral emotion are represented in the dataset. Each class is equally represented except for the neutral emotion, which has half of the sentences of other emotions in the dataset. EMO-DB [31] is the Berlin emotional speech database (EMO-DB), which covers seven emotional states: anger, boredom, disgust, fear, happiness,

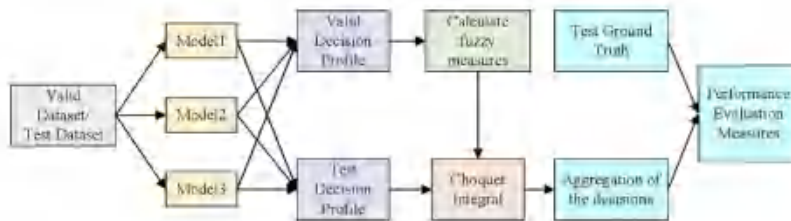


Fig. 3 Fuzzy integral fusion

neutral, and sadness. The verbal contents come from 10 Germans (5 males and 5 females), who provided predefined neutral utterances. EMO-DB [31] consists of approximately 535 sentences from seven emotions. The average duration of all audio files is 3 s. The EMOVO [32] corpus is the first affective dataset for the Italian language. This dataset was established by six professional actors who speak 14 sentences to simulate seven affective states: disgust, fear, anger, joy, surprise, sadness, and neutral. The Interactive Emotive Binary Motion Capture Database (IEMOCAP) [33] was developed by the team of Speech Analysis and Interpretation Laboratory (SAIL) from the University of Southern California (USC). This dataset contains five sessions lasting approximately 12 hours in total. The actors performed the chosen affective scripts and elicited five emotions (happiness, anger, sadness, fusion, and neutral states) under the designed imaginary settings. Finally, important information for all datasets is summarized in Table 3. In particular, 5531 utterances are selected from IEMOCAP as in Ref. [4, 12–14] and labeled as one of five categories: happy, angry, neutral, sad, and excited. Additionally, happy and excited are treated as separated category.

Metrics. To comprehensively understand the predictive performance of the models, we utilize four common metrics, i.e., weighted accuracy (WA), unweighted accuracy (UA), micro F1 score, and macro F1 score. The distinction between WA and UA (and between the micro and macro F1 score) is that the former takes label imbalance into account, while the latter does not.

$$WA = \frac{\sum_{i=1}^N TP_i}{\sum_{i=1}^N TP_i + FP_i} \quad (6)$$

where N represents the total number of categories, i represents the i^{th} category, TP_i represents the number of positive samples predicted as positive classes and FP_i represents the number of samples with negative classes when the predicted values are positive classes.

$$acc_i = \frac{TP_i}{TP_i + FP_i} \quad (7)$$

$$UA = \frac{1}{N} \sum_{i=1}^N acc_i \quad (8)$$

where acc_i represents the accuracy of each category, and N represents the number of sample categories.

$$precision = \frac{TP}{TP + FP} \quad (9)$$

$$recall = \frac{TP}{TP + FN} \quad (10)$$

$$MicroF1 = 2 * \frac{precision * recall}{precision + recall} \quad (11)$$

where TP represents the sum of TP_i , FP represents the sum of FP_i , FN represents the number of samples with positive classes that were predicted as negative classes.

$$MacroF1 = \frac{1}{N} \sum_{i=1}^N F1_i \quad (12)$$

where $F1_i$ represents the F1 score of the i^{th} category.

Implementation details. Since there is no uniform way to divide the datasets, a fixed random seed and 10-fold cross-validation are adopted to ensure the reliability of the experimental results. First, ten percent of the training dataset in each fold is used as the validation set. The mean and variance of the experimental results are taken as the final value. MFCCs are used as the feature input. The prediction results of all segments in the same utterance are averaged as the final result. The Adam optimizer is adopted with a weight decay rate of 10^{-5} . The learning rate is initialized to 10^{-3} and exponentially decayed with multiplicative factor of 0.95 until the value reaches 10^{-6} . All models are trained for 50 epochs, and the batch size is set to 32. The Mixup method for data augmentation has a hyperparameter α , which controls the distribution of the randomly generated λ . α of the Mixup method is set to 0.5 same as GLAM [8].

4.2 Hyperparameter testing

Data augmentation methods for segmented speech require two other hyperparameters, i.e., speech segment length S and

Table 3 Descriptions of each dataset

Datasets	Language	Year	Categories	Size	Speakers
CaFE [30]	French	2018	Anger, disgust, fear, happiness, sadness, surprise, neutral	936	12
EMO-DB [31]	German	2005	Anger, boredom, disgust, fear, happiness, sadness, neutral	535	10
EMOVO [32]	Italian	2014	Disgust, fear, anger, joy, surprise, sadness	588	8
IEMOCAP [33]	English	2008	Happiness, anger, sadness, neutral, excitement	5531	10

Table 4 Performance of GLAM with different segments settings

Datasets	Segment	WA	UA	macro F1	micro F1
CaFE [30]	S2 O1	56.21±5.23	57.41±5.45	54.29±4.54	55.32±4.68
	S2 O1.6	66.02±5.06	66.90±4.41	64.19±5.87	65.21±5.50
	S1.8 O1.6	69.34±3.97	70.85±4.45	67.76±3.94	68.68±4.02
EMO-DB [31]	S2 O1	75.46±6.59	71.90±7.56	70.10±8.76	72.98±6.55
	S2 O1.6	80.20±6.49	79.07±7.19	76.90±8.37	79.62±7.00
	S1.8 O1.6	80.38±6.70	77.99±7.64	77.13±8.55	79.53±7.18
EMOVO [32]	S2 O1	72.63±5.53	72.13±4.23	70.79±4.27	72.44±3.39
	S2 O1.6	79.92±5.00	80.52±4.44	79.17±4.72	79.63±5.10
	S1.8 O1.6	80.29±3.91	80.69±3.61	79.49±3.74	80.47±3.84
IEMOCAP [33]	S2 O1	65.11±1.88	67.17±1.84	65.7±1.98	64.96±1.91
	S2 O1.6	67.13±0.92	68.02±1.11	67.54±1.41	67.05±1.03
	S1.8 O1.6	67.17±1.84	68.61±1.88	67.80±1.79	67.05±1.84

Table 5 Performance results of different schemes for the GLAM model on the IEMOCAP dataset

	Conv 3x3	Conv 5x3	Weighted Accuracy	Number of parameters
GLAM [8]			67.17±1.84	904,078
GLAM-cs (ours)	✓		69.23±1.98	859,110 (4.97%)
GLAM-cs (ours)		✓	67.28±1.49	773,990 (14.39%)
GLAM-cs (ours)	✓	✓	65.49±1.52	729,022 (19.86%)

overlap length O . To verify the effect of these hyperparameters on model performance, different parameter combinations are tested in GLAM. Each utterance is divided into numerous segments of S -seconds of length and O seconds of overlapping segments. Starting from $S = 2$ and $O = 1$, the values for S and O are increased or decreased to find the appropriate settings. The representative results are shown in Table 4. It is obvious that the performance of GLAM is significantly improved as the overlap length increases and the segment length decreases. The settings near the boundary are selected in the following experiments.

4.3 Ablation study

To validate the degree of improvement with the proposed scheme, the IRM is added to different convolutional layers to address the feature and computational redundancy of GLAM [8] on the IEMOCAP [33] dataset. The symbol ✓ in Table 5 represents replacing a convolutional layer with the IRM. It is obvious that by replacing the 3×3 convolutional layer with the IRM, the highest weighted accuracy can be obtained. Additionally, the number of model parameters is the lowest.

Furthermore, different kinds of stacked IRMs are used to replace the 3×3 convolutional layer in the multiscale convolution module. The number of IRMs ranges from 1 to 5. The accuracy of models with different numbers of stacked IRMs is shown in Table 6. The number of IRMs has little effect on the accuracy. For most datasets, one IRM is the best choice.

Table 6 Weighted accuracy of GLAM-cs when stacking different numbers of IRM

Dataset	IRM#1	IRM#2	IRM#3	IRM#4
CaFE [30]	70.82±3.76	70.51±5.16	70.08±4.63	70.71±4.56
EMO-DB [31]	81.31±4.93	79.08±5.68	82.08±7.63	81.13±5.20
EMOVO [32]	83.50±5.93	81.64±4.31	83.00±5.78	79.75±4.56
IEMOCAP [33]	69.23±1.98	68.38±2.14	68.25±1.40	67.71±1.51

4.4 Kernel model comparison

First, our proposed model is compared with several state-of-the-art models, including MHCNN [9], AACNN [7] and GLAM [8]. From Table 7, it can be seen that GLAM-cs outperforms the others in four evaluation indices. For IEMOCAP [33], the largest dataset, the WA of GLAM-cs is approximately 2% higher than that of GLAM [8]. For the other three datasets, which are smaller than IEMOCAP, the proposed method also achieves improvements on all metrics. In particular, compared with that of GLAM [8], the WA of GLAM-cs for CaFE [30] and EMO-DB [31] is improved by about 1%. Additionally, the WA of GLAM-cs in EMOVO [32] is significantly improved by 3% over GLAM [8]. These experimental results show that our proposed multiscale convolution block with channel shuffle can extract speech emotion features more efficiently.

To compare GLAM-es with other models in more detail, the weighted accuracy is visualized for each fold of IEMOCAP [33] using each model in Figure 4. GLAM-es has the best performance for most folds. However, the weighted accuracy of GLAM-es is lower in certain cases, e.g., fold 1 and fold 5. This is because the partitioning of the data may have caused certain deviations in the model fitting during 10-fold cross-validation. For example, when a certain amount of data is folded, the feature distributions learned by the model differ greatly. Therefore, the average calculation over the 10-fold cross-validation can show the real performance of the model.

The results of AACNN [7], MHCNN [9], GLAM [8], and GLAM-es on IEMOCAP [33] are shown in Figure 5. The diagonal elements reflect the accuracy of the model for each class. GLAM-es outperforms the other models by 3%–6% when predicting the Happy class. Moreover, the accuracy of GLAM-es for the Sad class is also greater. When combined with the MFCC feature visualization map, the color shades for Anger and Sad are relatively different, while that for Happy speech is in the middle. The prediction accuracy of the four models for the Happy class is approximately 60%, while GLAM-es achieves the highest accuracy, i.e., 66.8%.

4.5 Fusion module test

To verify the effectiveness of the fusion stage, different combinations of several models are tested by comparing the fuzzy integral with two traditional fusion strategies, i.e., Mean and Vote, as shown in Table 8. It is obvious that fuzzy integral fusion (FI) is superior to the other methods for any combination of models on each dataset. This is because the fuzzy measure not only reflects the importance of a single model but also represents the interaction between models. It always promotes the performance of the best model in the final decision. In addition, the combinations with GLAM-es using fuzzy integral fusion are

superior to those without GLAM-es for all datasets. This means that GLAM-es can actively contribute to improving the performance of the fusion framework. All results are displayed in Figure 6. For most datasets except for CaFE, the best combination includes GLAM-es and GLAM. Thus, the two models can promote each other due to being derived from the same base model.

4.6 Framework comparison

To prove the efficiency of the fusion framework based on GLAM-es (FF-GLAM-es), it is compared with several recently published methods on the difficult dataset IEMOCAP. Ref. [4] proposed the CNN-GRU-SeqCap model, which uses a CNN to extract features from the spectrogram. Two parallel modules, a gated recurrent unit (GRU) and sequential capsules (SeqCap), are used for feature fusion. Ref. [12] presented a DRN+transformer model that can find the potential dependencies between different positions of features by applying Transformer in a dilated residual network (DRN). Ref. [14] proposed the TFCNN+DenseCap+ELM framework, which consists of three parts, a time-frequency CNN (TFCNN) to learn the local representation of the spectrogram; DenseCap, which adopts a deeper capsule network to address the shortcomings of the simple capsule network; and an extreme learning machine (ELM) for classification. Most previous researchers did not employ a standard data processing method with different partitioning schemes for IEMOCAP [33]. Therefore, the respective configurations and n-fold cross-validation schemes are set as shown in Table 9. The kernel-based GLAM-es and FF-GLAM-es all have better performance than the other models. In particular, the framework combined with the fusion module performs best, which validates the effectiveness and rationality of the proposed framework.

Fig. 4 WA for different algorithms on the IEMOCAP dataset



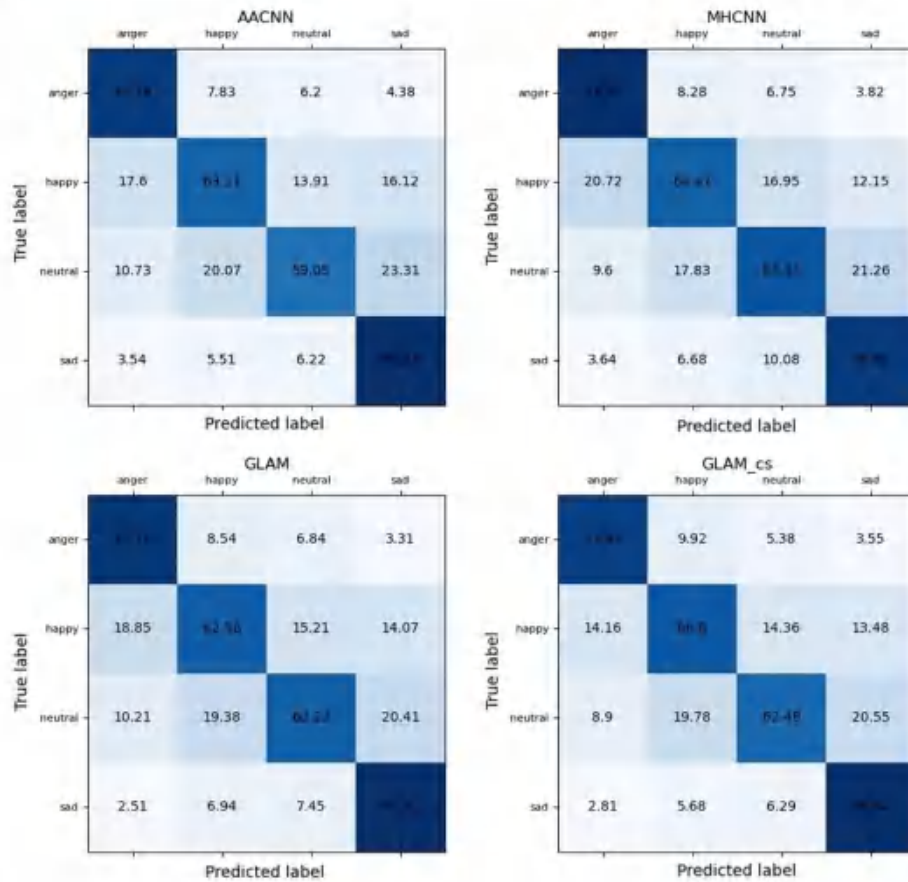


Fig. 5 Different model performance on the IEMOCAP dataset

5 Conclusions

In this paper, we proposed a fusion framework based on GLAM with channel shuffle (FF-GLAM-cs) for speech emotion recognition. GLAM-cs adopts a multiscale convolution module with channel shuffle to solve the feature redundancy and computational redundancy in the original model. GLAM-cs was validated to be superior to other models on

four datasets. Then, a fusion framework based on a fuzzy integral was constructed and verified. By fusing multiple models, the performance of the original framework can be effectively boosted. Experiments showed that fuzzy integral fusion is better than traditional fusion methods. The fuzzy measure was treated as a hyperparameter and optimized with the validation data, which adds some time to the computations. In future work, we will focus on how to determine the fuzzy measure faster.

Table 7 Comparison of evaluation metrics on different datasets with S1.8 O1.6 (S: segment length, O: overlap length)

Datasets	Model	WA	UA	macro F1	micro F1
CaFE [30]	MHCNN	66.14±3.83	66.99±3.93	64.18±4.87	65.66±4.40
	AACNN	67.74±5.51	68.61±6.75	65.99±5.47	67.33±5.61
	GLAM	69.34±3.97	70.85±4.45	67.76±3.94	68.68±4.02
EMO-DB [31]	GLAM-cs	70.82±3.76	71.84±5.14	69.75±4.79	69.99±4.48
	MHCNN	79.08±3.43	78.11±4.67	76.03±5.44	77.66±4.07
	AACNN	77.96±5.63	77.77±7.11	76.33±7.55	76.95±6.90
GLAM	80.38±6.70	77.99±7.64	77.13±8.55	79.53±7.18	
	GLAM-cs	81.31±4.93	78.89±7.24	77.84±7.93	80.55±5.14
	MHCNN	78.90±5.49	79.30±5.20	77.99±5.42	78.92±5.57
EMOVO [32]	AACNN	78.23±5.82	79.09±5.17	77.56±5.90	78.40±5.56
	GLAM	80.29±3.91	80.69±3.61	79.49±3.74	80.47±3.84
	GLAM-cs	83.50±5.93	83.46±4.39	82.27±5.25	83.32±5.85
IEMOCAP [33]	MHCNN	66.01±1.93	67.05±2.15	66.41±1.97	65.86±1.96
	AACNN	67.11±1.86	68.77±2.07	67.63±1.85	66.87±1.80
	GLAM	67.17±1.84	68.61±1.88	67.80±1.79	67.05±1.84
	GLAM-cs	69.23±1.98	70.53±2.14	69.80±1.99	69.11±2.00

Table 8 The unweighted accuracy (UA) of fusion framework with different combinations of models by different fusion tools. (1) AACNN, (2) MHCNN, (3) GLAM and (4) GLAM-cs

Dataset	Combination	Mean	Vote	F1
CaFE [30]	{(1), (2), (3)}	73.27±4.17	73.31±4.24	74.27±3.88
	{(1), (2), (4)}	74.43±4.72	72.24±3.76	75.6±4.09
	{(1), (3), (4)}	73.0±4.48	71.6±4.06	75.14±4.09
	{(2), (3), (4)}	73.92±4.51	73.82±4.07	74.66±3.81
EMO-DB [31]	{(1), (2), (3)}	79.51±7.63	78.45±7.91	81.7±7.71
	{(1), (2), (4)}	82.08±6.08	78.43±7.27	82.4±6.21
	{(1), (3), (4)}	78.76±6.43	77.58±7.76	83.03±6.09
	{(2), (3), (4)}	81.0±6.85	79.19±6.24	82.21±7.35
EMOVO [32]	{(1), (2), (3)}	81.23±5.19	80.78±4.42	82.83±3.76
	{(1), (2), (4)}	83.51±4.58	82.18±5.12	84.48±5.27
	{(1), (3), (4)}	83.17±5.19	82.18±4.28	84.03±3.68
	{(2), (3), (4)}	82.52±4.01	82.01±3.74	84.96±4.0
IEMOCAP [33]	{(1), (2), (3)}	71.32±2.37	70.37±2.07	70.93±2.22
	{(1), (2), (4)}	71.69±1.54	70.95±1.98	72.36±2.01
	{(1), (3), (4)}	71.63±1.39	70.77±1.74	72.71±1.35
	{(2), (3), (4)}	71.69±1.84	71.04±1.99	72.87±1.95

Acknowledgements The authors thank all editors and reviewers for their suggestions and comments.

References

- Schuller BW Speech emotion recognition: two decades in a nutshell, benchmarks, and ongoing trends 61(5):90–99. <https://doi.org/10.1145/3129340>
- Akçay MB, Oğuz K Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting

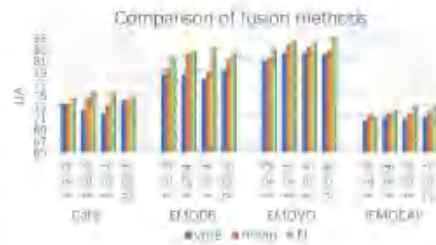


Fig. 6 Comparison of fusion methods: (1) AACNN, (2) MHCNN, (3) GLAM and (4) GLAM-cs

Table 9 Comparison of FF-GLAM-cs and other frameworks on the IEMOCAP dataset

Method	Unweighted Accuracy	K-fold
CNN GRU-SeqCap [4]	59.71%	5
DRN+ Transformer [12]	67.4%	10
TFCNN+DenseCap+ELM [14]	70.78%	–
GLAM [8]	68.61%	10
GLAM-cs(ours)	70.53%	10
FF-GLAM-cs(ours)	72.87%	10

- modalities, and classifiers 116:56–76. <https://doi.org/10.1016/j.specom.2019.12.001>
- Chen M, He X, Yang J, Zhang H 3-d convolutional recurrent neural networks with attention model for speech emotion recognition 25(10):1440–1444. <https://doi.org/10.1109/ISPL.2018.2860246>. Conference Name: IEEE Signal Processing Letters.

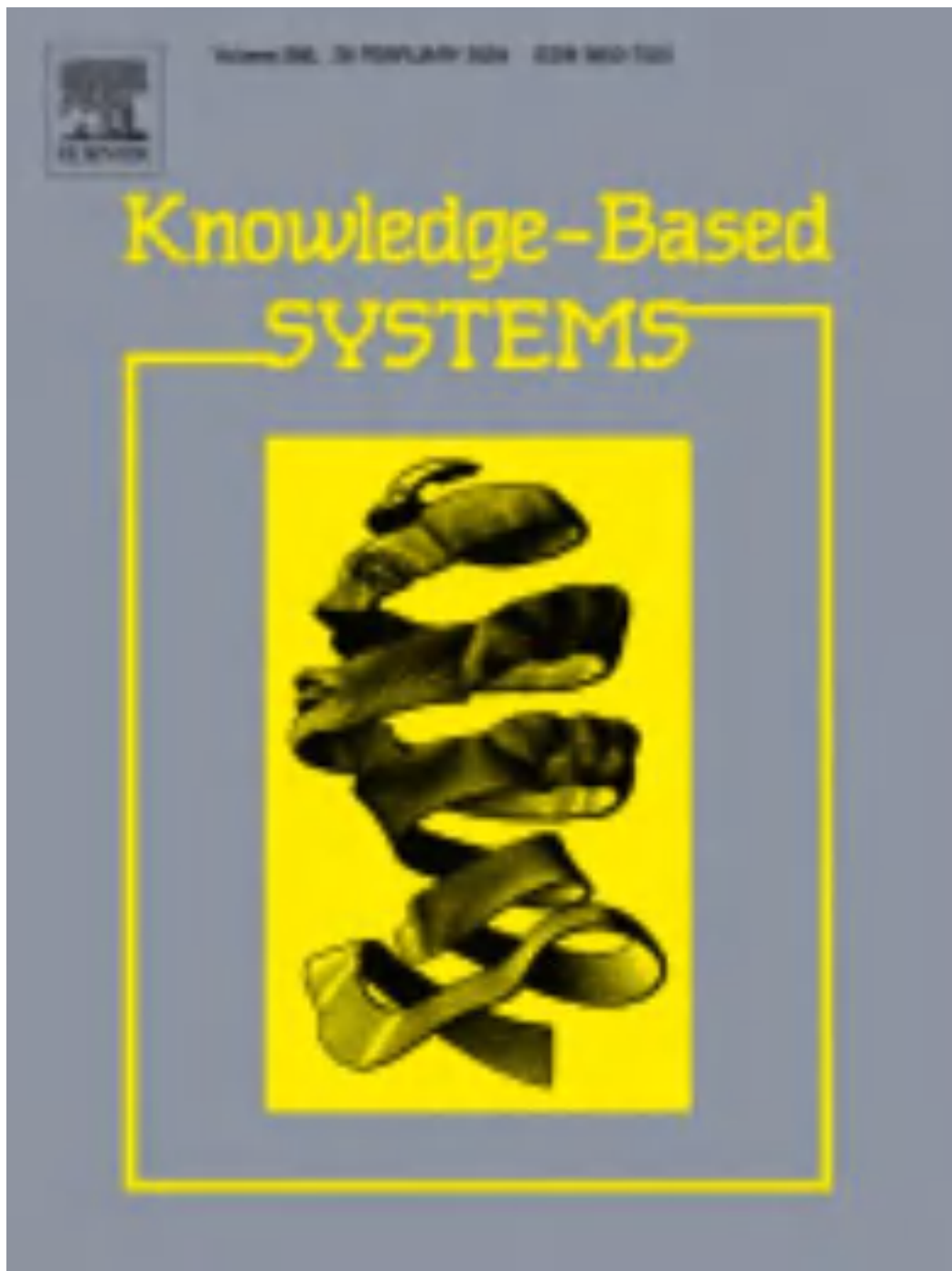
4. Wu X, Liu S, Cui Y, Li X, Yu J, Dai D, Ma X, Hu S, Wu Z, Liu X, Meng H Speech emotion recognition using capsule networks. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6695-6699. <https://doi.org/10.1109/ICASSP2019.8693163>. ISSN: 2379-190X
5. Meng H, Yan L, Yuan F, Wei H Speech emotion recognition from 30-sec mel spectrograms with deep learning network. 7-125808-125811. <https://doi.org/10.1109/ACCESS.2019.2938007>. Conference Name: IEEE Access
6. Issa D, Fathi Demter M, Yariqi A Speech emotion recognition with deep convolutional neural networks 58:101894. <https://doi.org/10.1016/j.speec.2020.101894>
7. Xu M, Zhang F, Cui X, Zhang W Speech emotion recognition with multiscale attention and data augmentation. In: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6319-6323. <https://doi.org/10.1109/ICASSP47728.2021.9414643>. ISSN: 2379-190X
8. Zhu W, Li X Speech emotion recognition with global-aware fusion on multi-scale feature representation. In: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6437-6441. <https://doi.org/10.1109/ICASSP4922.2022.9747517>. ISSN: 2379-190X
9. Xu M, Zhang F, Khan SU Improve accuracy of speech emotion recognition with attention head fusion. In: 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), pp. 1058-1064. <https://doi.org/10.1109/CCWC47524.2020.9031207>
10. Wang J, Xue M, Cullane R, Dao E, Ding J, Tanaka Y Speech emotion recognition with dual-sequence LSTM architecture. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6474-6478. <https://doi.org/10.1109/ICASSP40776.2020.9054629>. ISSN: 2379-190X
11. Poplan L, Riera P, Ferrer L Emotion Recognition from Speech Using Word2vec 2.0 Embeddings. arXiv. <https://doi.org/10.48550/arXiv.2104.05502>. arXiv:2104.05502
12. Lu Z, Cao L, Zhang Y, Chiu C-C, Fan J Speech sentiment analysis via pre-trained features from end-to-end ASR models. In: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7149-7153. <https://doi.org/10.1109/ICASSP40776.2021.9852957>. ISSN: 2379-190X
13. Cai X, Yuan J, Zheng R, Huang L, Church K (2021) Speech emotion recognition with multi-task learning. In: Interspeech, pp. 4508-4512. ISCA. <https://doi.org/10.21437/Interspeech.2021-1882>. https://www.isca-speech.org/archive/interspeech_2021/cd2/isca_interspeech.html
14. Liu J, Liu Z, Wang L, Guo L, Ding J Speech emotion recognition with local-global aware deep representation learning. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7174-7178. <https://doi.org/10.1109/ICASSP40776.2020.9054192>. ISSN: 2379-190X
15. Zehn W, Javed AR, Jaji Z, Khan HU, Gadikalu TR Cross corpus multi-lingual speech emotion recognition using ensemble learning 781:1805-1814. <https://doi.org/10.1016/j.speec.2021.101402>
16. Valles D, Marin R An audio processing approach using ensemble learning for speech-emotion recognition for children with ASD. In: 2021 IEEE World AI IoT Congress (AIoT), pp. 0055-0061. <https://doi.org/10.1109/AIoT52908.2021.9484174>
17. Wang X The Fuzzy Measure and Fuzzy Integral and Its Application in the Classification. Science Press
18. Grabusch M Fuzzy integral for classification and feature extraction
19. Banerjee A, Singh PK, Sarkar R Fuzzy integral-based CNN classifier fusion for 3d skeleton action recognition 3161-2206-2216. <https://doi.org/10.1109/TCSVT.2020.3019291>. Conference Name: IEEE Transactions on Circuits and Systems for Video Technology
20. Bhowal P, Sen S, Yoon JH, Geem ZW, Sarkar R Choquet integral and coalition game based ensemble of deep learning models for COVID-19 screening from chest x-ray images 2512:4326-4339. <https://doi.org/10.1009/OJHI.2021.3113415>. Conference Name: IEEE Journal of Biomedical and Health Informatics
21. Dey S, Bhattacharya B, Malakar S, Mirjalili S, Sarkar R Choquet fuzzy integral based classifier ensemble technique for COVID-19 detection 135-104385. <https://doi.org/10.1016/j.esump.2021.104385>
22. Pratik R, Desai M, Sen S, Saha Jimin L, Ad, Pava JP, Sarkar R A fuzzy distance-based ensemble of deep models for cervical cancer detection 219-106776. <https://doi.org/10.1016/j.amrb.2022.106776>
23. McFee B, Raffel C, Liang D, Ellis D, McCreary M, Buitrago B, Niño O Librosa: Audio and Music Signal Analysis in Python. <https://doi.org/10.25080/Majora7919e1ad300>. Pages: 24
24. El Agadi MMH, Kamel MS, Karay F Speech emotion recognition using gaussian mixture vector autoregressive models. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, vol. 4, pp. 957-960. <https://doi.org/10.1109/ICASSP2007.367230>. ISSN: 2379-190X
25. Dai W, Han D, Dai Y, Xu D Emotion recognition and affective computing on social social media 52(7):777-788. <https://doi.org/10.1016/j.mcm.2015.02.007>
26. Zhang H, Cisse M, Dauphin YN, Lopez-Paz D mixup: Beyond empirical risk minimization. <https://doi.org/10.26434/chemrxiv-1110-09682>
27. Zhang X, Zhou X, Lin M, Sun J ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. arXiv. <https://doi.org/10.48550/arXiv.1707.01883>. arXiv:1707.01883
28. Lee K-M, Lee Kwang H Identification of α -fuzzy measure by genetic algorithm 75(3):301-309. <https://doi.org/10.1016/j.am.2014.09.007>
29. Akiba T, Sano S, Yanase T, Ohno T, Koyama M Optuna: A Next-generation Hyperparameter Optimization Framework. arXiv. <https://doi.org/10.48550/arXiv.1907.10902>. arXiv:1907.10902
30. Goumay P, Lattae O, Lefebvre R A canadian french emotional speech dataset. In: Proceedings of the 9th ACM Multimedia Systems Conference, MMSys '18, pp. 399-402. Association for Computing Machinery. <https://doi.org/10.1145/3204949.3208121>
31. Burkhardt F, Pieschke A, Rietter M, Sendlinger WP, Weiss B (2005) A database of german emotional speech. In: Interspeech, pp. 1517-1520. ISCA. <https://doi.org/10.21437/Interspeech.2005-416>. https://www.isca-speech.org/archive/interspeech_2005/world-audio5k_interspeech.html
32. Contardi G, Iadrola L, Prodi A, Todisco M (2014) Emoto corpus: an italian emotional speech database. In: International Conference on Language Resources and Evaluation (LREC 2014), pp. 3501-3504. European Language Resources Association (ELRA)
33. Busso C, Bullot M, Lee C-C, Kazemzadeh A, Mower E, Kim S, Chang JN, Lee S, Narayanan S S IEMOCAP: interactive emotional dyadic motion capture database 4249-3355-3359. <https://doi.org/10.1007/s10879-008-9076-6>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g., a society or other person) holds exclusive rights to this article under a publishing agreement with its author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

3.以通讯作者发表本专业论文情况

3.1. EduLGCL: Local-global contrastive learning model for education recommendation



Knowledge-Based Systems

Supports open access

Submit your article

Menu

Volume 286

28 February 2024

Download full issue

< Previous vol/issue

Next vol/issue >

Receive an update when the latest issues in this journal are published

Sign in to set up alerts

Editorial board Full text access

Editorial Board

Article 111486

 View PDF



Research Article

Research article Full text access

Research on the attribute reduction method based on the best approximation set

Fachao Li, Kunyan Li, Chenxia Jin, Yuanjian Lin

Article 111362

 View PDF Article preview 

Research article Full text access

EduLCCL: Local-global contrastive learning model for education recommendation

Yijun Zhao, Fajian Jiang, Yin Pang, Yunxi Deng, ... Jinfeng Wang

Article 111357

 View PDF Article preview 

Research article Full text access

Knowledge-based Dual External Attention Network for peptide detectability prediction

Xiaocai Zhang, Hui Peng, Tao Tang, Yuansheng Liu, ... Jianjia Zhang

Article 111378

 View PDF Article preview 

Research article Full text access



EduLGCL: Local-global contrastive learning model for education recommendation

Yijun Zhao^a, Fajian Jiang^a, Yin Pang^b, Yunxi Deng^a, Youyou Han^b, Jinfeng Wang^{a,b,*}

^a College of Mathematics and Informatics, South China Agricultural University, Guangzhou, 510642, China

^b Key Laboratory of Smart Agriculture of Guangzhou, South China Agricultural University, Guangzhou, 510642, China

ARTICLE INFO

Keywords:
Self-supervised learning
Recommendation
Graph convolutional network
Supervised learning
Collaborative filtering

ABSTRACT

Contrastive learning has been used in recommendation to learn user and item representations from sparse and long-tail user-item interaction histories. Recommendations applied in education have several key challenges. Firstly, the frequency of student-course interactions is lower, resulting in limited data. Secondly, existing models rely on multimodal data for multi-objective recommendations and they limit the model's generalizability. Lastly, it is important to consider the impact of static features on different students. Aiming at addressing these challenges, a novel course recommendation system based on Local-Global Contrastive Learning called as EduLGCL is proposed, in which the mutual information between course features and student features is computed to select course features. After expanding the connections, the relationship between students and courses is described from different perspectives. Multiple graph convolutional operations are performed on the interaction history graph which can connect students and courses and represent student clusters. Through the feature extraction layer, the final student representation is obtained. Extensive experiments were conducted on three public datasets and a course dataset collected from a real-world environment to validate the performance of the EduLGCL. The results demonstrate that EduLGCL outperforms baseline methods, particularly in handling user cold-start problems. Our code is available at <https://github.com/mh17K/EduLGCL>.

1. Introduction

Recommendation systems proactively suggest items that users may like or find interesting, alleviating user concerns caused by information overload. In the context of the rapid development of the Internet, online learning has become an essential part of the educational journey for students. Online education helps students break through the limitations of time and space, bringing them a higher degree of learning freedom. The phenomenon of information overload due to the huge amount of educational resources on the Internet prevents students from quickly finding courses and learning directions that interest in [1,2]. Learning in the wrong direction can easily lead to student dropout and a decline in learning interest. How to reduce students' selection costs, meet their personalized learning needs, and increase their interest in learning is the goal of recommendation systems in education [3].

Existing recommendation systems which are widely used in e-commerce scenarios [4,5], music scenarios [6], and video scenarios [7] mostly learn users' personalized preferences from their interaction history with items. However, in the educational scenario, student-course

interactions are cold behaviours, i.e. the act of choosing between students and courses exists only for a short period time in each semester. This means that it is difficult for recommender systems to accurately represent a student's learning interests from short interaction histories, and thus recommend courses to students that may be of interest to them. Good examples in other domains cannot be directly applied in the education scenario.

At the same time, there exists user cold-start problems in the education scenario, such as a freshman entering school without interaction with any courses, making it difficult to capture students' interests and preferences. No interaction or interactions less than 5 times are often removed in other recommender systems, which is not applicable to education. Therefore, learning a more accurate representation of students and courses based on the sparse history of student-course interactions becomes the key problem.

There is also a clear long-tail problem in the education scenario, where most students tend to opt for popular courses and there are some cold and excellent courses that tend to be overlooked. Because it is difficult to accurately express students' interests, it is difficult for

* Corresponding author at: College of Mathematics and Informatics, South China Agricultural University, Guangzhou, 510642, China.
E-mail address: wangjinfeng@scau.edu.cn (J. Wang).

<https://doi.org/10.1016/j.kbsys.2023.111357>

Received 12 October 2023; Received in revised form 11 December 2023; Accepted 28 December 2023

Available online 30 December 2023

0950-7051/© 2023 Elsevier B.V. All rights reserved.

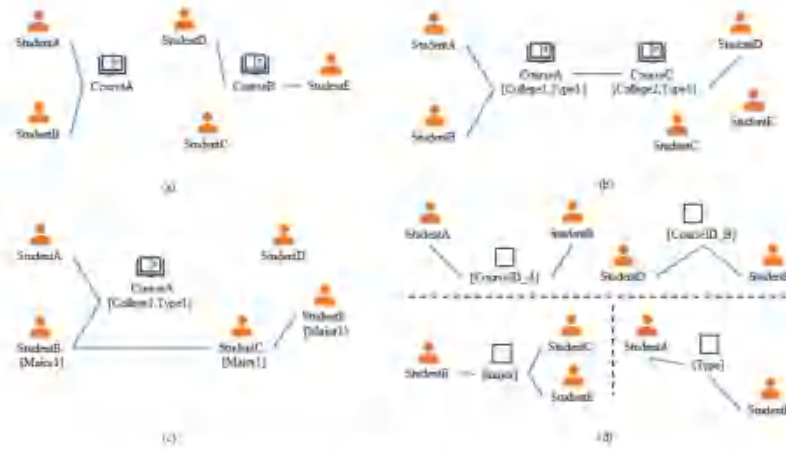


Fig. 1. Enhance connection to existing student-course interaction history using basic features of students and courses.

traditional recommender systems to help students explore these cold courses.

Recent researchers have also provided several solutions for the recommender systems applied in educational scenarios including leveraging social networks [6], constructing knowledge graphs with external knowledge sources [9], and learning student preference representations from multimodal data [10]. These methods have been adjusted to different recommendation objectives and have achieved promising results. However, this external knowledge is often not easily obtained and applied in educational scenarios. Therefore, the use of static profiles of students and courses can be considered to complement the interaction history of students and courses to learn more comprehensively the representation of students' interests. At the same time, there are student community relationships in the education scenario like social networks, as there are similarities in interests among students in the same majors. These community relationships are extracted from the static profile of students and do not need to be additionally collected like social network data. The use of static profiles of students and courses to supplement the interaction history is illustrated in Fig. 1.

Fig. 1(a) demonstrates a traditional recommendation approach that utilizes only the interaction history between students and courses for recommendations, limiting the representation of courses to mere ID numbers. Given the same courses taken by students A and B, and students D and E, interest similarity arises. Nonetheless, the system remains unaware of student C's interests. By the static attributes of courses, denoted as Type, it is possible to find the similarity between courses A and B. Consequently, this reveals that students D and E might have similar preferences and interests, as illustrated in Fig. 1(b). Utilizing the static profiles of students in the educational scenario, such as Major, to construct a student relationship network can reveal the connections between students B, C, and E. These three students might have similar preferences and interests, as illustrated in Fig. 1(c). Not only can this approach help identify the potential interests of student C, but it also reveals the interest relationships between students A, B, D, and E. Finally, to reduce the over-smoothing issue in the graph convolutional computation [11], the expanded interaction history is divided into multiple connected subgraphs based on different descriptive features. These subgraphs represent the connection relationships between students from different perspectives. As illustrated in

Fig. 1(d), a connection relationship among students is constructed from the perspectives of major, type, and ID.

Another key problem is how to choose the appropriate static features to better represent the student's preferences. Manual selection methods would require extensive experimentation. There are methods to measure feature importance based on average impurity [12] and gradient boosting decision tree selection based on gradient contribution [13,14]. These methods require more computer memory resources, and the effect of feature selection depends on the degree of dispersion of the data set. In this paper, we propose calculating the mutual information value between student ID features and course features as the basis for feature selection. By choosing the multiple features that have the most significant impact on student interests, we generate connected subgraphs from different perspectives.

To address the above problem of applying recommender systems in the education scenario, this paper proposes to complement the history of student-course interactions using static profiles of courses to describe student interest representations from different perspectives. At the same time, using static profiles of students to extract their community relationships and explore the learning interests of potential students. Constructing cross-graph contrastive learning to integrate information and mitigate long-tail problems in scenarios.

In this article, a recommendation model based on graph contrastive learning with mutual information is proposed to provide with students some courses in which may be interested. The main contributions of this work are listed as follows.

- A recommendation model using a local-global learning model with mutual information filtering features called EduLGCLE is designed to solve the data sparsity and user cold start in the task of course recommendation in educational scenarios.
- More accessible static characteristics of students and courses are proposed to extract group relationships among students and connection relationships under different perspectives and expand the connections between students and courses without the other data sources, which can improve the generalizability of the model.
- Mutual information is used to filter course statics and select appropriate feature combinations for extending existing student-course interaction history. This process generates a global relational view for contrastive learning and provides explanations for the recommended results.

- Extensive experiments on three real-world datasets demonstrate the improvement of the proposed model in recommendation accuracy. For application in real-world educational scenarios, two datasets of student control groups are collected for experiments. This further validates the feasibility and effectiveness of the proposed model in education.

The rest of the paper is organized as follows. Section 2 provides an overview of recent technologies. Section 3 presents the general form of the course recommendation problem in an educational scenario and the implementation details of EduLGCN. Section 4 reports the findings of experiments conducted to validate the enhancement of the model. Section 5 concludes the paper.

2. Related work

2.1. Content-aware recommendation

Employing embedding vectors to represent users and items, this study optimizes the learning of user-item interaction data. Similar to MF [35], it maps user and item IDs to low-dimensional embedding vectors and calculates their inner products to predict interactions. However, this method solely learns from user-item interaction data, while contextual features (e.g., location, time) also impact user interaction behavior. Recently, deep learning methods have emerged to leverage contextual feature interactions for improved recommendation performance [36–38]. For instance, NFM [36] proposes click operations and pooling on vectors, while AFM [37] introduces an attention mechanism to highlight the importance of feature interactions. GCM [38] encodes the context information through the encoder and adds it to the graph convolutional filter for learning. Despite their effectiveness, most of these works focus only on second-order feature interactions. The proposed EduLGCN extends the original graph-structured interaction data with static features of students and courses. By reducing the number of nodes at a fixed distance, it brings students closer together while better capturing higher-order collaborative filtering signals.

2.2. Graph convolutional network for recommendation

In early recommendation systems based on matrix factorization methods [35], sparse matrices were used to record the interaction relationships between users and items. The graph convolutional neural network (GCN) treats the interaction matrix as a bipartite graph that records the connectivity between user nodes and item nodes. By performing multiple graph convolution operations on this interaction graph and propagating knowledge between neighboring nodes, higher-order relationships between users and items can be captured. It can also handle data with asymmetric length of user interactions. NGCF [19] extends the traditional collaborative filtering approach by applying traditional graph convolutional networks to recommendation systems by capturing higher-order neighbor relationships in the interaction graph. LightGCN [20] considers the specificity of the recommendation problem to remove the unnecessary nonlinear activation and feature transformation matrices in traditional graph convolutional networks, which simplifies the computational process and uses the weighted sum to simulate the residual connections. A random-walk method was used in some models, such as PinSage [21] and GraphSage [22], to generate sub-graphs that sample and aggregate graph structure for efficient information transfer in large-scale graphs. However, this method is similar to node dropout and may not consider some cold nodes when walking. DiffNet [23] uses a differential attention mechanism to distinguish the degree of contribution of edges between nodes to improve recommendation performance. These approaches share the commonality of using graph structures to capture relationships between nodes, thereby improving the performance of recommendation algorithms.

2.3. Contrastive learning for recommendation

In recommender systems, low-frequency data will lead in sparse supervised signals that the model can learn. By introducing contrastive learning, contrasting views with different neighbor structures are generated based on the main view by learning the representation of the same node on different views. Nodes in the main view can extract useful information from the contrastive view, which can effectively mitigate the data-sparsity problem in the scenario [24]. The contrastive learning methods currently applied in recommender systems can be divided into two types based on generating contrastive views: methods with structural dropout and methods with local-global relationships. Contrastive learning based on structural dropouts, such as SGL [25], and SimGCL [26], can generate contrastive views in the main view or introduce random noise in the embedding space. This approach tends to introduce more natural noise and further exacerbates the sparsity of connections to cold nodes which are already sparse. Contrastive learning generating local-global relationships, such as NGL [27], HCCN [28], and LightGCL [20], may use external knowledge sources like clustering relationships or knowledge graphs to generate global representations of nodes as contrastive views. It can also perform matrix factorization on the main view to generate contrastive views including global information. While this approach effectively addresses the cold-start problem, its effectiveness relies on the quality of external knowledge. It may lose model generalization. The proposed EduLGCN constructs a global relational view using easily accessible static features of students and courses to describe students' preferred interests from different perspectives.

2.4. Application of recommendation system in education

In recent years, researchers have applied recommendation technologies in education scenarios to improve students' motivation and engagement in learning [3]. These methods often describe students' learning preferences from different perspectives based on knowledge graphs [20] constructed from external knowledge sources and multi-modal data such as video and audio [30]. However, these external knowledge sources typically require specific scenarios for information collection, which may result in poor generalization in different educational contexts. At the same time, these models make recommendations for objectives such as learning motivation in offline classes, and lack a quick filtering method for online classes. The proposed EduLGCN uses easily accessible static profile information to complement the connections of cold users by using potential community relationships between groups of students to complement. Freshman have not yet selected any university courses, which may explore in the wrong learning direction early leading to an increase in absence rates and a decrease in motivation [31]. EduLGCN can use collective recommended preferences to help students explore new learning directions more accurately.

3. Methodology

3.1. The definition of course recommendation problem

The recommendation system [32] determines the user's preference for an item by analyzing the user's interaction history. This level of preference can be reflected by an item's rating made by the user, such as a course's rating. Using this paradigm, the historical interaction data between students and courses can be represented in the form of a rating matrix. However, in actual recommendation scenarios, students are only explicitly interested in a small portion of the courses by rating and reviewing and implicitly interested in a portion of the courses by browsing or clicking, but there is still a large portion of the courses without interaction with students, which is a long-tail phenomenon. The rating matrix is usually sparse and asymmetric. Using a bipartite graph to represent the interaction data can save storage space.

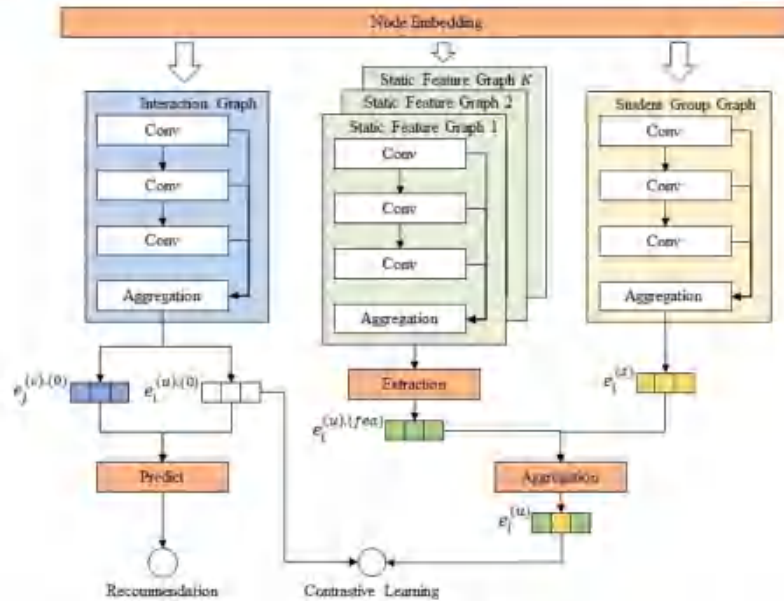


Fig. 2. The overall structure of EduAGCL.

The graph-based course recommendation problem can be defined in the following form. Let U and C denote the sets of N students and M courses, where $|U| = N$ and $|C| = M$. The interaction history of a student and a course can be represented as a student-course interaction graph G_{ij} , which is defined as $\{(u, v_{ui}, i) | u \in U, i \in C\}$, with edge $v_{ui} > 0$ indicating an observed interaction between student u and course i . The interaction behavior includes both explicit and implicit interactions. Conversely, $v_{ui} = 0$ indicates no observed interaction between student u and course i . The set of course features is denoted as D using K static features of the course to complement the course connection, where $|D| = k$. The course graph G_c records the relationships between course and their feature information, defined as $\{(i, link_{i,f}, f) | i \in C, f \in D\}$. When $link_{i,f} = 1$, it indicates that course i can be described by static feature f . $link_{i,f} = 0$ means no relationship between them. However, using static features of different aspects to describe a course at the same time may lead to feature blurring [33]. The set of course features is denoted as $\{D_1, \dots, D_T \in D\}$ by classifying the static features into T groups with different indirect nodes according to the node types. The number of features contained in each indirect node is denoted as $\{K_1, \dots, K_T \in K\}$. Different feature-course sub-graphs can be obtained for each indirect node, which describes the connections between students and courses from different perspectives. These sub-graphs can be represented as $\{G_{c_1}, \dots, G_{c_T} \in G_c\}$.

Additionally, within the scenarios, there exist implicit connections among students beyond social relationships, such as shared interests between two students studying in the same major. The student graph, denoted as $G_s = \{(u, link_{uv}, friend) | u \in U\}$, captures the connectivity among students. Here, the edge $link_{uv} = 1$ signifies an observed connection between student u and their friend, while the edge $link_{uv} = 0$ indicates an unobserved connection between student u and their friend.

The inputs and outputs of the course recommendation model can be expressed as follows:

Input: a student-course bipartite interaction graph G_{ij} , a course graph G_c containing static features of the course, and a student group graph G_s .

Outputs: The edge weight v_{ui} between student u and course i will be predicted and sorted. The first several courses ranked in descending order by course prediction scores are used as a recommendation list.

3.2. The overall structure of EduAGCL

The overall structure of the EduAGCL proposed in this paper is shown in Fig. 2. The embedding layer provides trainable representations of student embeddings and embeddings of individual course features. All views share the same set of student embedding representations. The graph convolution process can be divided into three channels, including a main view channel with interaction history graphs, a static feature channel with all student-course connection graphs, and a student group channel containing student community information graphs. Three graph convolution operations are performed in each channel for propagating and aggregating low-order and high-order neighbour information for each node. In the main view channel, the learned embedding representations of student and course are fed into the prediction layer. An inner product operation is performed in the prediction layer to generate a similarity score indicating the student's preference for the course. The student embeddings learned in the static feature channel go through a fully connected feature extraction layer to produce the final embedding representation of student with global information according to each student's interests. A contrastive learning task with the global representation of the student is performed. The embedding representation of the student in the main view channel is considered as a local representation to extract useful information from the global representation of the student.

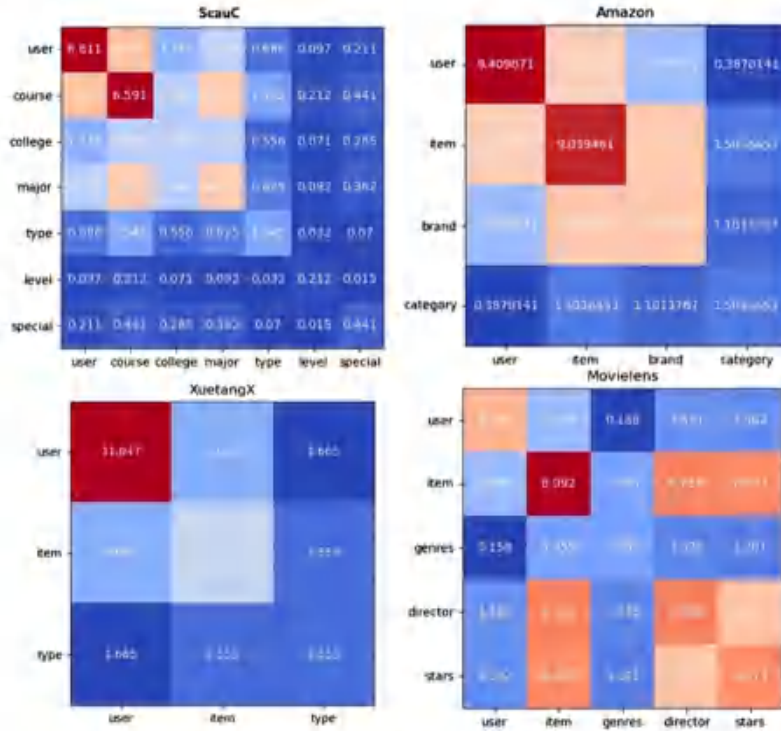


Fig. 3. Heat map of mutual information values for each dataset.

3.2.1. Embedding layer

Using a unique identifier maps the student, course, and feature labels into a vector, an embedding finding table is initialized as the corresponding embedding representations $e_i^{(s)}$, $e_j^{(c)}$, and $e_l^{(f)}$, for the student node u_i , the course ID node c_j and course static feature node $l_l^{(f)}$. All student nodes and course nodes can be represented as $E^{(s)} \in \mathbb{R}^{N \times d}$ and $E^{(c)} \in \mathbb{R}^{M \times d}$, where N and M is the number of students and courses; d denotes the dimension of the embedded representation.

3.2.2. Local relation learning

Neighbour information propagation is performed three times on the student-course interaction graph G_0 which represents the interaction history of the student and the course. The first-order neighbour and high-order neighbour information of the nodes are learned. Local representations of student nodes and course nodes are obtained. Referring to the LightGCN and MFGCL [34], EduLGCN is designed using a graph convolutional encoder with the feature transform matrix, non-linear activation method, and initial layer embedding removed. At layer $(l + 1)^{th}$, the information aggregation process can be represented in the following form.

$$e_{i,j+1}^{(s)} = \sum_{c \in N_i^{(s)}} \frac{1}{\sqrt{|\sum_j R_{ij}^{(s)}|} \sqrt{|\sum_j R_{ij}^{(c)}|}} e_j^{(c)} \tag{1}$$

$$e_{j,l+1}^{(c)} = \sum_{u \in N_j^{(c)}} \frac{1}{\sqrt{|\sum_u R_{ju}^{(s)}|} \sqrt{|\sum_u R_{ju}^{(c)}|}} e_u^{(s)} \tag{2}$$

where $e_{i,j+1}^{(s)}$ and $e_{j,l+1}^{(c)}$ denote the results of student node u_i and course node c_j after aggregating and propagating the neighbour information in G_0 in the $(l + 1)^{th}$ layer; $\sum_j R_{ij}^{(s)}$ and $\sum_j R_{ij}^{(c)}$ denote the degree of student node u_i and course node c_j in G_0 ; R is defined as a degree matrix; $N_i^{(s)}$ and $N_j^{(c)}$ denote the aggregation of neighbour nodes of student node u_i and course node c_j in G_0 . The final representation of student node u_i and course node c_j in G_0 is obtained by aggregating the node representations of all layers in G_0 . The calculations can be expressed as the following form.

$$e_i^{(s)} = a_1 e_{i,1}^{(s)} + a_2 e_{i,2}^{(s)} + a_3 e_{i,3}^{(s)} \tag{3}$$

$$e_j^{(c)} = a_1 e_{j,1}^{(c)} + a_2 e_{j,2}^{(c)} + a_3 e_{j,3}^{(c)} \tag{4}$$

To simplify the calculation, the weights in EduLGCN are set as $a_1 = a_2 = a_3 = \frac{1}{3}$. It is also possible to use the attention mechanism to generate the weights [35,36].

The final representations of student node u_i and course node c_j are subjected to an inner product to generate a similarity score for predicting the weights y_{ij} of the edges between student node u_i and course node c_j in G_0 . After sorting the values y_{ij} in similarity, the top 10 and top 20 courses are selected and output as a recommendation list. The calculations are shown below.

$$y_{ij} = e_i^{(s)} \cdot e_j^{(c)} \tag{5}$$

3.2.3. Global relation learning

In order to reduce feature blurring and over-smoothing in graph convolution computation, the student-course connectivity graph is divided into multiple student-course sub-graphs in different perspectives, denoted as $\{G_{c_1}, \dots, G_{c_T}\} \in \mathcal{G}_T$ in which T denotes the total number of selected course feature nodes and G_{c_i} is denoted as G_{ij} . The number of feature nodes in each graph can be noted as $\{K_0, \dots, K_T\} \in \mathcal{K}$, where \mathcal{K} represents the set of static course characteristics, such as course, college, and major. By calculating the mutual information value between course characteristics and student identification features, T distinct features are selected. The heat maps of mutual information values for each dataset are illustrated in Fig. 5. It was found that selecting the feature group with the largest mutual information value and the smallest worked best. Because this kind of combination is least affected by the feature blurring problem.

A three-layer graph encoder is applied to each student feature graph to learn and aggregate higher-order neighbor relationships for each node. Referring to SocialGN [33], an information extraction mechanism is designed. It utilizes a fully connected layer to aggregate the final representations of student nodes and feature nodes. The aggregation is performed on T student-course graphs. This process generates a personal global representation of student nodes. The calculations can be expressed as the following form.

$$z_c^{global} = w_{c_{max}} \left\{ \sigma \left(w_{c_1}^{global} z_c^{local} \right) \right\} \left\{ \sigma \left(w_{c_2}^{local} z_c^{local} \right) \right\} - \left\{ \sigma \left(w_{c_{T+1}} z_c^{local} \right) \right\} \quad (6)$$

$$z_c^{local} = \frac{f^{(c_1)}(u_i)}{\sum_{j=1}^T f^{(c_j)}(u_i)} \quad (7)$$

where σ denotes the nonlinear activation function \tanh ; $w_{c_1}, w_{c_2}, \dots, w_{c_{T+1}} \in \mathbb{R}^{d \times d}$ is the trainable matrix used for adaptive extraction of personal preferences; $w_{c_{max}} \in \mathbb{R}^{(d \times T) \times d}$ is the trainable matrix used for dimension reduction of the concatenated node representation. z_c^{local} denotes the final representation of student node u_i in the student-feature G_{c_i} , and particular z_c^{local} is $z_c^{local} = z_c^{local} \cdot z_c^{local}$ denotes the final representation of student node u_i . The normalization operation prevents z_c^{local} from being too large.

Similarly, the connections between groups of students can effectively broaden their interests, especially the interactions among freshmen. The model performed three rounds of graph convolution on the student group graph to uncover both direct and second-order connections among students. After that, the model combined this information with the ultimate global representation of student nodes, creating node representation for comparative learning. The specific calculation process can be presented in the following.

$$z_i^g = \text{aggregation} \left(z_i^{local} \otimes z_c^{local} \right) \quad (8)$$

where z_i^{local} represents the representation of student node u_i obtained from the student graph G_i , encompassing the student node u_i as well as information from the higher-order neighbors, which is computed as in Eq. (7). To streamline computations, the aggregation here is performed by taking the average.

3.2.4. Local-global contrastive learning

In the joint learning approach constructed in this paper, the main view performs the recommendation task, using BPR Loss [17] as the loss function. The contrast view performs the contrast learning task using InfoNCE loss [28] shown below.

$$L_{InfoNCE} = - \sum_{(i,j) \in N_p \cup N_n} \ln \sigma \left(\gamma \left(z_i^g - z_j^g \right) \right) \quad (9)$$

During training in a mini-batch, the unobserved student-course connections are treated as negative samples, while the observed connections are treated as positive samples. N_p denotes the set of all courses with

Table 1
The comparison of time complexity.

	LightGCN	LGCAF	SGL-ED	EduLGCN
Adjacency matrix	$2 A $	$4 A + 2K A $	$4 A + 4 A $	$4 A $
Graph traversal	$O(2 A)$	$O(2 + 2K) \cdot A $	$O(2 + 4) \cdot A $	$O(2) \cdot A $
Prediction	$O(2 A)$	$O(2 A)$	$O(2 A)$	$O(2 A)$
Constant	-	-	$O(BN)$	$O(BN)$

which student u_i interacts in the current mini-batch; σ denotes the similarity scores of negative sample pairs (i, j) ; σ denotes the sigmoid activation function.

By employing contrastive learning loss, the distances between the local and global representations of the same node in the embedding space are reduced, while the distances between different nodes are amplified. This method allows the extraction of useful information from global signal representations to enhance the local signal representations, thereby mitigating the issue of data sparsity. The contrastive learning loss function is shown below.

$$L_{CL} = - \sum_{i \in \mathcal{N}} \log \frac{\exp(\langle z_i, z_i^{local} \rangle / \tau)}{\sum_{j \in \mathcal{N}} \exp(\langle z_i, z_j \rangle / \tau)} \quad (10)$$

where z_i and z_i^{local} denote the global and local representations of the same node i respectively; z_j denotes a negative sample node, referring to other nodes of the same type within the current mini-batch, excluding the node itself; τ represents the temperature parameter of the softmax function, which controls the degree of the contribution of the negative sample pairs during the training process.

The joint learning method for the recommendation task and the contrastive learning task is shown below.

$$L = \lambda_1 L_{BPR} + \lambda_2 L_{CL} + \lambda_3 \|\Theta\|_2^2 \quad (11)$$

where λ_1 controls the strength of contrastive learning; λ_2 controls the strength of L2 regularization; Θ is the set of model parameters, including the initialized embedding of all nodes and the matrix of trainable parameters for the fully connected layer.

3.3. Time complexity analysis

In this subsection, the theoretical time complexity of LightGCN [20], LGCAF [22], SGL-ED [25], and EduLGCN will be analyzed in a mini-batch shown in Table 1. Let $|A|$ denotes the number of edges observed in an interaction graph; K denotes the number of selected features; r denotes the probability of edge discard; L denotes the number of layers of convolution; d denotes the dimension of the embedding space; B denotes the mini-batch size; and N denotes the number of nodes in the current mini-batch.

- Both EduLGCN and LGCAF utilize item description information to expand the interaction graph, and the regularized adjacency matrix constructed has $2|A| + 2K|A|$ non-zero elements. Due to the feature ambiguity problem, not all description information can bring positive improvements. LGCAF attempts various information combinations through manual methods, while EduLGCN uses mutual information for feature selection. For SGL-ED, because it uses edge dropout to construct two contrastive learning channel views, each contrast channel's adjacency matrix has $2r|A|$ non-zero elements. LightGCN does not use graph augmentation operations, so the regularized adjacency matrix has only $2|A|$ non-zero elements.
- In the graph traversal, since SGL-ED is a three-channel graph comparison learning framework, the time consumption is almost three times higher compared to the single-channel LightGCN.

LGCACF and EduLGCCL, on the other hand, experience increased time consumption because they use decomposition of multiple sub-graphs.

- The prediction layers of the above methods are trained in each mini-batch using BPR Loss, resulting in similar time consumption for this part.
- Only SGL-ED and EduLGCCL utilized contrastive learning frameworks, which results in time consumption due to the acquisition of positive and negative samples. Treating the different representations of nodes themselves as positive samples requires $O(MN)$ time, while considering other nodes within the same mini-batch as negative samples takes $O(M^2N)$ time, where M is the ratio of negative samples in positive samples.

4. Experiments and analysis

To verify the effectiveness of the proposed EduLGCCL, we perform massive experiments to explore the following research questions. EduLGCCL performs on different datasets compared with Baseline. EduLGCCL performs with different numbers of features. EduLGCCL performs for user cold-start problems on different datasets. Different parameter settings affect the performance of EduLGCCL.

4.1. Experimental setting

4.1.1. Datasets

The experiments are conducted on three real-world datasets: MovieLens [40], Amazon [41] and XuetangX [42]. EduLGCCL is proposed for educational applications, so the real data is collected from different grades of one university. Offline questionnaires are adopted to collect information from students and their intention to interact with the courses. The used public datasets are sourced from various application scenarios, exhibiting variations in both dataset size and feature dimensions. Divide the data into training set, validation set, and testing set in the ratio of 7:1:2. The detailed information of each dataset is shown in Table 2.

- **MovieLens** [40]: It is a film dataset widely used to evaluate the performance of recommender systems. Film genre, film director, and film rating were chosen as the static profiles of the movies. After removing invalid films with an interaction count of less than 5, the user's rating behavior is considered as implicit interaction data.
- **Amazon** [41]: A large and widely used e-commerce dataset, of which Amazon Electronics was selected to evaluate the performance of EduLGCCL. All interactions between users and products are treated as implicit data, except for products with less than 5 interactions and users with less than 10 interactions.
- **Taobao** [43]: A dataset from the e-commerce domain, collected from Taobao.com, was utilized in this study. The interaction data encompassed "purchase" and "add to cart" behaviors. Users with less than 10 interactions were excluded.
- **XuetangX** [42]: Data was collected in 2018 at XuetangX, which was one of China's largest MOOC platforms, including 1,302 courses, 82,535 users, and 458,454 user-course interactions. All interactions were recorded with the same weight as invisible interests.
- **ScauC** [Ours]: This dataset was collected from our university including 2002 core courses taken as the course set. Students' interests in the courses were gathered through off-line questionnaires. The students were divided into two groups: the freshman group consisting of 445 students and the senior group consisting of 350 students. The active interest preferences and existing course enrollment information were adopted from both groups of students. After filtering out courses without interaction at all, all expressed interests between students and courses were considered as explicit interests and recorded with equal weight.

Table 2
The details of each dataset.

Dataset	Users	Items	Rating	Density	Features	Primary count
MovieLens [40]	610	9606	10000	0.0171	genre director actor	21 2942 1034
Amazon [41]	1220	14004	24060	0.002	brand category	2771 15
XuetangX [42]	82535	102	40840	0.0002	type	19
Taobao [43]	89,008	89,406	91833	0.00008	category	167
ScauC [Ours]	897	1614	7277	0.0029	course type level quest	21 31 3 4 3

4.1.2. Baselines

The following models are used as baselines for comparison, including Matrix Factorization (MF) [16], LightGCN [20], LGCACF [26], SGL-ED [23], BUB [44], SimGCL [36], and MFGCL [34].

- **MF** [16]: Matrix Factorization is a classic and widely used technique in recommendation systems to decompose a user-item interaction matrix into latent features. It enables the identification of hidden patterns and preferences, allowing personalized recommendations based on these latent factors.
- **LightGCN** [20]: The application of graph convolutional networks was proposed in recommendation models by extending traditional collaborative filtering methods and addressing the inability to effectively capture latent high-order relationships. Unnecessary nonlinear activations and feature transformation matrices have been removed for the recommendation task.
- **LGCACF** [26]: Using LightGCN as the foundational encoder, distinct interaction graphs are constructed for various static item features and user identifier features to represent user embeddings from different perspectives. Global user representations are generated through averaging, allowing exploitation of item connections from different viewpoints.
- **SGL-ED** [23]: This is a joint recommendation model applying graph contrastive learning. To address the long-tail and sparse supervision issues in recommendation scenario, contrastive learning was introduced by SGL. Different contrastive learning views were generated using edge dropout, establishing a three-channel contrastive learning mode comprising a main view and contrastive learning views. A unified learning mode was constructed to simultaneously optimize the recommendation loss and contrastive loss. ED was the best-performing variant to represent the method using edge dropout in the SGL model.
- **BUB** [44]: The recommendation model, using only first-order collaborative filtering signals, adopts a two-tower architecture for the target network and online network. Self-supervised learning is applied to positive sample pairs to extract augmented views from neighborhood information, mitigating the issue of data sparsity. However, negative sample pairs are not considered in the optimization learning process.
- **SimGCL** [36]: Unlike SGL-ED, this model proposes the addition of uniform noise in the embedding space to generate different representations of the same node, enabling contrastive learning without changing the original graph structure. As this method introduces perturbations in the embedding space, it can be considered a form of local-global contrastive learning model.
- **MFGCL** [34]: A Top-K recommendation model based on course-awareness and contrastive learning is proposed. LightGCN serves as the base encoder, dividing the binary user-feature interaction

graphs into multiple parts based on node types. Distinct representations are obtained from different perspectives on these interaction graphs, which are then aggregated to form the final representation. A three-channel contrastive learning structure is constructed, wherein two contrastive learning channels generate disturbed representations by adding random noise to the embedding space.

4.1.3. Hyper-parameter setting

For the embedding layer setup, the Xavier initializer is used to create a 512-dimensional initialized user and message embedding. All models are optimized using Adam optimizer to search for appropriate learning rates in $\{0.01, 0.005, 0.001\}$. The coefficient λ_2 of the L2 regularization is fixed to 10^{-4} . The batch size is set as 2048. The appropriate temperature coefficients in the contrastive learning models, SGL-ED and EduLGC, are determined in $\{0.2, 0.175, 0.15, 0.125, 0.1\}$, and the contrastive learning strengths λ_1 are searched for the appropriate scale in $\{0.2, 0.1, 0.05, 0.01\}$. All experiments were conducted on a computing platform with an Intel®5 12600K processor and an RTX-3060 12G.

4.1.4. Evaluation metrics

Two evaluation matrices are used to measure recommendation results $\{1, 0, 0.5\}$: $Recall@K$ and $NDCG@K$. K is set as 10 and 20, which means recommending the top 10 or 20 courses that are most likely to be interested. $Recall@K$ measures how many relevant courses interesting to students are included in the recommendation list. It is used to assess whether the recommendation model successfully recalls courses interesting to students. Indicators are calculated as shown in Eq. (12). $NDCG@K$ refers to Normalized Discounted Cumulative Gain at K , which takes into account the relationship between the first K recommendation results and the correlation used to measure the quality of the recommendation. The higher the indicator is, the closer the output to the real preference of students is, indicators are calculated as shown in Eq. (13). All reported experimental results are averages after 5 runs.

$$Recall@K = \frac{TP@K}{TP@K + FN} \tag{12}$$

$$NDCG@K = \frac{DCG@K}{IDCG@K} = \frac{\sum_{i=1}^K \frac{rel_i}{2^{D(i-1)}}}{\sum_{i=1}^K \frac{rel_i^*}{2^{D(i-1)}}} \tag{13}$$

where $TP@K$ denotes the number of courses with real interactions in the top K results of the recommendation list; FN denotes the number of courses with real interactions that are not in the recommendations list; $DCG@K$ refers to the discounted cumulative gain in the top K recommendation results; $IDCG@K$ indicates the discounted cumulative gain in the top K recommendation results in the ideal cases; rel_i refers to the relevance of the i th recommendation result; rel_i^* refers to the relevance of the i th recommendation result in descending order of relevance.

4.2. Performance comparison

The experimental results are presented in Table 3, Tables 4 and 5, in which $R@K$ represents $Recall@K$ and $N@K$ represents $NDCG@K$. In order to simulate the recommendation effect in the real environment, the XuetangX and ScauC datasets were simply cleaned, and the cold-start user items with less than 10 interactions were retained. In all experiments, the contrastive learning methods yielded better results. Compared to traditional models, SGL-ED and EduLGC consistently demonstrated great advantages owing to that contrastive learning can obtain embeddings with a more uniform distribution.

As shown in Table 3, the performance of EduLGC surpasses all baseline methods. The results indicate that expanding the observed interaction data by utilizing selected item static features can represent the

Table 3
Performance comparison with baselines on multi-feature datasets.

Dataset	Moyilems				Amazon			
	R@10	N@10	R@20	N@20	R@10	N@10	R@20	N@20
MF	0.1072	0.2222	0.1762	0.2389	0.0459	0.1625	0.0399	0.0945
LightGCN	0.1360	0.3382	0.2893	0.3392	0.0568	0.0792	0.0810	0.0995
IGCAGF	0.1169	0.3172	0.1825	0.3447	0.0636	0.0794	0.0851	0.0844
R3R	0.1438	0.4601	0.3039	0.4161	0.0655	0.0862	0.0872	0.0861
SGL-ED	0.1479	0.3415	0.2859	0.3408	0.0709	0.0825	0.0932	0.0917
SimGC	0.1491	0.3429	0.2718	0.3480	0.0685	0.0794	0.0811	0.0886
MRGC	0.1466	0.3468	0.2776	0.3531	0.0667	0.0728	0.0815	0.0832
EduLGC	0.1572	0.3601	0.2392	0.3501	0.0768	0.0845	0.0886	0.0943
Improvs	+5.03%	+5.01%	+2.27%	+2.34%	+3.37%	+2.42%	+5.74%	+3.54%

Table 4
Performance comparison with baselines on single-feature datasets.

Dataset	TaoBao				XuetangX			
	R@10	N@10	R@20	N@20	R@10	N@10	R@20	N@20
MF	0.4761	0.2876	0.2797	0.4225	0.4781	0.3363	0.2797	0.4215
LightGCN	0.6479	0.6343	0.6710	0.6419	0.5210	0.4076	0.6158	0.4796
R3R	0.6110	0.6261	0.6468	0.6057	0.4948	0.3945	0.5246	0.4911
SGL-ED	0.6090	0.6020	0.6065	0.6015	0.4719	0.4161	0.4991	0.5091
SimGC	0.6083	0.6020	0.6050	0.6006	0.2952	0.4730	0.4374	0.5070
MRGC	0.6077	0.6012	0.6068	0.6010	0.5312	0.4527	0.6067	0.4669
EduLGC	0.6089	0.6023	0.6065	0.6014	0.5197	0.4701	0.6421	0.5038
Improvs	-0.64%	-0.13%	-0.08%	-0.08%	+0.39%	-0.61%	+0.77%	-0.61%

Table 5
Performance comparison with baselines on real educational scenario dataset.

Dataset	ScauC			
	R@10	N@10	R@20	N@20
MF	0.1904	0.1175	0.2381	0.1275
LightGCN	0.1767	0.1192	0.2558	0.1402
IGCAGF	0.1839	0.1190	0.2645	0.1410
R3R	0.1705	0.1170	0.1448	0.1460
SGL-ED	0.1829	0.1196	0.2708	0.1414
SimGC	0.1810	0.1191	0.2497	0.1406
MRGC	0.1793	0.1209	0.2534	0.1451
EduLGC	0.1895	0.1238	0.2724	0.1490
Improvs	+1.61%	+2.99%	+2.99%	+3.23%

distance between two target nodes from different perspectives. This effectively discovers users with similar interests, thus better representing user preferences and enabling personalized recommendations.

As illustrated in Table 5, employing a single static feature to expand the interaction paths alone renders no better method to explore and minimize the paths between two user nodes. In this case, utilizing the expanded interaction graph for global information extraction is significantly hampered by the ambiguity issue. Furthermore, challenges such as data noise and overly short interaction sequences persist. Consequently, the overall performance of EduLGC on this dataset is somewhat weaker.

As presented in Table 5, EduLGC enhances the performance of recommendation metrics in the real-world interactive dataset ScauC. The method of combining existing student-course interaction history with static features to generate a comprehensive relationship view strengthens the interpretability of the recommendation outcomes, enabling the provision of socially responsive and personalized course recommendations for students.

4.3. Ablation experiment

To explore the importance of contrastive learning and static feature enhancement in EduLGC, ablation experiments are performed to validate and analyse the performances of the following variants: EduLGC without static feature enhancement (w/o info) and EduLGC without

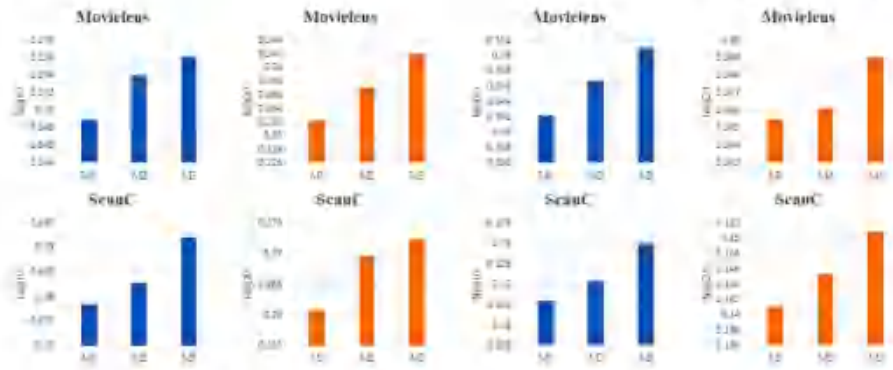


Fig. 4. The results of EduLGCN on MovieLens and ScauC datasets with various feature combinations.

Table 6

The ablation results of EduLGCN and its variants on the XuetangX and Amazon

Dataset	Metric	w/o CL	μ -value	w/o Info	μ -value	EduLGCN
Amazon	AUC	0.5010	0.0043	0.0944	0.0033	0.0740
	F1	0.0557	0.0020	0.0592	0.0045	0.0545
	AP	0.0533	0.0047	0.0510	0.0012	0.0580
XuetangX	AUC	0.5208	0.0019	0.5222	0.0001	0.5407
	F1	0.4543	1.62E-06	0.4479	2.87E-05	0.4760
	AP	0.5286	0.0003	0.6177	1.52E-07	0.6451
ScauC	AUC	0.4903	1.40E-02	0.4865	1.13E-02	0.5306

contrastive learning (w/o CL) for ScauC, XuetangX, and Amazon. Especially, the situation without the supplementation of student graph information (w/o student) is tested for the ScauC dataset. The results of the ablation experiments are shown in Table 6 and 7, where a p -value less than 0.05 indicates that there is a statistically significant difference between the EduLGCN and the recommended metrics of each variant in terms of their significance in the context of statistical testing.

The variant without the contrastive learning (w/o CL) was employed, retaining the information expansion module, making the EduLGCN structure closely resemble LGCAF. Further removal of the connection expansion module from EduLGCN (w/o info) resulted in a structure similar to LightGCN. It can be observed that the removal of the contrastive learning module led to the direct utilization of global node representations for similarity computation across the four metrics on the Amazon and XuetangX datasets, resulting in a decrease in performance due to the presence of global noise. Moreover, the removal of the connection expansion module resulted in a scenario where only the interaction history was available for model learning. A further decline in performance is caused by missing comprehensive representations.

The variant without student graph (w/o student) for the ScauC dataset includes only the contrastive learning module and the information expansion module. Due to the presence of a substantial number of Freshman courses with interaction history lengths less than five, the model could not effectively validate whether the node representations are learned, thus affecting the optimization performance in subsequent iterations. Consequently, a notable decrease in performance was observed under this circumstance.

4.4. Performance of EduLGCN with different features

In order to explore the importance of mutual information static feature selection in EduLGCN, we used different numbers of feature

combinations on different datasets in order to conduct the experiments. In the MovieLens dataset, there are four available item features: movie, genre, director, and star. Different combinations of features were selected for experimentation under the condition of the fixed selection of "Movie". There are five available features in the ScauC dataset and many combinations of features cannot be manually filtered, even with a fixed selection of "course" features. The experimental results on MovieLens and ScauC are illustrated in Fig. 4 for the three cases of the highest theoretical mutual information value, the lowest theoretical mutual information, and the best performing feature combination.

As depicted in Fig. 4, M1 denotes the situation with the lowest theoretical mutual information values, specifically indicating feature combinations below one or two-feature. M2 represents the situation with the highest theoretical mutual information values, encompassing full feature combinations or those greater than one. M3 represents the optimal performing feature combination.

Overall, a higher number of classifying features in a given type indicates a greater mutual information value between the feature and user characteristics. However, selecting the feature combination with the highest mutual information value does not necessarily yield the best results. This is because feature blurring may occur, causing difficulties in distinguishing the differences between two closely related challenging negative samples in the embedding space. For instance, the combination (college1, major1) and (college1, major2) might result in being lack of sufficient training samples for the model to differentiate between major1 and major2 effectively within the embedding space, which led to blurring.

Furthermore, inherent noise may be introduced from various scenarios. So, the combination of a few features with the highest mutual information value and minority features with the lowest mutual information value can often yield favourable results.

4.5. Validation on user cold start problem

In the context of recommendation systems, users who have fewer interactions with items are referred as "cold users". Recommending items to cold users who have limited interaction history is known as a "cold start problem". In educational scenarios, cold users specifically embody a freshman entering school without interaction with any courses, making it difficult to capture students' interests and preferences. The users in the XuetangX and ScauC datasets are grouped according to the length of their interaction history. The experimental results are shown in Fig. 5.

Different user groups are defined based on varying levels of user interaction. Group 0 represents users whose interaction frequencies

Table 7
The ablation results of EduLGCN and its variants on the ScauC.

Dataset	Metric	w/o CL	ρ -value	w/o info	ρ -value	w/o student	ρ -value	EduLGCN
ScauC	RMSE	0.1712	$1.10E-06$	0.1757	$0.26E-07$	0.1767	$3.14E-06$	0.1695
	NDC@10	0.1094	$5.92E-07$	0.1162	$1.89E-07$	0.1006	$5.12E-07$	0.1277
	Rec@20	0.2648	$3.79E-07$	0.2613	$8.12E-07$	0.2220	$1.87E-05$	0.2924
ScauC	NDC@20	0.1403	$2.22E-07$	0.1445	$2.36E-07$	0.1270	$1.22E-06$	0.1474

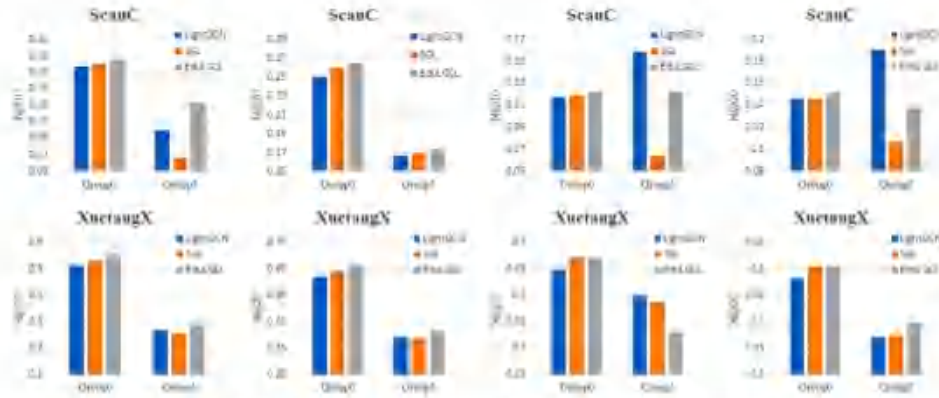


Fig. 5. The comparison results of various models on different user groups in the ScauC and XuetangX datasets.

range from 1 to 5 times, which corresponds to freshman in the educational scenario. Group 1 comprises users whose interaction frequencies range from 5 to 10 times, which represents a mix of seniors and some freshmen.

It is obvious in Fig. 3 that EduLGCN exhibits excellent performance in the freshman group's metrics for the ScauC dataset, which extends the relationship between student nodes using student clustering relationships and course static features. However, its performance in recommendation quality is weaker than that of the conventional LightGCN model for the senior group. Although EduLGCN recommends courses of interest to students, these courses might not accurately represent the students' true preferences. It could reflect differences between the inferred similar preferences of student clustering and the actual preferences learned by the model from existing data.

Even though EduLGCN is weaker than SGL in overall performance for XuetangX dataset, it can assist in solving the cold-start problem. Despite the relatively limited number of available course features for expansion, these features are helpful for connecting with cold-start users. Moreover, since the number of cold-start students in XuetangX is small, SGL's edge dropout has a limited impact on this user category. However, an interesting phenomenon arises in the senior group, that is, traditional models outperform contrastive learning models in recommendation quality. It could be attributed to the number of students in this range, as the limited additional information introduced by contrastive learning might result in redundancy for the recommendation task within this range.

4.6. Hyperparameter analysis for educational scenario

In order to investigate the effect of adding contrastive learning intensity on the performance of EduLGCN, the proportion of adding contrast learning λ_c is set as the set of $\{0.2, 0.1, 0.05, 0.01\}$ and the strength

of adding negative sample r is set as the set of $\{0.2, 0.175, 0.15, 0.1\}$. The experimental results with different proportions and strengths are demonstrated in Fig. 5. The comparison is made by fixing the same negative sampling strength while varying the proportions, and by fixing the same proportions while varying the strengths.

As shown in Fig. 6, when incorporating the same proportion of negative samples in the ScauC and XuetangX datasets, the model's performance tends to be improved as the proportion of contrastive learning decreases. It validates that adding a small amount of the global interest representation of students can promote the performance of recommendations. With a fixed proportion of contrastive learning, a higher strength of negative sample incorporation leads to more difficult negative samples being added. In this case, as the number of negative samples decreases, the model's performance also degrades. This phenomenon may be related to the sparsity of the educational dataset. Additionally, the issue of feature ambiguity caused by multiple feature combinations may affect the model's ability to distinguish difficult negative samples.

5. Conclusion

This study addresses the data sparsity and user cold-start issues faced by recommendation systems in educational scenarios, presenting a graph-based contrastive learning recommendation model named EduLGCN. The model employs the static features of students and courses to expand the sparse interaction data, adding more neighbor nodes to student nodes with fewer interactions, and supplementing their learning interest representations. It describes students' interest preferences from diverse perspectives based on the static features of courses, mitigating the user cold-start problem in recommendation systems. By introducing a contrastive learning method, the model integrates local and global student interest representations, alleviating the data sparsity

- [19] Xiang Wang, Xingqun He, Meng Wang, Fuli Feng, Tai-Seng Chua, Neural graph collaborative filtering, in: Proceedings of the 42nd international ACM SIGIR conference on Research and development in information retrieval, 2019, pp. 165–174.
- [20] Xiangqun He, Kun Deng, Yong Wang, Yan Li, Yongqiang Zhang, Meng Wang, LightGC: Simplifying and powering graph convolution network for recommendation, in: Proceedings of the 43rd International ACM SIGIR conference on research and development in information retrieval, 2022, pp. 639–648.
- [21] Rex Ying, Raining He, Kaifeng Chen, Peng Eksombatchai, William L. Hamilton, Jure Leskovec, Graph convolutional neural networks for web-scale recommender systems, in: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, 2018, pp. 974–983.
- [22] Will Hamilton, Zhifu Yang, Jure Leskovec, Inductive representation learning on large graphs, in: Advances in neural information processing systems, vol.30, 2017.
- [23] Le Wu, Xiehe Sun, Yanjie Fu, Wotang Hong, Xing Wang, Meng Wang, A neural influence diffusion model for social recommendation, in: Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, 2019, pp. 235–244.
- [24] Tianchang Yin, Xinyang Yu, Derek Zhiyuan Cheng, Yibo Yu, Ting Chen, Aditya Menon, Lichun Huang, Ed H Chi, Steve Joo, Jinyi Kang, et al., Self-supervised learning for large-scale item recommendations, in: Proceedings of the 20th ACM International Conference on Information & Knowledge Management, 2021, pp. 4321–4330.
- [25] Jianan Wu, Xiang Wang, Fuli Feng, Xiangqun He, Liang Chen, Jiaxian Lian, Xing Xie, Self-supervised graph learning for recommendation, in: Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval, 2021, pp. 726–735.
- [26] Jianliang Yu, Huihui Yin, Xun Xun, Tong Chen, Lichen Cui, Quoc Van Hung Nguyen, Are graph augmentations necessary? simple graph contrastive learning for recommendation, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 1294–1303.
- [27] Zhan Lin, Guanglin Tian, Yupeng Hou, Wayne Xin Zhao, Improving graph collaborative filtering with neighborhood-enriched contrastive learning, in: Proceedings of the ACM Web Conference 2022, 2022, pp. 2320–2329.
- [28] Lingshao Xia, Chao Huang, Yang Xu, Jianshi Zhao, Lowsi Yu, Jianyi Huang, Hypergraph contrastive collaborative filtering, in: Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval, 2022, pp. 70–79.
- [29] Xuheng Chi, Chao Huang, Lingshao Xia, Xuhui Ren, LightGC: Simple yet effective graph contrastive learning for recommendation, 2023, arXiv preprint arXiv:2302.08191.
- [30] Xiaohui Zou, A survey on application of knowledge graph, J. Phys. Conf. Ser. 1487 (2020) 012016.
- [31] Hao Zhang, Heng Yang, Tao Huang, Guoqing Zhao, DBNCR: Personalized course recommendation system based on DBN in MOOC environment, in: 2017 International symposium on educational technology, ISET, 2017, pp. 106–108.
- [32] Jie Lian, Wei Zhou, Fengli Luo, Junhao Wen, Min Gao, Xuhua Li, Jun Zeng, SocialIGN: Light graph convolution network for social recommendation, Inform. Sci. 589 (2022) 595–607.
- [33] Zhihui Zhou, Lili Zhang, Ning Yang, Contrastive collaborative filtering for cold-start item recommendation, 2023, arXiv preprint arXiv:2302.02131.
- [34] Yijun Zhao, Fajian Jiang, Jiafeng Wang, MPGL: Light graph contrast learning with multi-format information for recommendation, in: 2023 International Joint Conference on Neural Networks, IJCNN, 2023, pp. 01–08.
- [35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, in: Advances in neural information processing systems, vol.30, 2017.
- [36] Minghao Tang, Peng Zhou, Shaobo Li, Yuanmeng Zhang, Jianyi He, Aoxi Zhang, Multi-head multimodal deep matrix recommendation network, Knowl.-Based Syst. 130(5) 110989.
- [37] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, Lars Schmidt-Thieme, BPR: Bayesian personalized ranking from implicit feedback, 2012, arXiv preprint arXiv:1205.2618.
- [38] Michael Gormann, Aapo Hyvärinen, Noise-contrastive estimation: A new estimation principle for unnormalized statistical models, in: Proceedings of the thirtieth international conference on artificial intelligence and statistics, 2016, pp. 297–304.
- [39] Donghui Mei, Niu Huang, Xin Li, Light graph convolutional collaborative filtering with multi-object information, IEEE Access 9 (2021) 34433–34441.
- [40] F. Maxwell Harper, Joseph A. Konstan, The movielens dataset: history and context, ACM Trans. Interact. Intell. Syst. (TIIS) 5 (4) (2015) 1–19.
- [41] Raining He, Julian McAuley, Up and down: Modelling the visual evolution of fashion trends with one-class collaborative filtering, in: Proceedings of the 25th international conference on world wide web, 2016, pp. 507–517.
- [42] Jang Zhang, Bowen Hao, Be Chen, Cuiqing Li, Heng Chen, Jianing Sun, Hierarchical reinforcement learning for course recommendation in MOOCs, in: Proceedings of the AAAI conference on artificial intelligence, 2019, pp. 435–442.
- [43] Han Zhu, Xiang Li, Pengye Zhang, Guozhong Li, Ji He, Han Li, Kim Gu, Learning tree-based deep model for recommender systems, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 1079–1088.
- [44] Dongha Lee, Seongju Kang, Hyeonjun Ju, Chanyoung Park, Hwanil Yu, Bootstrapping user and item representations for one-class collaborative filtering, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 317–328.
- [45] Jiepeng Tang, Chib-Ming Chen, Chuan-Hsi Wang, Ming-Feng Tsai, HDP-IR: High-order proximity for implicit recommendation, in: Proceedings of the 12th ACM conference on recommender systems, 2018, pp. 140–144.

3.2. DFCCNet: A Dense Flock of Chickens Counting Network Based on Density Map Regression





animals



Article

DFCCNet: A Dense Flock of Chickens Counting Network Based on Density Map Regression

Jinze Lv, Jinfeng Wang, Chaoda Peng and Qiong Huang

Special Issue

Novel Technology in Poultry Production and Nutrition: Role in Disease Prevention, Performance and Welfare Improvement and Sustainability

Edited by

Dr. Vasilios Tsiouris, Dr. Georgios C. Papadopoulos, Dr. Ilias Giannenas and Prof. Dr. Paschalis Fortomaris



<https://doi.org/10.3390/ani13233729>

Article

DFCCNet: A Dense Flock of Chickens Counting Network Based on Density Map Regression

Jinze Lv ¹, Jinfeng Wang ^{1,2,*} , Chaoda Peng ¹ and Qiong Huang ^{1,2,*} 

¹ College of Mathematics and Information, South China Agricultural University, Guangzhou 510642, China; 1170062@stu.scau.edu.cn (J.L.); chadapeng@scau.edu.cn (C.P.)

² Key Laboratory of Smart Agriculture of Guangzhou, South China Agricultural University, Guangzhou 510642, China

* Correspondence: wangjinfeng@scau.edu.cn

Simple Summary: In large-scale chicken farming, monitoring flock density can help optimize feeding management and improve animal welfare. Traditional manual counting methods are not only prone to errors and omissions but also can cause harm to chickens. In this paper, a counting model for dense chicken flocks is proposed. The method can effectively estimate the number of chickens in real dense scenes.

Abstract: With the development of artificial intelligence, automatically and accurately counting chickens has become a reality. However, insufficient lighting, irregular sizes, and dense flocks make this a challenging task. The existing methods cannot perform accurate and stable counting. In this article, a dense flock of chickens counting network (DFCCNet) is proposed based on density map regression, where features from different levels are merged using feature fusion to obtain more information for distinguishing chickens from the background, resulting in more stable counting results. Multi-scaling is used to detect and count chickens of various scales, which can improve the counting accuracy and ensure stable performance for chickens of different sizes. Feature convolution kernels are adopted to convolve feature maps, which can extract more accurate target information, reduce the impact of occlusion, and achieve more reliable and precise results. A dataset of dense flocks of chickens (namely Dense-Chicken) has been collected and constructed, which contains 600 images of 99,916 chickens, with labeled points and boxes. It can be accessed by researchers as benchmark data. The proposed method was compared with some state-of-the-art algorithms, to validate its effectiveness. With its robustness being verified by counting in three kinds of density situations, with the mean absolute error being 4.26, 9.85, and 19.17, respectively, and a speed of 16.15 FPS, DFCCNet provides an automatic and fast approach to counting chickens in a dense farming environment. It can be easily embedded into handheld devices for application in agricultural engineering.

Keywords: artificial intelligence; chicken counting; density map regression; feature fusion; multi-scaling



Citation: Lv, J.; Wang, J.; Peng, C.; Huang, Q. DFCCNet: A Dense Flock of Chickens Counting Network Based on Density Map Regression. *Animals* **2023**, *13*, 3729. <https://doi.org/10.3390/ani13233729>

Academic Editor: Sebastian Buzvinski

Received: 23 September 2023

Revised: 23 November 2023

Accepted: 26 November 2023

Published: 1 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Chicken is one of the most popular meats in the world. Therefore, the chicken breeding industry has great economic value. In the process of chicken breeding, cageless rearing is more conducive to the growth of chickens than cage rearing [1,2]. Proper breeding density can improve the growth performance of chickens, as well as their immunity and carcass yield [3–5], while too intensive feeding can negatively affect the health of chickens [6,7]. Moreover, many steps such as welfare breeding [8], feed feeding, stocking, and slaughtering are needed to obtain accurate quantities. Therefore, rapid and accurate estimation of flock density is a very important research field.

Traditional poultry farming requires artificial observation of behavior and health status, manual feeding, and counting. This requires a lot of labor, which increases the cost

of breeding. An irregular and inaccurate manual operation may cause harm to chickens. Moreover, due to the small space and high density of chickens, if the staff do not carry out complete disinfection before entering the chicken house, they may pass germs to the chickens, which spread quickly, causing serious losses [9]. It is obviously difficult for chicken farm managers to count manually, which can lead to wrong amounts, low efficiency, and subjective influences. In particular, overlapping chickens and different perspectives can cause large errors with complex backgrounds. These problems are common in large-scale breeding enterprises. Therefore, it is necessary to develop and implement efficient, automated, and highly accurate counting methods, so as to fully improve efficiency and truly realize intelligent and automated management.

Digital technology with intelligent monitoring methods is widely used in poultry health and welfare management [10], which can realize rapid, accurate, automatic, non-invasive monitoring in the process of poultry breeding, and help replace some labor-intensive tasks in poultry breeding. A novel fully automated and non-invasive monitoring system was proposed to detect lame behavior according to the important correlation between characteristic variables and broiler gait scores [11]. A real-time automatic feed weight monitoring system was developed to automatically detect the intake and weight of a single turkey; to study the feed conversion rate and feeding behavior of a single turkey in a population environment [12]. A segmentation algorithm could effectively separate a broiler from the background, extract the pose information of the broiler, and accurately and quickly identify the health status of the broiler with an accuracy rate of 99.47% [13]. However, these methods belong to shallow learning, which limits performance when dealing with complex problems, due to a lack of deep topology and big data support.

Deep learning is a method of data representation-based learning that consists of multiple processing layers, to learn data representations with multiple levels of abstraction [14]. An attention encoder and convolutional neural network (CNN) were used to detect and classify chicks in different living stages, with an accuracy of 95.2% [15]. CNN was used to monitor the opening beak behavior, spatial dispersion, and movement of chickens [16]. A deep neural network (DNN) and cross-entropy in information theory were used to train rooster and hen classifiers with an average accuracy of 96.85%, which provided a feasible method for the estimation of sex ratio [17]. A camera-based system was developed to monitor chickens and detect injuries caused by pecking at each other using neural networks [18]. A DNN-based pose estimation method was first applied to classify and identify the poses of broilers [19]. Experiments showed that chickens in standing, walking, running, eating, and resting states could be identified.

Counting refers to estimating the number of objects in a target area, to obtain information and perform timely control operations, and is currently widely used in various fields [20–22]. In the field of chicken counting, some researchers have conducted related studies. One study applied a localized fully convolutional network (LCFCN) to count and locate chickens from images of the pens with an average absolute error (MAE) of 0.56 [23]. A fully convolutional network termed DenseFCN was designed, and a point-supervised method was used to count chickens in the image, with an accuracy of 93.84% and speed and 9.27 frames per second (FPS) [24]. A novel framework called YOLOX-birth growth death (Y-BGD) was proposed for automatic and accurate cage-free broiler counting, with an average accuracy of 98.13% [25]. A automatic system based on YOLOv5 was applied to count chickens in the image with an average accuracy rate of 95.87% [26]. We analyzed the datasets used in these studies, in which [23–25] counted chickens in sparse scenes, with an average number of chickens per image of 27.89, 24.36, and 2.38, respectively. Ref. [26] counted chickens in both sparse and dense scenes, with an average number of chickens per picture of 155.37. However, the counting objects in all of these studies were large chickens. In real farming environments, large and small chickens are not the same in farming scenes; the same number of small chickens are farmed in smaller areas, and small chickens are more likely to congregate together, thus creating more serious occlusion and shadowing. More research is needed to address this situation.

At present, the mainstream counting technologies can be divided into two types, namely object detection and density map regression. Object detection can identify objects in images and determine their size and position using prediction boxes. The number of prediction boxes can be used to count. Object detection is mainly applied to two counting scenes: one is to directly detect and count chickens in sparse scenes [26], and the other is to count chickens passing through the counting line in the corridor by combining tracking algorithms [25]. But the accuracy is greatly affected in dense environments with heavy occlusion. Density map regression is mainly used in high-density scenes, which can obtain accurate counting by learning the mapping relationship between the picture and the real density map. This method has been applied in aquaculture. A lightweight fish counting model based on density plot regression was used for counting high-density fish [27]. Multi-scale modules and attention mechanisms were integrated into a network based on density map regression in fish counting, with an accuracy of 97.1% [28]. Although these experiments were conducted in a dense environment, their experimental environment was also ideal. Serious occlusion and complex backgrounds could still lead to poor stability and accuracy of the model, and it is necessary to further improve the performance.

Lightweight models are crucial in practical applications. Deep learning models usually require a large amount of computational resources, such as CPU, GPU, and so on. If the computational burden is too heavy, this will cause the model to run slowly, or even not be able to run on resource-constrained devices. This will make the model unable to meet the requirements of real-time calculation and portability, hindering its application in practice. Therefore, it is important to take into account the weight of the model, while ensuring the correctness of the model.

Aiming at avoiding heavy computational effort and further improving the chicken counting accuracy under serious occlusion and complex environments, a lightweight model based on density map regression is proposed. The main contributions can be summarized as follows:

- (1) A lightweight framework, DFCCNet, was designed by improving a feature convolution kernels module and proposing a density map module, satisfying the requirements for fast computing power and high detection accuracy in dense-flock chicken-counting tasks;
- (2) Feature fusion is adopted to obtain more feature information with insufficient lighting; a multi-scaling mechanism can be used for solving irregular sizes; and feature convolution kernels are employed to address serious occlusion;
- (3) A self-built dataset called Dense-Chicken was collected and constructed in a complex feeding environment, which contains 600 images of 99,916 chickens, in-dot annotation files and box annotation files. It can be shared for researchers.

The rest of this article is organized as follows: Section 2 provides an overall description of the proposed models and methods. Section 3 describes the details of dataset preparation, which was used to train and evaluate the proposed network. The experimental results and a performance analysis of the proposed network are presented. Finally, Section 4 concludes the work by providing a summary and prospects of the proposed technique.

2. Materials and Methods

In this section, the overall framework of DFCCNet is first introduced, and then the feature convolution kernel generation module and counting module are explained. Finally, the loss function is demonstrated. The code is available at <https://github.com/2226450990/DFCCNet> (accessed on 22 November 2023).

2.1. Overall Framework

The overall framework of DFCCNet is shown in Figure 1. First, the original image and the cropped image are fed into the backbone module for feature extraction, to obtain image features and cropped image features, respectively. Then, in the correlation map generation module, the feature map of the cropped image is used as a feature convolution kernel

to convolve the image features, to generate the correlation map. Finally, in the density map generation module, the resolution of the correlation map is restored to the original image size and the density map is generated. In particular, the parameters of the feature convolution kernel remain constant throughout training and testing, and it only needs to be generated once before the counting model is trained. The model makes direct calls to the feature convolution kernel during training and testing. Therefore, depending on whether the model parameters are changed and the order of execution, DFCCNet is divided into two parts: the feature convolution kernel generation module, and the counting module.

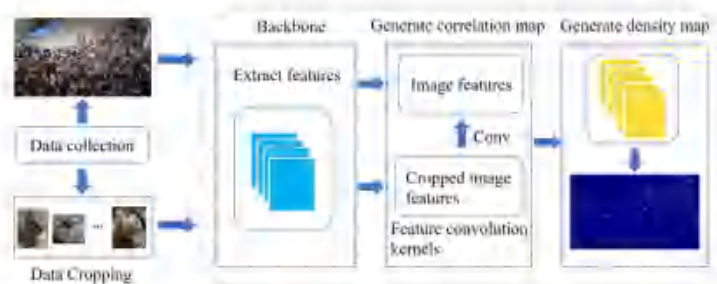


Figure 1. The overall framework of DFCCNet.

2.2. Feature Convolution Kernel Generation Module

In this study, the images contain a lot of useless background information that can adversely affect the counting of chickens. In order to count objects from different categories, Ref. [29] proposed an adaptive method. The input information for this method consists of the image and the box information of the counting target. Before generating the density map, image features located within these boxes are used as convolution kernels to perform convolution operations with the overall image features, to generate a correlation map focused on the counting objects. Inspired by this approach, feature convolution kernels are generated for enhancing the correlation between counting objects and feature maps, while filtering useless background information. Unlike the original method, in this study, since the counting objects are chickens of the same breed and age, and the morphological differences between different chickens in the images are small, generating different convolution kernels to improve the correlation of the feature maps during training was unnecessary. In this paper, the feature convolution kernel generation module is independent as the pre-module for counting, and the structure of this module is shown in Figure 2, which first adopts multiple images of a single chicken for feature extraction, and then the feature maps of different images are resized to the same size and fused to form a feature convolution kernel. During the whole model training and testing period, the feature convolution kernel only needs to be generated once. Although this can reduce the computational consumption, the quality of the feature convolution kernel needs to be considered. ResNet [30] has a wide range of applications in agriculture due to its excellent model performance [31–33]. ImageNet is a large-scale dataset, and by pretraining on this dataset, a model can be given good feature extraction capability [34]. Therefore, the pretrained ResNet in the feature convolution kernel generation module is used for feature extraction, to ensure the quality of the convolutional kernel. Various versions of ResNet were used for comparative experiments, and finally ResNet50 was selected.

Although the experimental subjects were all chickens of the same breed and age, they presented different sizes in the images. Single-scale feature extraction can lead to the loss of multi-scale information. In order to solve this problem, multi-scale scaling methods are widely adopted [35,36]. In this paper, multi-scale scaling is applied to the feature convolution kernel, which ensures that the subsequently generated correlation

maps can focus on the multi-scale chicken information, and thus accurately count chickens of different sizes in the images.

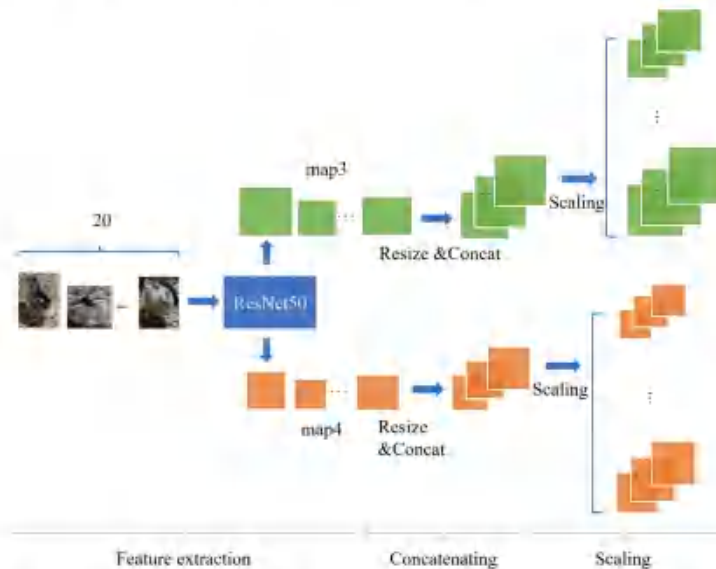


Figure 2. The structure of the feature convolution kernel generation module. The whole process is divided into three stages: feature extraction, concatenating, and scaling. Where the green and orange blocks represent the output of the third and fourth residual blocks, respectively.

2.3. Counting Module

The framework of the counting module is presented in Figure 3. First, the image is input to ResNet50 to extract features, and feature convolution kernels are used to generate the correlation maps. Then, the correlation maps are adjusted to the same size and concatenated together for input into the density map prediction module. Finally, the predicted density map is generated. During the training phase, the loss between the predicted density map and the ground truth is calculated to update the model parameters. During the testing phase, the sum of the pixel values of the predicted density map is used as the counting result. In the counting module, two strategies are used to improve the counting performance.

(1) Transfer learning. In this research, data annotation became a time-consuming and cumbersome task, due to the high density and complex environment. Although a significant amount of time and effort was spent on data annotation, limited data still affected the performance of the model. Transfer learning provided us with new possibilities and can effectively apply knowledge learned in one task or field to other related tasks or fields, thereby greatly improving the learning ability and generalization of the model [37]. Through transfer learning, the trained models could be used as a base to quickly retrain new tasks, without the need to train the entire network from scratch. This method not only saved a lot of time and computational resources, but also improved the performance of the model under limited data conditions [38,39]. As with the feature convolution kernel generation module, the pretrained ResNet50 is used for feature extraction.

(2) Feature fusion. The feature maps output from different depths of the convolutional layers have different information: shallow feature maps are rich in spatial information, and deep feature maps are rich in semantic information. Fusion of feature maps from

different layers can improve network representation and enhance information acquisition, effectively improving model performance [40,41]. In this paper, the feature maps and feature convolution kernels output from the third residual block and the fourth residual block of ResNet50 are utilized to generate the correlation maps of different layers, which are fused to obtain the final correlation map. This correlation map can be used to generate a high-quality density map for counting.

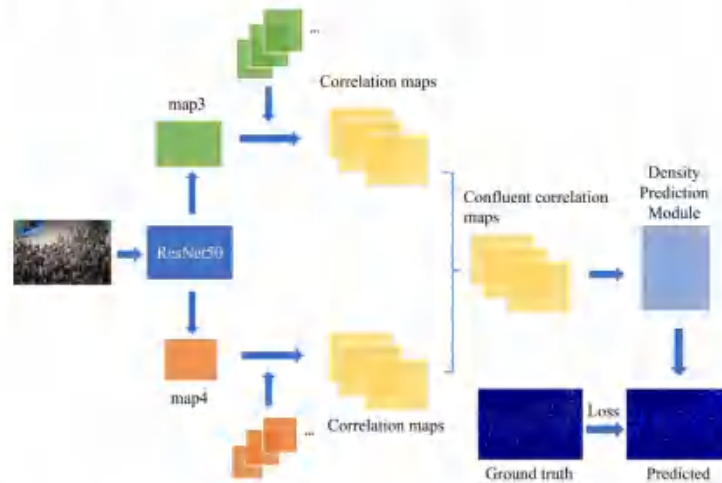


Figure 3. The structure of counting module.

The details of the counting module are shown in Figure 4, which is divided into three modules; i.e., feature extraction, correlation map generation, and density map prediction. The feature convolution kernels and feature map obtained by the third block are used to obtain a series of correlation maps. The same operation is performed on the fourth block. In order to finally concatenate the correlation maps, the correlation maps output by the fourth block need to be upsampled. In the density map prediction module, five convolution kernels and three upsampling operations are used, where the first four convolution kernels are designed to restore the output size to the input size and the last convolution kernel is designed to predict the density map. Finally, the sum of the pixel values of the density map is used as the counting result.

2.4. Loss Function

In this paper, mean squared error (MSE) is used as the main body of the loss function, which means the average of the squared distance between the predicted value and the true value. MSE is defined in Equation (1).

$$MSE = \frac{1}{N} \sum_{i=1}^N (Z_i - \hat{Z}_i)^2 \quad (1)$$

where N represents the number of images, Z_i represents the real density map of the i -th image, and \hat{Z}_i represents the predicted density map. However, MSE is sensitive to outliers and susceptible to noise (including different sizes, occlusion, etc.) in the dense-counting field, which leads to inaccurate prediction in high-density areas. Thus, we added a constraint to the MSE that uses the random occlusion target for additional MSE to reduce

noise interference. The random occlusion target is labeled on each image in the training set. The loss function is defined in Equation (2).

$$Loss = MSE + \lambda MSE(occlusion_target) \quad (2)$$

The former MSE calculates the error between the predicted density map and the true density map. The latter part is the error between the predicted density map corresponding to the random occlusion target and its real density map. λ is a hyperparameter set as 1×10^{-9} .

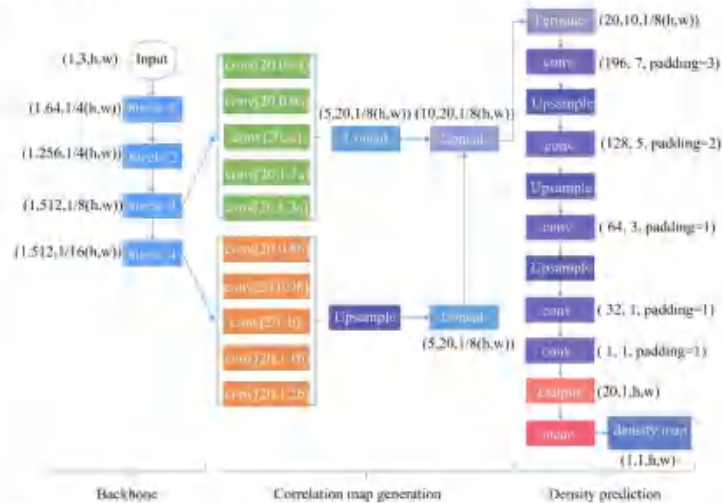


Figure 4. Details of the counting module. In this figure, h and w refer to the height and width of the input image, and a and b refer to the characteristic convolution kernel sizes of the output of the third and fourth blocks. Since the feature convolution kernel generated earlier uses 20 cropped images, the dimension of the convolution kernel is 20 in the correlation map generation module. The scale_factor for all upsampling operations was set to 2.

3. Experiments and Discussion

In this section, a series of experiments are reported. (1) Each strategy was incrementally added to the counting network to determine its effectiveness, and the most appropriate parameter settings were identified. (2) The validation and testing sets were divided into different densities to verify the robustness of the model. (3) Comparison experiments between DFCCNet and some state-of-the-art methods were executed, to show the counting performance of DFCCNet.

3.1. Dataset

The experimental data were collected from a chicken farm in Guangdong Province in China in October 2021. The collected subjects were week-old chickens. The equipment for data acquisition is shown in Figure 5. The height of the chicken farm fence was 50 cm. During the data collecting process, our devices mainly included a mobile phone with the function of recording video and a bracket. The bracket was used to fix the mobile phone during the video process, and the length of the bracket could be adjusted in the range of 50–100 cm. Then, diverse data could be obtained from different angles in different breeding areas, which enhanced the generalization ability of the model. Finally, we obtained

20 videos with a resolution of 1280×720 pixels and a duration of 3 min, which had a frame rate of 30 frames per second.



Figure 5. Experimental data acquisition device.

3.1.1. Dataset Construction

A total of 1800 images were intercepted in videos by cutting 1 frame every 60 frames. All images were then examined and images with motion blur and a similar chicken distribution were removed, retaining a total of 600 images. The data were divided into a training set, validation set, and testing set, in a ratio of 7:1:2. And the images in the different sets were from different videos. The distribution of the dataset is shown in Table 1. In the data labeling stage, two image annotation tools, i.e., Labelme and LabelImg, were adopted to obtain JSON files containing point coordinate information and XML files containing box coordinate information. The representative dataset is available at https://stuscaueducn-my.sharepoint.com/:x/g/personal/3170062_stu_scau_edu_cn/ETT-vDigmvZBu6EgSRtSn0sBnNHLojY_tDmiVaoZteVP3g?e=rGa2yO (accessed on 15 December 2022).

Table 1. The amounts and distribution of the proposed dataset.

Datasets	Images	Chickens	Average
Complete dataset	600	99,916	166
Training Set	416	70,777	170
Validation Set	61	9657	158
Testing Set	123	19,482	158

Some scenes in the dataset are shown in Figure 6. The challenges of counting a dense flock of chickens are presented as follows:

- (1) Complex background in the pictures: Chickens in different areas are exposed to different intensities of light, in which it is difficult to distinguish chickens from the background;
- (2) Occlusion situation: Since some feeders are placed in the breeding area, these devices can block the chickens to some extent. On the other hand, the high density of chickens and their habit of gathering also cause occlusion;
- (3) Pixel blur: Chickens like to move and sometimes move quickly, which can cause pixel blur;
- (4) Different scales: Since the object has the characteristics of being large in proximity and small in the distance, there will be many chickens of different sizes in the image. This requires a model to have strong feature extraction capabilities;
- (5) Different numbers: The density and number of chickens in different regions are different, which will challenge the generalization of a model.



Figure 6. Images and annotation files in different scenes.

3.1.2. Data Preprocessing

The spatial distribution information of the chickens needed to be generated according to the ground truth before training. Gaussian smoothing with an adaptive size window was used to generate a density map. First, a point annotation map was given, in which each point was located in the approximate center of the object. Second, the distances between each point and its nearest neighbor were calculated and averaged. Finally, this average distance was used as the size of the Gaussian window to generate the target density map.

In the model, feature convolution kernels fusing the features of single chickens were adopted, to generate correlation maps with better correlation with chickens.

Therefore, single chickens with different sizes and orientations were obtained by cropping images. The process of cropping image is shown in Figure 7.



Figure 7. Image cropping. The chickens cut out are intact, with different sizes and directions.

3.2. Experimental Environmental and Metrics

The experimental environment of DFCCNet is listed in Table 2. The PyTorch framework for deep learning was adopted.

Table 2. Experimental environment.

Configuration	Parameter
CPU	Intel Core i7-12700
GPU	NVIDIA RTX A5000
Operating system	Windows10
Deep learning framework	Pytorch 1.11.0
Programming language	Python 3.8

In this paper, mean absolute error (MAE), root mean squared error (RMSE), and mean normalized absolute error (NAE) were used as the main evaluation metrics in the counting experiments [42–44]. MAE is the average error between the predicted values and ground truth in counting, which was used to evaluate the accuracy of the model. RMSE is the dispersion of the error between the predicted values and ground truth, reflecting the stability of the model. NAE is the normalized MAE. They are defined in Equations (3), (4) and (5), respectively.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (4)$$

$$NAE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \quad (5)$$

where N is the number of images, y_i represents the real number of chickens in the i -th image, and \hat{y}_i represents the predicted number of chickens in the i -th image. The smaller MAE and RMSE, the higher the accuracy and stability of the model.

In addition, the parameter amount and frames per second (FPS) were used for evaluating the performance of the model. Parameter amount determines the size of the model. If a model is for use on a mobile platform, fewer parameters should be set to meet lightweight requirements. FPS was used to test the efficiency and speed of the model.

3.3. Model Validation with Different Strategies

In this section, transfer learning, feature convolution kernel, feature fusion, and improved loss function were sequentially added to the network for counting.

3.3.1. Transfer Learning

In this paper, a pretrained ResNet50 network was used to extract image features. ResNet is a deep learning network with many different versions, such as ResNet18, ResNet34, ResNet50, and ResNet101, which have different depths and sizes. The performance of a model improves as the depth of the network increases, but the model becomes more

complex and has a larger size. In order to select the most appropriate version of ResNet and to verify the effectiveness of transfer learning, different models were used to extract image features. The experimental results are shown in Table 3, where the MAE of the models were all >20 without using transfer learning, while the MAE decreased dramatically when transfer learning was used. In addition, the MAE of ResNet50 and ResNet101 were significantly smaller than those of ResNet18 and ResNet34 when transfer learning was used. Although the error of ResNet101 was slightly smaller than ResNet50, its size was much larger. Therefore, considering the performance and size of the models, the pretrained ResNet50 was used to extract image features.

Table 3. Model performance using different ResNets. A ✓ indicates that transfer learning is used.

Transfer Learning	Method	Val MAE	Val RMSE	Test MAE	Test RMSE	Size (MB)
	ResNet18	25.08	32.68	32.40	42.30	44.7
	ResNet34	27.77	34.14	37.56	46.01	83.3
	ResNet50	24.01	30.95	30.28	41.01	97.8
	ResNet101	24.43	34.56	27.61	39.41	171
✓	ResNet18	17.21	25.15	22.67	34.59	44.7
✓	ResNet34	15.67	19.80	20.79	26.56	83.3
✓	ResNet50	12.15	19.18	17.72	25.15	97.8
✓	ResNet101	12.05	17.75	16.63	25.87	171

3.3.2. Feature Convolution Kernel

In order to improve the relevance of the image features to the chickens, and to reduce the interference of the complex background, the ordinary convolution kernel was replaced with a feature convolution kernel. During the feature convolution generation process, features from multiple single chicken images were fused and multi-scale scaling was performed. The following experiments explored these factors:

(1) Feature Fusion. Although the counting objects in this paper were chickens of the same breed and age, the feature map of one image did not provide an accurate representation of all chickens, due to their different postures. A feature convolution kernel that fuses enough features is beneficial for improving the relevance of the image features to chickens during the counting process. To find the most suitable number of features, a series of experiments were executed. The experimental results are shown in Table 4. The results show that the performance of the model improved as the number of features increased. However, the model performance no longer improved after a certain number was reached. This is because extracting a large number of image features is conducive to enriching the feature information of the chickens, but excessive information will lead to redundant feature information. Finally, the features of twenty single chicken images were extracted for the subsequent experiments.

Table 4. Model performance that fuses different numbers of features.

Number	Val MAE	Val RMSE	Test MAE	Test RMSE
10	12.58	17.90	14.62	21.04
15	10.31	16.20	14.87	23.72
20	10.71	15.45	14.41	22.55
25	10.50	15.81	14.65	23.40
30	11.06	15.25	15.36	23.27

(2) Features Multi-scaling. A multi-scaling strategy was adopted to obtain rich target features by adding convolution kernels with different scales. The length and width of most chickens in the image were between 50 and 150 pixels. In the stage of feature convolution kernel generation, the feature maps could represent the features of single chickens with 100 × 100 pixels. Therefore, the scales were set from 0.5 to 1.5, to allow the model to

accurately count chickens with different scales. A series of experiments were implemented to find the most suitable strategy. The results are shown in Table 5. The magnification and reduction factors were set consistently. New scales were added to the combination of scales with the best result. This is conducive to counting chickens of different scales, through adding feature convolution kernels with multiple scales. But too much accumulation can produce redundant information, such that the performance cannot be improved. Finally, the best result was obtained by setting multiple scales of 0.8, 0.9, 1.0, 1.1, and 1.2.

Table 5. Model performance with different scale feature convolution kernels.

Scale	Val MAE	Val RMSE	Test MAE	Test RMSE
1.0	10.71	15.45	14.41	22.55
0.5, 1.0, 1.5	10.47	14.98	14.42	20.70
0.6, 1.0, 1.4	10.09	14.61	14.08	20.08
0.7, 1.0, 1.3	9.94	14.91	14.54	22.91
0.8, 1.0, 1.2	9.93	15.01	13.74	22.02
0.9, 1.0, 1.1	10.52	15.57	13.99	20.70
0.5, 0.8, 1.0, 1.2, 1.5	10.19	14.82	13.34	20.57
0.6, 0.8, 1.0, 1.2, 1.4	9.42	15.47	16.45	24.78
0.7, 0.8, 1.0, 1.2, 1.3	9.95	14.98	16.08	24.81
0.8, 0.9, 1.0, 1.1, 1.2	10.11	14.61	12.74	19.40
0.5, 0.8, 0.9, 1.0, 1.1, 1.2, 1.5	9.96	14.85	14.21	22.11
0.6, 0.8, 0.9, 1.0, 1.1, 1.2, 1.4	10.92	15.99	13.72	20.82
0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3	10.29	14.56	13.95	20.39

3.3.3. Feature Fusion

In addition to the feature fusion used due to the introduction of feature convolution kernels, feature fusion was used to concatenate features from the different layers during the correlation map generation stage. Extracting features using the fifth residual block of ResNet50 resulted in a significant increase in the number of model parameters; therefore, in order to reduce the computational consumption, the outputs of the third and fourth residual blocks were used for the experiments. The experimental results are shown in Table 6. It is obvious that concatenating the correlation maps of the third block and the fourth block obtained the best result. The richer the feature information obtained, the higher the quality of the density map generated using the fusion strategy.

Table 6. Model performance when extracting features from different blocks.

Block	Val MAE	Val RMSE	Test MAE	Test RMSE
3	15.58	22.45	21.13	30.71
4	10.11	14.61	12.74	19.40
3, 4	9.32	12.43	12.39	18.81

3.3.4. Improved Loss Function

The loss function in this article used additional MSE. Different values of the hyperparameter (λ) were set for the experiments. The experimental results are shown in Table 7, the best results were obtained when the value of λ was set to 1×10^{-9} . Compared to the original MSE, the addition of the extra MSE reduced the MAE by 0.29 for the validation set and 0.32 for the test set.

Table 7. Model performance for different values of λ .

λ	Val MAE	Val RMSE	Test MAE	Test RMSE
0	9.32	12.43	12.39	18.81
1×10^{-7}	9.43	12.81	12.64	18.46
1×10^{-8}	9.19	12.61	12.24	18.33
1×10^{-9}	9.03	12.09	12.07	18.20
1×10^{-10}	9.29	12.46	12.41	18.77

In the series of experiments above, various strategies were sequentially added to the network to perform the counting task, and the experiments verified their effectiveness. DFCCNet adopted all of these strategies and achieved accurate counting results, as shown in Figure 8.

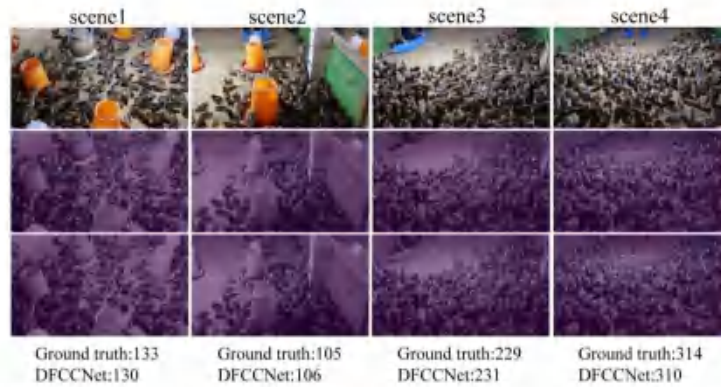


Figure 8. Visualization of counting results. The first line shows the original image, the second line shows the ground truth, and the last line shows the visualization output of DFCCNet. For better observation, the gray scale map was combined with the density map.

3.4. Robustness Testing

In order to verify the robustness of DFCCNet, the validation set and testing set were divided into different groups with different density levels, according to the number of targets. The detailed distribution of the data is shown in Figure 9. The data with three density levels were used as validation sets, to verify the robustness of the model. CSRNet [45] was compared as a baseline. The results are presented in Table 8. DFCCNet achieved smaller counting errors than CSRNet at the same level, and the NAE was less than 0.1 for all levels. Therefore, DFCCNet could maintain stable performance at different densities.

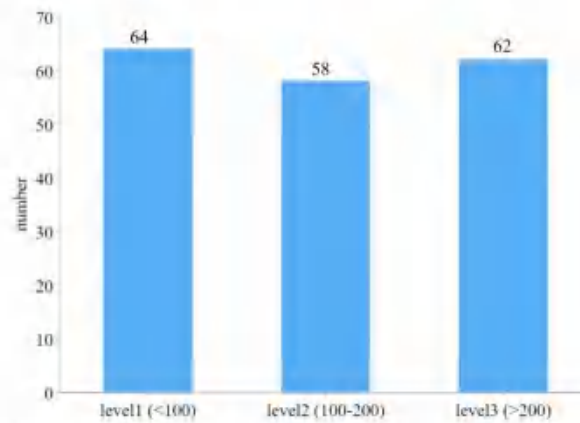


Figure 9. Distribution of data for different density levels.

Table 8. The performance of DFCCNet and CSRNet at different density levels.

Method	Density Level	MAE	NAE
CSRNet (baseline)	level1	7.03	0.121
	level2	10.10	0.073
	level3	21.46	0.077
DFCCNet	level1	4.26	0.074
	level2	9.85	0.071
	level3	19.17	0.068

3.5. Comparison with Other Methods

To further analyze the performance of DFCCNet, six state-of-the-art methods, including YOLOv5x [46], YOLOv7x [47], multi-column convolutional neural network (MCNN) [48], CSRNet, context aware network (CAN) [49], and segmentation guided attention networks (SGANet) [50] were used to compare with DFCCNet on the same training set, validation set, and testing set. The experimental results are shown in Table 9.

Table 9. Experimental results of chicken counting using different methods.

Method	Val MAE	Val RMSE	Test MAE	Test RMSE	Parameters (MB)	FPS
yolov5x	11.69	16.88	15.48	24.92	87.20	36.70
yolov7x	15.61	24.14	18.54	29.06	70.78	41.39
MCNN	12.83	17.71	17.04	22.92	0.13	56.68
CSRNet	10.15	13.78	14.20	21.10	16.26	8.43
CANNet	10.78	14.31	14.17	18.78	18.10	7.90
SGANet	9.77	13.07	14.77	22.62	21.79	5.90
Ours	9.03	12.09	12.07	18.20	17.89	16.15

First, DFCCNet was compared with two advanced object detection algorithms, YOLOv5x and YOLOv7x. However, it was difficult for these two methods to detect accurately individual objects in dense groups, due to occlusion phenomena. It is obvious that they had large MAE values in the experiment. Compared to YOLOv5x, the MAE of DFCCNet was reduced by 2.66 for the validation set and 3.41 for the test set. Although DFCCNet's FPS decreased by 20.55 compared to YOLOv5x, its number of parameters decreased by 69.31 MB. For the other density map regression methods, it can be seen that the parameters of MCNN were only 0.13 MB, and FPS reached 56.68. However, the three-column convolution kernels used by MCNN were too large in size. It could not extract features well for some small targets, resulting in a low accuracy. CSRNet, CANNet, and SGANet had a higher accuracy. But they are more complex, making them slow, and had a FPS of less than 10. Compared to CSRNet, DFCCNet showed a significant improvement in performance. The MAE was reduced by 1.12 on the validation set and 2.13 on the test set, and the FPS was improved by 7.72, while the parameters were increased by only 1.63 MB. Counting tasks in real scenery need to be implemented on mobile devices, which requires important considerations such as accuracy, size, and speed. In comparison, DFCCNet had the best accuracy, while maintaining acceptable parameters with 17.89 MB and FPS with 16.15, which can better meet the requirements of mobile devices. As shown in Figure 10, some renderings of DFCCNet, YOLOv5x, and CSRNet are presented. The first row shows some images were counted, and the last three rows list the counting results of YOLOv5x, CSRNet, and DFCCNet, respectively. It can be seen that YOLOv5x could obtain good counting result when the occlusion was not serious in scene1, but accurate counting was hindered in the case of serious occlusion in scene3 and scene4. In sparse scenes, density map regression, such as in DFCCNet and CSRNet, was not superior to object detection method, i.e., YOLOv5x. But they could achieve a higher counting accuracy than YOLOv5x

in dense scenes. It is obvious that the counting results of DFCCNet were better than those of the classical model, i.e., CSRNet.

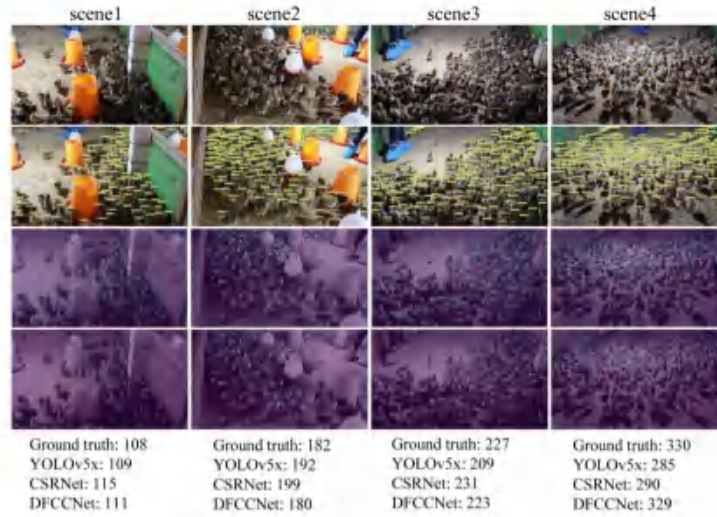


Figure 10. Comparison of yolov5x, CSRNet, and DFCCNet. For the same image, the result of DFCCNet was closer to the ground truth.

In addition, the values of MAE with the parameters and FPS respectively for the different models are shown in Figure 11. The methods with a density map have fewer parameters and smaller errors, while the object detection methods are faster. It can be seen that DCCNet is more suitable for the counting task for dense flocks of chickens.

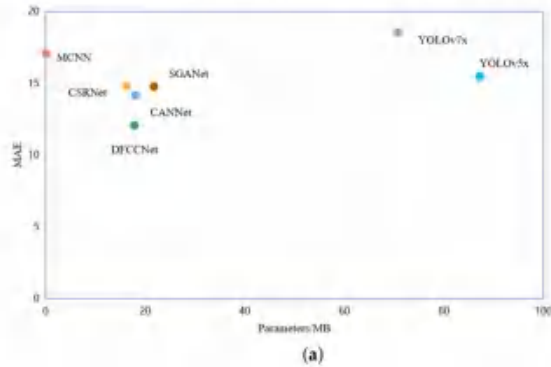


Figure 11. Cont.

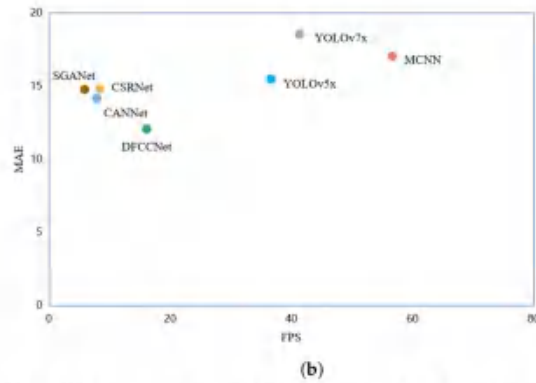


Figure 11. Comparing the test MAE, parameters, and FPS of the different methods. (a) The relationship between MAE and parameters. (b) The relationship between MAE and FPS.

3.6. Mobile Application

To enable DFCCNet to run on mobile devices and make the system easily accessible to end users, a mobile application (named Chickens-Counting) was developed. During the deployment of the model, the format of the model was transformed into Core ML (IOS), a format that can be used by mobile frameworks. In order to improve the performance of the model and reduce memory usage, quantization methods were used to compress the trained model. Sample screens of the app are shown in Figure 12. The app can ask the camera to take a picture or select a picture from the album, then the picture is passed to the API of the counting module. The API returns the time consumed by the counting process and the result of counting, and a visual image of the counting result is displayed in the screen. The app was tested on an iPhone 12 and the time required to count a single picture was 0.3 ± 0.05 s, which meets the counting requirements for mobile devices.

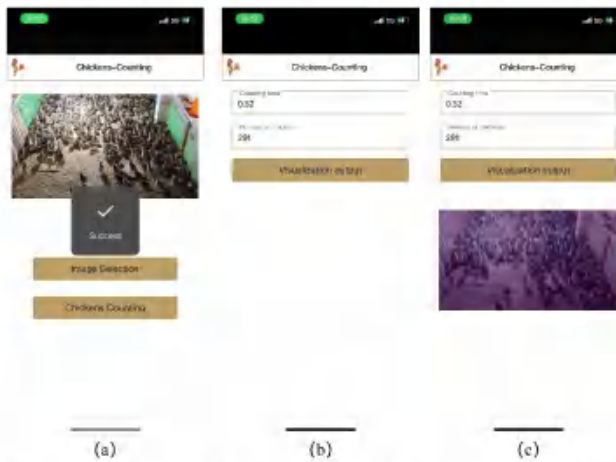


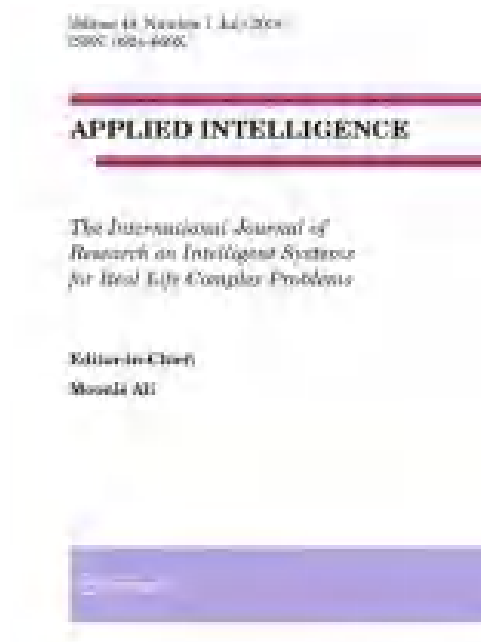
Figure 12. Chickens counting mobile app user interface. (a) Picture acquisition screen, (b) counting result, (c) visualization result.

11. Aydın, A. Development of an early detection system for lameness of broilers using computer vision. *Comput. Electron. Agric.* **2017**, *136*, 140–146. [[CrossRef](#)]
12. Tu, X.; Du, S.; Tang, L.; Xie, H.; Wood, B. A real-time automated system for monitoring individual feed intake and body weight of group housed turkeys. *Comput. Electron. Agric.* **2013**, *75*, 313–320. [[CrossRef](#)]
13. Zhuang, X.; Bi, M.; Guo, J.; Wu, S.; Zhang, T. Development of an early warning algorithm to detect sick broilers. *Comput. Electron. Agric.* **2018**, *144*, 102–113. [[CrossRef](#)]
14. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
15. Ren, Y.; Huang, Y.; Wang, Y.; Zhang, S.; Qu, H.; Ma, J.; Wang, L.; Li, L. A High-Performance Day-Age Classification and Detection Model for Chick Based on Attention Encoder and Convolutional Neural Network. *Animals* **2022**, *12*, 2425. [[CrossRef](#)] [[PubMed](#)]
16. Chen, B.L.; Kuo, Y.F. Early Warning System for Open-beaked Ratio, Spatial dispersion, and Movement of Chicken Using CNNs. In Proceedings of the 2023 ASABE Annual International Meeting, Omaha, NE, USA, 9–12 July 2023; p. 1. [[CrossRef](#)]
17. Yao, Y.; Yu, H.; Mu, J.; Li, J.; Fu, H. Estimation of the gender ratio of chickens based on computer vision: Dataset and exploration. *Entropy* **2020**, *22*, 779. [[CrossRef](#)]
18. Volkman, N.; Zelenka, C.; Devaraju, A.M.; Brünner, J.; Sittack, J.; Spindler, B.; Kemper, N.; Koch, R. Keypoint detection for injury identification during turkey husbandry using neural networks. *Sensors* **2022**, *22*, 5188. [[CrossRef](#)]
19. Yang, C.; Zhang, T.; Zheng, H.; Huang, J.; Cui, K. Pose estimation and behavior classification of broiler chickens based on deep neural networks. *Comput. Electron. Agric.* **2021**, *180*, 105863. [[CrossRef](#)]
20. Rong, J.; Zhou, H.; Zhang, F.; Yuan, T.; Wang, P. Tomato cluster detection and counting using improved YOLOv5 based on RGB-D fusion. *Comput. Electron. Agric.* **2023**, *207*, 107741. [[CrossRef](#)]
21. Tian, M.; Guo, H.; Chen, H.; Wang, Q.; Long, C.; Ma, Y. Automated pig counting using deep learning. *Comput. Electron. Agric.* **2019**, *163*, 104840. [[CrossRef](#)]
22. Zhang, L.; Li, W.; Liu, C.; Zhou, X.; Duan, Q. Automatic fish counting method using image density grading and local regression. *Comput. Electron. Agric.* **2020**, *179*, 105844. [[CrossRef](#)]
23. Abuasadah, D.; Switzer, A.; Bosa, M.; Liu, Y. Automatic counting of chickens in confined area using the LFCNN algorithm. In Proceedings of the IEEE International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 18–20 May 2022; pp. 1–7. [[CrossRef](#)]
24. Cao, L.; Xiao, Z.; Liao, X.; Yao, Y.; Wu, K.; Mu, J.; Li, J.; Fu, H. Automated chicken counting in surveillance camera environments based on the point supervision algorithm: LG-DenseFCN. *Agriculture* **2021**, *11*, 493. [[CrossRef](#)]
25. Li, X.; Zhao, Z.; Wu, J.; Huang, Y.; Wen, J.; Sun, S.; Xie, H.; Sun, J.; Cao, Y. Y-BGD: Broiler counting based on multi-object tracking. *Comput. Electron. Agric.* **2022**, *202*, 107547. [[CrossRef](#)]
26. Zhu, X.; Wu, C.; Yang, Y.; Yao, Y.; Wu, Y. Automated Chicken Counting Using YOLO-v5x Algorithm. In Proceedings of the 2022 8th International Conference on Systems and Informatics (ICSAI), Kunming, China, 10–12 December 2022; pp. 1–6. [[CrossRef](#)]
27. Zhao, Y.; Li, W.; Li, Y.; Qi, Y.; Li, Z.; Yao, J. LFCNet: A lightweight fish counting model based on density map regression. *Comput. Electron. Agric.* **2022**, *203*, 107496. [[CrossRef](#)]
28. Yu, X.; Wang, Y.; An, D.; Wu, Y. Counting method for cultured fishes based on multi-modules and attention mechanism. *Aquac. Eng.* **2022**, *96*, 102215. [[CrossRef](#)]
29. Ranjan, V.; Sharma, U.; Nguyen, T.; Hoai, M. Learning to count everything. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 3394–3403. [[CrossRef](#)]
30. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
31. Degu, M.Z.; Simen, G.L. Smartphone based detection and classification of poultry diseases from chicken fecal images using deep learning techniques. *Smart Agric. Technol.* **2023**, *4*, 100221. [[CrossRef](#)]
32. Sreeravulu, K.; Khan, H.A.S.; Damin, K.; Akhila, M.; Bharathi, G. Detection of Chicken Disease Based on Day-Age Using Pre Trained Model of CNN. In Proceedings of the International Conference on Mining Intelligence and Knowledge Exploration, Kristiansand, Norway, 28–30 June 2023; pp. 13–24. [[CrossRef](#)]
33. Wu, D.; Ying, Y.; Zhou, M.; Pan, J.; Cui, D. Improved ResNet-50 deep learning algorithm for identifying chicken gender. *Comput. Electron. Agric.* **2023**, *205*, 107622. [[CrossRef](#)]
34. Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Li, F.F. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255. [[CrossRef](#)]
35. Lu, L.; Li, H.; Ding, Z.; Guo, Q. An improved target detection method based on multiscale features fusion. *Microsc. Opt. Technol. Lett.* **2020**, *63*, 3051–3059. [[CrossRef](#)]
36. Yang, J.; Wu, C.; Du, B.; Zhang, L. Enhanced multiscale feature fusion network for HSI classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10328–10347. [[CrossRef](#)]
37. Ray, A.; Kolekar, M.H.; Balteubramanian, R.; Hafiane, A. Transfer learning enhanced vision-based human activity recognition: A decade-long analysis. *Int. J. Inf. Manag. Data Insights* **2023**, *3*, 100142. [[CrossRef](#)]
38. Gulzar, Y. Fruit image-classification model based on MobileNetV2 with deep transfer learning technique. *Sustainability* **2023**, *15*, 1906. [[CrossRef](#)]
39. Mamal, N.; Othman, M.F.; Abdalghafar, R.; Alwan, A.A.; Gulzar, Y. Enhancing image annotation technique of fruit classification using a deep learning approach. *Sustainability* **2023**, *15*, 901. [[CrossRef](#)]

40. Fan, X.; Luo, P.; Mu, Y.; Zhou, R.; Tjahjadi, T.; Ren, Y. Leaf image based plant disease identification using transfer learning and feature fusion. *Comput. Electron. Agric.* **2022**, *196*, 106892. [CrossRef]
41. Ouyang, H.; Zeng, J.; Leng, L. Inception Convolution and Feature Fusion for Person Search. *Sensors* **2023**, *23*, 1984. [CrossRef]
42. Ma, Z.; Wei, X.; Hong, X.; Gong, Y. Bayesian loss for crowd count estimation with point supervision. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6142–6151. [CrossRef]
43. Ranjan, V.; Le, H.; Hoai, M. Iterative crowd counting. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 270–285. [CrossRef]
44. Wang, Q.; Gao, J.; Lin, W.; Li, X. NWPU-crowd: A large-scale benchmark for crowd counting and localization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 2141–2149. [CrossRef]
45. Li, Y.; Zhang, X.; Chen, D. Csmnet: Dilated convolutional neural networks for understanding the highly congested scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1091–1100. [CrossRef]
46. Glenn, J. YOLOv5. Git Code. 2020. Available online: <https://github.com/ultralytics/yolov5> (accessed on 1 October 2022).
47. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, QC, Canada, 17–24 June 2023; pp. 7464–7475.
48. Zhang, Y.Y.; Zhou, D.S.; Chen, S.Q.; Gao, S.H.; Ma, Y. Single-Image Crowd Counting via Multi-Column Convolutional Neural Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Cvpr), Las Vegas, NV, USA, 27–30 June 2016; pp. 589–597. [CrossRef]
49. Liu, W.; Salzmann, M.; Fua, P. Context-aware crowd counting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5099–5108.
50. Wang, Q.; Breckon, T.P. Crowd counting via segmentation guided attention networks and curriculum loss. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 15233–15243. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

3.3. DDTCN: Decomposed dimension time-domain convolutional neural network along spatial dimensions for multiple long-term series forecasting





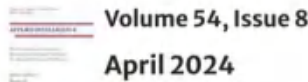
Applied Intelligence

Publishing model: Hybrid

South China Agricultural University Explore open access funding Change institution

Journal menu

[Search all Applied Intelligence articles](#) →

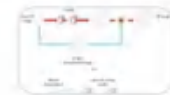


Volume 54, Issue 8

April 2024

49 articles in this issue

Scene text image super-resolution using multi-scale convolutional neural network with skip connections



OriginalPaper | 03 May 2024 | Pages: 5931 – 5943

Multiple remote sensing image encryption scheme based on saliency extraction and magic cube circular motion



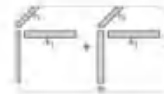
OriginalPaper | 03 May 2024 | Pages: 5944 – 5960

A conditioned feature reconstruction network for few-shot classification

OriginalPaper | 21 May 2024 | Pages: 6592 – 6605

DDTCN: Decomposed dimension time-domain convolutional neural network along spatial dimensions for multiple long-term series forecasting

OriginalPaper | 21 May 2024 | Pages: 6606 – 6623

Image classification based on tensor network DenseNet model

OriginalPaper | 21 May 2024 | Pages: 6624 – 6636

Analyzing the robustness of decentralized horizontal and vertical federated learning architectures in a non-IID scenario

OriginalPaper | Open access | 21 May 2024 | Pages: 6637 – 6653

Relationship constraint deep metric learning

OriginalPaper | 21 May 2024 | Pages: 6654 – 6666

Learning using granularity statistical invariants for classification

OriginalPaper | 22 May 2024 | Pages: 6667 – 6681

BNN-SAM: Improving generalization of binary object detector by Seeking Flat Minima

OriginalPaper | 22 May 2024 | Pages: 6682 – 6700



DDTCN: Decomposed dimension time-domain convolutional neural network along spatial dimensions for multiple long-term series forecasting

Kaihong Zheng¹ · Jinfeng Wang^{1,2} · Yunqiang Chen¹ · Rongjin Jiang³ · Wenzhong Wang⁴

Accepted: 10 May 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Time series analysis is widely applied in action recognition, anomaly detection, and weather forecasting. Time series forecasting remains a key challenge due to the complexity of temporal patterns, overlapping changes within sequences, and the need for advanced predictive models to forecast longer sequences in many scenarios. In this study, a model named the decomposed dimension time-domain convolutional neural network (DDTCN) is proposed. This model is specifically designed to address the challenges associated with long time series data. This paper presents a dimension temporal convolutional network (DTCN) module, which has a strong ability to capture variable correlations, and an adaptive strategy is introduced. Specifically, the model proposed in this paper first decomposes time series trends and is combined with the DTCN to extract variable correlations, thus achieving accurate predictions for complex time series and providing a powerful solution for long-term series forecasting. Experiments are conducted on multiple long-term series datasets covering five practical applications: energy, transportation, economics, weather, and healthcare. The proposed model is extensively evaluated and compared with traditional time series prediction methods and several benchmarks. The experimental results demonstrate that the proposed model outperforms state-of-the-art methods in most tasks involving multiple long-term series forecasting. Additionally, a pig price dataset is generated to predict agricultural economic trends, where compared to that of state-of-the-art methods, the DDTCN achieves a reduction in prediction error of 25.36%. Hence, this model holds promising prospects for wide-ranging applications.

Keywords Long-term series forecasting · Convolutional neural network · Adaptive modeling · Time trend decomposition

1 Introduction

Time series analysis serves as a valuable tool for uncovering underlying patterns in data, predicting future trends,

detecting anomalies, guiding decision-making and strategies, and enhancing operations in various domains. A time series refers to a sequence of data that occurs in chronological order. By using time series data, we can perform various tasks, such as classification, clustering, anomaly detection, and prediction. Time series forecasting holds significant importance and value in a multitude of fields, including energy [1], weather [2], transportation [3], economics [4], and health care [5].

When providing a formal definition of a time series prediction task, it is assumed that y_t represents specific time series data. For the time series prediction task, the main problem is how to use the observed values y_t from the time point $t - m$ to the time point t , i.e., $\mathcal{Y}_{t-m:t} = \{y_{t-m}, \dots, y_t\}$, to predict the value y_t from the time point $t + 1$ to time point $t + l$, $\mathcal{Y}_{t+l:t+l} = \{y_{t+l}, \dots, y_{t+l}\}$. As shown in Fig. 1, the processes of predicting a single variable with univariate data and multivariate data and predicting multiple variables

✉ Jinfeng Wang
wangjinfeng@scau.edu.cn

- ¹ College of Mathematics and Informatics, South China Agricultural University, Wushan, Guangzhou, Guangdong 510642, People's Republic of China
- ² Key Laboratory of Smart Agriculture of Guangzhou, Wushan, Guangzhou, Guangdong 510642, People's Republic of China
- ³ Department of Digital Process, Guangdong Wens Food Group Co., Ltd., Xixing, Yunfu, Guangdong 527300, People's Republic of China
- ⁴ College of Economics and Management, South China Agricultural University, Wushan, Guangzhou, Guangdong 510642, People's Republic of China

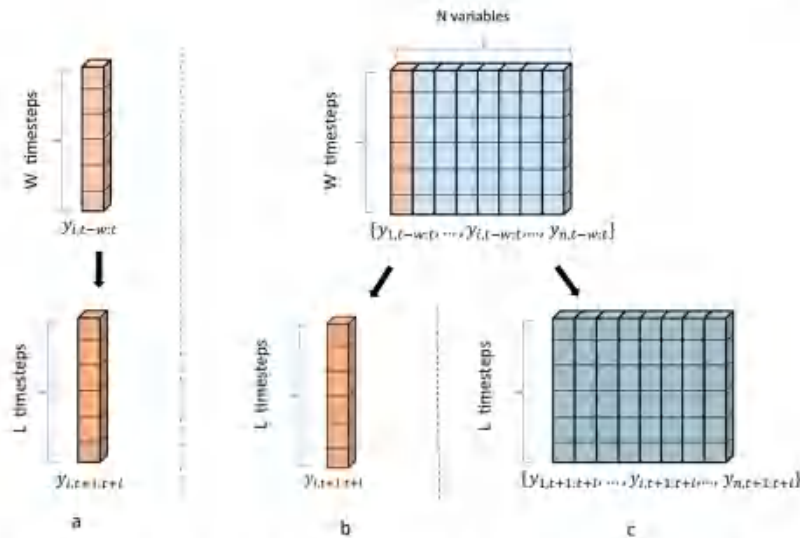


Fig. 1 Time series prediction task. **a**: From univariate time series to univariate time series; **b**: From multivariate time series to univariate time series; **c**: From multivariate time series to multivariate time series

with multivariate data can be summarized as follows:

$$y_{i,t+l:t+l} = f(y_{i,t-w:t}) \quad (1)$$

$$y_{i,t+l:t+l} = f(y_{1,t-w:t}, \dots, y_{i,t-w:t}, \dots, y_{N,t-w:t}) \quad (2)$$

$$\begin{aligned} & y_{1,t+l:t+l}, \dots, y_{i,t+l:t+l}, \dots, y_{N,t+l:t+l} \\ & = f(y_{1,t-w:t}, \dots, y_{i,t-w:t}, \dots, y_{N,t-w:t}) \end{aligned} \quad (3)$$

where \bar{i} represents the input channel, which is a time series variable, w is the length of the backtracking window, and l is the length that needs to be predicted.

Time series data refer to continuously record data at each time point that typically retain only a limited set of scalar values, distinguishing them from sequential data such as language or video data. Individual time points in time series data often lack semantic information to the extent that many studies primarily focus on temporal changes. The richness of information within these changes determines the extent to which intrinsic properties of time series, such as continuity, periodicity, and trends, can be effectively reflected. Changes in real-world time series data often involve complex temporal patterns, where multiple variations intertwine and overlap, rendering the modeling of temporal changes highly challenging.

In recent years, deep learning models have been rapidly developing, demonstrating excellent performance in both

prediction tasks and representation learning. The complex changes in real-world time series have led people to associate them with the powerful nonlinear modeling ability of deep learning models. Various deep learning models have been proposed to capture complicated fluctuations in real-world time series. Both long time series prediction and NLP involve modeling sequence data. Transformers [6] are considered the most successful solution for extracting semantic correlations between elements in long sequences. In recent years, Transformers, which have achieved great success in natural language processing, have received much attention in time series prediction. In natural language processing, sequences are usually words or characters in text, while in time series prediction, sequences are usually measurement data at time points. In time series analysis, a Transformer model uses an attention mechanism to capture pairwise time dependencies between time points. However, its position information processing ability is insufficient, making it difficult to capture fine-grained time dependencies between time points. Moreover, the computational complexity of the self-attention mechanism rapidly increases with increasing sequence length, which may lead to expensive training and reasoning on long sequences. To overcome these challenges, many researchers have proposed various variants of the Transformer [7–9]. These variants include embedding temporal information into models or using improved

attention mechanisms. Although using positional encoding and embedding subseries in the Transformer using tokens can help preserve some sorting information, the inherent invariance of the self-attention mechanism still leads to the loss of temporal information, making it difficult for the attention mechanism to directly obtain long-term dependencies from complex temporal patterns.

Before the Transformer model was developed, RNN-based recurrent network models or their variants were the first choice for time series prediction. These models are designed for processing sequence data, and with their recursive structure, the hidden state information is highly related to the current input and the information of the previous state, with stronger flexibility [10–13]. Another type of method utilizes convolutional neural networks along the time dimension [14–16]. However, in the time dimension, due to the locality of one-dimensional convolution, although models can effectively capture the relationship between adjacent time points, it is difficult to model long-term dependencies. The model in this article also adopts a one-dimensional convolution module. However, the model begins in the spatial dimension and improves the network architecture, which can effectively model variable correlations and temporal relationships.

The complexity of time patterns and overlapping changes within time series data poses significant challenges, especially when longer sequences need to be forecasted in various application scenarios. Long-term time series forecasting remains a key challenge in the field of time series analysis. In the realm of multivariate long-term time series forecasting, capturing prolonged dependencies within complex temporal patterns and efficiently capturing variable correlations along sequences are major challenges. To address these challenges, our model leverages two key methods and has achieved state-of-the-art performance in multivariate multistep forecasting tasks across various popular time series datasets. Our first idea, inspired by Autoformer, is to incorporate time series decomposition methods [17, 18] into the model. By decomposing the time series into seasonal and trend parts, global features and prolonged dependencies can be effectively extracted, reducing additional interference when capturing variable correlations in subsequent steps. Our second idea is to address local features based on a Temporal Convolutional Network (TCN), which has been applied to time series analysis [19, 20]. The convolution-based module proposed in this paper is effective at capturing local correlations and gathering information from each variable by expanding the receptive field. In contrast to traditional TCNs, the proposed model adopts a spatial perspective, which accomplishes trend decomposition and treats time as a channel to better capture the global attributes of time series. Thus, our model accounts for both global temporal properties and local features, offering a more explicit and precise

approach and outperforming many of the currently popular models in terms of both efficiency and accuracy.

The contributions of this study are as follows.

- (1) A convolution-based framework is constructed, and a time trend decomposition module is introduced to capture the global characteristics of the time series in a multivariate time series to achieve better performance.
- (2) The Dimension Temporal Convolutional Network (DTCN) module based on convolutions of spatial dimensions is proposed to capture correlations between variables.
- (3) An adaptive time series variable sensing strategy is proposed. This strategy is applied in the DTCN module. With this strategy, the DTCN can dynamically adjust the size of the convolution kernel or the dilation factor in the DTCN according to the variables of the time series to change the spatial receptive field of the DTCN, making the model more flexible and more generalized.
- (4) From the experiments conducted on seven benchmark time series datasets, significant improvements in the proposed method are observed compared to nine popular baseline models. In addition, a new real-world dataset named PPJice is constructed. This dataset is related to agriculture and the economy and is constructed to further demonstrate the good generalizability of the DDTCN model.

The remainder of this paper is organized as follows. Section 2 reviews related work. In Section 3, the problem definition and symbol notations are presented, followed by a description of the proposed approach. Section 4 presents an analysis of the experimental results and the model. Finally, in Section 5, conclusions and avenues for future work are offered.

2 Related work

Traditionally, statistical methods have been an effective means to forecast time series problem. However, these approaches often rely on numerous assumptions about the data, such as stationarity, linear correlation, normal distribution, and independence. When dealing with real-world time series data, complex and diverse scenarios are often encountered involving various unknown changes and trends. In recent years, Neural Network has gained prominence due to the nonlinearity, which can provide more robust and effective solutions for time series forecasting. Meanwhile, Transformer-based model can capture correlations within sequential data through the self-attention mechanism, offering superior global information interaction capabilities. Notably, some approaches have achieved satisfactory results due to incorporating time series decomposition techniques.

So, this article is inspired by these methods. Some related work will be described briefly as follows.

Traditional statistical method Statistical models assume that time changes follow a given pattern. For example, ARIMA [2], [22] solves a prediction problem by transforming a nonstationary process into a stationary process through differencing. The vector autoregressive (VAR) model [23–25] is mainly used for multivariate time series, and lag terms are used to represent the dependence between variables. Vector autoregressive moving average (VARMA) [26] uses the vector autoregressive (VAR) and vector moving average (VMA) models to capture the autoregressive structure and moving average structure of time series, respectively.

Neural network model With the advent of the big data era, various techniques and methods that demonstrate impressive potential for processing time series data have emerged. Machine learning, pattern recognition, intelligent computing, and data mining algorithms are continually emerging, offering an expanding toolkit for addressing the complexity of time series data. In particular, neural network models have proven to be more powerful and flexible than traditional statistical models in a variety of applications.

Models based on RNNs utilize a recurrent structure to implicitly capture temporal changes through state transitions between time steps [27, 28]. On the foundation of autoregressive models combined with recurrent networks, the DeepAR method has been proposed for probabilistic forecasting [29]. DeepAR is trained on a large number of related time series. Additionally, there are models that combine improved attention mechanisms with RNNs to explore the long-range dependencies of multivariate time series [30–32]. The adversarial truncated Cauchy self-attentive time variant neural network (ASATVN) and multiscale adaptive attention-based time-variant neural network (MATVN) incorporate novel advancements in attention mechanisms within the framework of time-variant neural network models [33, 34]. MLP-based models adopt MLPs along the time dimension, processing each time step of the time series data as part of the input to capture temporal dependencies, which are encoded as parameters of the MLP layer [35–37]. A method based on the TCN employs convolutional kernels sliding along the temporal dimension to capture temporal variations. Another approach to prediction utilizing convolutional neural network models involves the integration of additional spatial relational data for modeling. These approaches often require extra spatial relational data, employing various convolutional models, including multiconvolutional models [38] and graph convolutional network models [39]. Therefore, these approaches frequently require more extensive spatiotemporal analytics, which are particularly crucial in fields such as weather forecasting, environmental monitoring, and transportation management.

Transformer, as special neural network architectures, has achieved outstanding performance in time series forecasting. Initially, designed for NLP for sequence modeling, Transformers, with their attention mechanisms, can effectively capture dependencies between time points. Autoformer, which utilizes an autocorrelation mechanism, was proposed to capture sequential temporal dependencies based on learning periods. In addition, to address complex temporal patterns, Autoformer also utilizes a deep decomposition architecture to obtain the seasonal and trend components of the input series. Subsequently, FEDformer [40] incorporates an expert ensemble design to enhance seasonal and trend decomposition and presents sparse attention in the frequency domain. Crossformer [41] emphasizes variable correlations and highlights the utilization of cross-dimensional dependencies for multiple time series (MTS) prediction. In addition, Crossformer introduces a two-stage attention (TSA) layer to effectively capture dependencies between embedded segments.

The proposed model incorporates convolutional modules, with convolutional kernels sliding along the spatial variable dimension to capture variable correlations while considering all the temporal information as comprehensively as possible. Meanwhile, residual or skip connections are utilized to prevent the loss of crucial information and mitigate vanishing or exploding gradient problems during network propagation. Additionally, our work incorporates Time Series Decomposition technique introduced as follows.

Time series decomposition Time series decomposition is a standard method in time series analysis that decomposes a time series into multiple components, each representing one of the underlying categories of more predictable patterns. This method is primarily used to explore historical changes over a period of time and can highlight the inherent properties of a time series amidst complex temporal patterns [42]. For forecasting tasks, decomposition, such as Prophet [43] with trend and seasonality decomposition, N-BEATS based on MLP [44] with basis expansion, and DeepGLO [45] with matrix decomposition, is always used as a preprocessing step for historical sequences before predicting future sequences. The proposed model is inspired by the idea of decomposition to decompose time series into trend and seasonality components, thereby retaining the global characteristic.

3 Models and methods

A framework named the DDTCN is constructed as shown in Fig. 2, which mainly consists of two DTCN modules, the trend dimension temporal convolutional network (TDTCN) and the seasonal dimension temporal convolutional network (SDTCN). These modules are constructed to process the trend

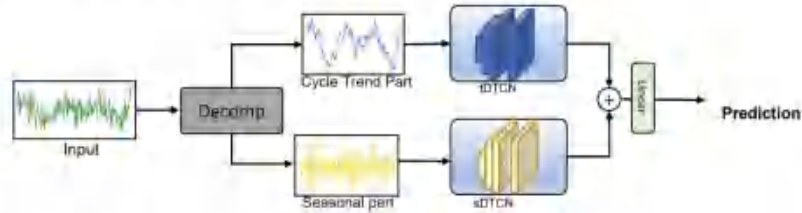


Fig. 2 The overall network architecture of the input and output of the model

part and the seasonal part of the time series information, respectively.

3.1 Time trend decomposition

To learn complex temporal patterns in the context of long-term forecasting, the proposed model adopts a decomposition approach. This method separates a sequence into its trend part and nontrend part. Given that the nontrend part carries a substantial amount of seasonal information, it is referred to as the seasonal part, as illustrated in Fig. 3. These two parts reflect the long-term progression and seasonality of the series. After the sequence decomposition is completed, these two parts will enter two dedicated DTCN modules, the tDTCN and sDTCN, to better preserve the global characteristics. For a time series with a time window length of w as the input of the model, the process can be presented as follows:

$$y_t = \text{AvgPool}(x) \quad (4)$$

$$y_s = x - y_t \quad (5)$$

where y_t and $y_s \in R^{w \times D}$ represent the seasonal part and the extracted trend period part, respectively. $\text{AvgPool}(x)$ refers to moving the average of a time series $x \in R^{w \times D}$ with a time window length of w and applying padding operations to maintain the same sequence length.

This paper uses y_s and $y_t = \text{Decomp}(x)$ to summarize the equation above.

3.2 DTCN

The convolutional neural network has achieved good results in many tasks, such as image classification, image segmentation, and speech recognition. As a convolutional architecture, the TCN model has been proven to play a good role in sequence modeling tasks and has considerable clarity and simplicity. The notable characteristics of the temporal convolutional network (TCN) model are as follows. The architecture employs causal convolutions, ensuring that there is no information leakage from the future to the past. The architecture can handle input sequences of arbitrary lengths and map them to an output sequence of the same length, akin to a recurrent neural network (RNN). The DDTCN builds upon these characteristics. However, what sets the dimension temporal convolutional network (DTCN) module in our model apart from traditional convolutional network models is that our DTCN is an adaptive convolution module along the spatial dimension. For instance, the traffic dataset describes road occupancy measurements from different sensors on highways in the San Francisco Bay Area, where each time series represents a sensor's record. These sensors have spatial relationships, which refer to the relationships between variables. Figure 4 illustrates the differences in sliding along the temporal and spatial dimensions. Within the graph, $\{y_1, \dots, y_t, \dots, y_w\}$ denotes data points at different time intervals for the same variable, while $\{t_1, \dots, t_s, \dots, t_N\}$ represents different variable values at the same time interval. The DTCN module operates from the spatial dimension,

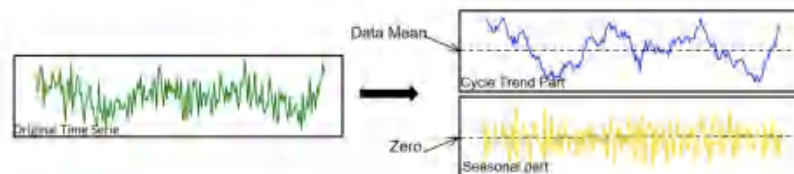


Fig. 3 Time trend decomposition module

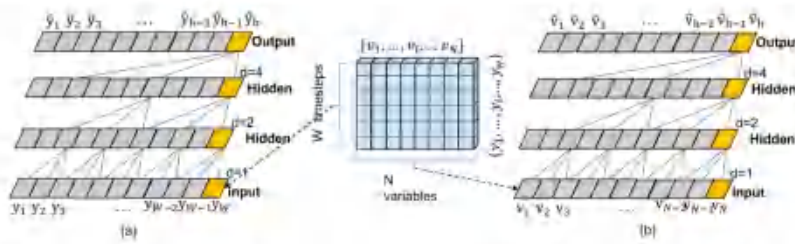


Fig. 4 The difference between TCN modules sliding along the temporal dimension and DTCN modules sliding along the spatial dimension when increasing the receptive field by changing the dilation factor. (a): TCN module sliding along the time dimension. (b): DTCN modules sliding along the spatial dimension

displaying a strong ability to capture correlations between variables, treating time series information as channel information, and sliding convolutional kernels along the spatial dimension. The combination of the DTCN module and our proposed network architecture enables our model to map to output sequences of arbitrary lengths.

As shown in Fig. 5, in the DTCN module, there are several convolutional layers, which are referred to as dimensional convolutional networks (DCNs) in this paper. Since the DTCN module focuses on the variables in the spatial dimension and the kernel slides between the variables, the number of DCNs should depend on the number of variables, that is, the size of the spatial dimension. The relationship between the DCN and spatial dimension is associated with how to ensure that the model has a receptive field along the spatial dimension, that is, how to ensure that there are filters that can reach all variables on the spatial dimension and that the deep network can be used to model data with large spatial dimensions.

To meet this condition, an adaptive DCN layer and two methods for increasing the spatial receptive field are designed. In the adaptive DCN layer, the number of DCN layers will be adjusted according to the spatial dimension of the data, that is, the number of variables. When the number of characteristic variables of the time series increases, the number of DCN layers should also increase. However, the spatial receptive field of the DTCN depends on the network depth n as well as the filter size k and dilation factor d , so the stability of increasingly deep DTCNs is important. The spatial receptive field of the DTCN is changed by adjusting the size of the convolution kernel (Fig. 5 (b)) or the dilation factor (Fig. 5 (c)) to ensure that the convolution kernel reaches each spatial variable of the data.

In the DCN, the receptive field of the n^{th} layer is:

$$RF_n = RF_{n-1} + (k - 1) * s \tag{6}$$

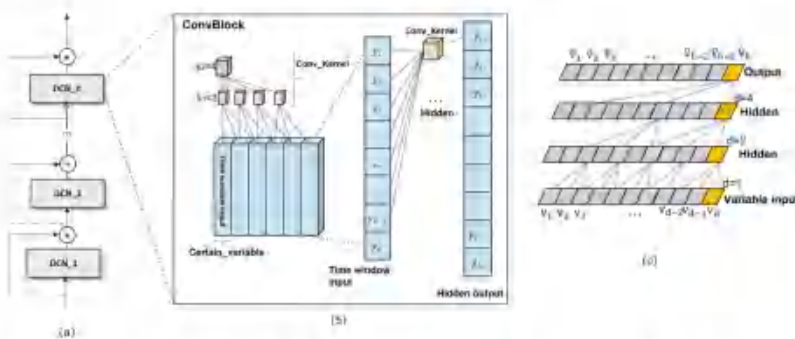


Fig. 5 DTCN module and its internal components: (a): DTCN residual block. (b): Internal mechanism of the DCN and an example of expanding the spatial receptive field by changing the convolutional kernel size k . (c): An example of expanding the spatial receptive field by changing the dilation factor d

where k is the size of the convolution kernel, RF_n is the receptive field of the n^{th} layer and s is the stride of the convolution kernel. As shown in Fig. 5 (b), when setting s to 1 and changing the size of the convolution kernel to make the DTCN module obtain a spatial receptive field matching the data, the following must be satisfied:

$$D \leq \sum_{i=1}^n (k_i - 1) + 1 \tag{7}$$

where D represents the spatial dimension of the data (the number of variables), k_i denotes the kernel size of the DCN, (the i^{th} DCN) layer, and n represents the number of DCN layers. In Fig. 5 (b), when $k_1 = 3$ and $k_2 = 4$, n is automatically determined to be 2 to satisfy the time series of 6 variables. Once the relationships between kernel sizes for each layer are determined, n can automatically adjust based on the size of the spatial dimension of the data.

For datasets with higher dimensionality, our solution is similar to some traditional TCN approaches. Dilated convolutions are used to achieve exponentially larger receptive fields. For a one-dimensional input sequence $x \in R^n$ and a filter $f: \{0, \dots, k-1\} \rightarrow R$, the dilated convolution operation F on an element x_i of the sequence is defined as follows:

$$F(x) = \sum_{j=0}^{k-1} f(j) * x_{i+d*j} \tag{8}$$

where d is the dilation factor, k is the filter size, and $s = d * j$ represents the past direction. Therefore, dilation is equivalent to introducing a fixed stride between every two adjacent filter taps. When d is equal to 1, the dilated convolution simplifies to a regular convolution. Using larger dilation values allows the output of the top layer to represent a wider range of inputs, effectively expanding the spatial receptive field of the DTCN.

When we vary the dilation factor to adapt the DTCN module, enabling it to acquire a spatial receptive field that matches the data, from a receptive field perspective

$$k' = d * (k - 1) + 1 \tag{9}$$

$$RF_n = RF_{n-1} + (k' - 1) * s \tag{10}$$

where k' represents the size of the expanded convolutional kernel, k represents the original convolutional kernel size, RF_n represents the receptive field of the n^{th} layer and s represents the stride of the convolution kernel. However, this receptive field may not always be fully effective. To ensure that every input related to essential variables is covered, a common practice in most dilated convolution-related works is followed, that is, d increases exponentially with the network depth (i.e., in the i^{th} layer of the network, $d = D(2^i)$). As shown in Fig. 5(c), when a filter size of 3 is determined,

it is necessary to satisfy the following condition:

$$D \leq 2^{n+1} - 1 \tag{11}$$

where n refers to the number of layers of the DCN. The DTCN also uses the above formula to automatically adjust the number of layers of the DCN, as shown in Fig. 5(c). The dilation factor is $d = 1, 2, 4$, the filter size is $k = 3$, the number of effective input variables is 1, and the stride of the convolution kernel sliding on the input sequence is determined to be 1. The output reaches every input in a valid variable. Changing the size of the padding at the same time ensures that the output length of each layer is the same.

$$D = (D + 2 * p - d * (k - 1) + 1) / s \pm 1 \tag{12}$$

where D is the spatial dimension of the data (number of variables), p is the padding size, d is the dilation factor, k is the convolution kernel size, and s is the sliding step of the convolution kernel. In this paper, an attempt is made to increase the convolution kernel to obtain a more powerful representation ability. If the computing resources are limited and the data variable is large, increasing the dilation factor is considered to expand the receptive field.

The DTCN module is composed of n DCN layers, where the value of n is automatically adjusted based on the data dimension D ,

$$y_i = DCN_1(y) \tag{13}$$

$$y_n = DCN_n(y_{n-1}) * y_{n-1} \tag{14}$$

where DCN refers to the convolution block convoluted along the spatial dimension and y_i represents the output of the DCN convolution block of the n^{th} layer. Additionally, the output of each DCN layer is added to the input of the next DCN block, as illustrated in Fig. 5 (a). This helps preserve the original features and information, allowing the network to better utilize and convey useful information, and has been repeatedly shown to be very useful for deeper networks.

The dDTCN and sDTCN modules in the proposed model differ from traditional TCN models. In the proposed model, the DTCN module takes the time series information window as the input channel for the convolution layers, with convolution kernels sliding in the variable direction. The dDTCN module processes the trend data, while the sDTCN module handles the seasonal data. The patterns and characteristics of the trend data and seasonal data exhibit significant differences. Although the dDTCN and sDTCN modules share similar designs, they handle different types of data, and their parameters are not shared. However, in the standalone DTCN module, the parameters of the convolution kernels are shared, allowing the module to continuously capture all the information in the time window. To address correlations between

variables, the spatial receptive field of the DTCN can be expanded by adjusting the kernel sizes or dilation factors and automatically adjusting the DCN layers based on the data dimensions, enabling the model to capture these relationships.

3.3 Overall input and output of the model

Figure 2 represents the overall process of the model from input to output. In simple terms, historical time window information is input into the model, and the Decomp module is used to decompose it into a seasonal part and a cycle trend part. These two parts are then separately processed through the sDTCN and tDTCN modules, followed by fusion. Finally, the data are mapped to the desired dimensions through the output layer. The process is as follows:

$$y_s, y_t = \text{Decomp}(x_{input}) \tag{15}$$

$$y_{s_output} = \text{sDTCN}(y_s) \tag{16}$$

$$y_{t_output} = \text{tDTCN}(y_t) \tag{17}$$

$$y_{predicted} = \text{MLP}(y_{s_output} + y_{t_output}) \tag{18}$$

where y_s and y_t represent the seasonal part and cycle trend part, respectively, of the historical time window information x_{input} after being decomposed by the Decomp module, y_{s_output} is the output of y_s after being processed by the sDTCN module, y_{t_output} is the output of y_t after being processed by the tDTCN module, and $y_{predicted}$ represents the model's final output value after passing through a linear layer.

4 Experiment

4.1 Data and metrics description

As shown in Table 1, eight datasets covering the fields of energy, transportation, economics, weather, health care, and agriculture are used in this study.

The Electric Power Transformer Temperature (ETT) dataset contains two years of power-related data collected from two sites in China, with two different sampling frequencies: 15-minute data (ETTm1) and hourly data (ETT1). Each data point contains six power load characteristics and target oil temperature values. The Electricity Load Electricity dataset contains the hourly electricity consumption of 321 customers from 2012 to 2014. The Exchange_rate dataset contains daily exchange rates for eight different countries from 1990 to 2016. The traffic dataset is a collection of hourly data from the California Department of Transportation and describes road occupancy measured by different sensors on freeways in the San Francisco Bay area. The Weather dataset includes Max Planck Institute climate data collected every

10 minutes in 2020. Each data point mainly contains 21 climatic features, such as temperature and air pressure. The ILL dataset includes data on flu-like illnesses (ILL) patients recorded weekly by the Centers for Disease Control and Prevention in the United States from 2002 to 2021, describing the ratio of ILL patients to the total number of patients. Pig Prices (PPrice) is a dataset collected in the Chinese market that contains data related to the prices of live pigs. The dataset covers the time span from June 2018 to June 2023, with a daily sampling frequency, and includes 45 data variables, encompassing prices for various categories of live pigs, feed prices, pork market prices, and price indices related to meat products.

All datasets are divided into training sets, validation sets, and testing sets with a ratio of 0.7:0.1:0.2, except for the ETT dataset, which is split at a ratio of 0.6:0.2:0.2. Four metrics are used for evaluating models, which include the mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). These metrics are chosen due to their wide use in the time series forecasting literature [34, 46, 47]. Both the MSE and RMSE measure the average difference between model predictions and actual observations. Therefore, the MSE and RMSE are used selectively based on actual situations. The MAE measures the average absolute difference between model predictions and actual observations. The MAPE measures the percentage prediction error of the model and is more sensitive to the relative error. The lower the values of these four metrics are, the better the model performance.

$$MSE = \frac{\sum (\text{actual} - \text{predicted})^2}{n} \tag{19}$$

$$RMSE = \sqrt{MSE} \tag{20}$$

$$MAE = \frac{\sum |\text{actual} - \text{predicted}|}{n} \tag{21}$$

$$MAPE = \frac{1}{n} \sum \frac{|\text{actual} - \text{predicted}|}{\text{actual}} \tag{22}$$

where *actual* represents the time series of real-world data and *predicted* denotes the time series predictions output by the model.

4.2 Implementation details

According to the adaptive strategy proposed in this article, when the size of each convolutional kernel k and other parameters are fixed, the number of DCN layers and the dilation factor d of each DCN layer in the DTCN are automatically adjusted based on the data. On the Electricity and Traffic datasets, this strategy is used in this paper, with a hidden layer $H = 256$, DCN convolution kernel size $k = 3$ for each layer, and stride of the convolution kernel $s = 1$. On the other datasets employ another adaptive strategy, which automatically adjusts the number of DCN layers and the size

Table 1 Dataset Description

DATASET	FREQUENCY	TIME SERIES	TIME STEPS	W_S	HORIZON
ETTh1	1 hour	7	17420	96	96,192,336,720
ETTm1	15 minutes	7	60680	96	96,192,336,720
Exchange_rate	1 day	8	7589	96	96,192,336,720
Electricity	1 hour	321	26304	96	96,192,336,720
Traffic	1 hour	892	17544	96	96,192,336,720
Weather	10 minutes	21	52696	96	96,192,336,720
ILJ	1 week	7	966	36	24,36,48,60
PPrice	1 day	45	1827	36	24,36,48,60

FREQUENCY is the data sampling frequency, TIME SERIES is the number of time series in the dataset, HORIZON is the number of variables, TIME STEPS is the total length of the time series in a dataset, W_S is the length of the time window, which is the length of the past time series for each input into the model, HORIZON is the predicted length, which is the length of the output time series of the model

k of convolution kernels in each layer of the DCN based on the data after fixing the dilation factor d and other parameters, is applied. In this strategy, the hidden layer $H = 128$ or $H = 256$, the expansion factor $d = 1$, and the stride of the convolution kernel $s = 1$. The Adam optimizer is used for training, with a learning rate of 0.0001. The batch size is set to 32. All the experiments are implemented in PyTorch and trained on an NVIDIA 4070 12 GB GPU. The code will be publicly available.

4.3 Performance in multivariate long time series forecasting

In the comparative experiments for long-term forecasting of multivariate time series, several models are considered for comparison. These models include enhanced transformer-based models, including Informer [48], Autoformer, and FEDformer, MLP-based models, including LightTS [37] and DLinear [50], the traditional Transformer, traditional LSTM, and the general time series analysis model TimesNet [49].

The self-attention distilling operation of the Informer highlights the dominance of attention by halving the cascading layer input and efficiently handles extremely long input sequences. In Autoformer, the autocorrelation mechanism replaces the self-attention mechanism, surpassing the previous self-attention series and improving both the computational efficiency and information utilization. At the same time, it also uses a time trend decomposition module internally. Thus, the model is capable of progressive decomposition. To better capture the global characteristics of time series, FEDformer uses a frequency-enhanced decomposition transformer architecture with a mixture of experts for seasonal trend decomposition. In the FEDformer structure, Fourier enhancement blocks and wavelet enhancement blocks replace cross-attention and self-attention blocks.

The MLP-based models include LightTS and DLinear. LightTS uses interval sampling and continuous sampling

methods to more effectively capture long-/short-term patterns and improve the efficiency of long input sequences. It is a lightweight model based on an MLP. DLinear is a combination of decomposition schemes used in the Autoformer and FEDformer and linear layers, outperforming many transformer-based models. Compared with the original paper, the Transformer and LSTM have basically not undergone any modifications. TimesNet, as a task-general foundation model, converts a one-dimensional time series into a two-dimensional one and captures its leapwise changes through a parameter-efficient admission block.

All experiments are performed on the same machine. The experimental results are recorded in Tables 2, 3 and 4, with the best prediction results highlighted in bold and the second-best results underscored. The tables summarize the evaluation results of the DDTCN and other methods on these eight real-world datasets. From the tables, the DDTCN achieves state-of-the-art performance in most cases. At the same time, the DTCN model performs much better than other models on the Exchange_rate and PPrice datasets. This is because in price or exchange rate data, which fluctuate within a certain range and have very small fluctuations, the convolution module can capture the correlation between variables and the regular pattern of change over time well while preserving global features. Therefore, it is believed that the DDTCN demonstrates outstanding predictive capabilities in a variety of domains, especially in areas related to economics and finance. On the ILJ dataset, due to the small amount of data and related spatial variables, the DDTCN does not have a strong ability to capture the correlation of variables and cannot achieve the optimal effect. However, this model achieves a suboptimal effect in approximately 60% of the tasks, and the error level remains stable as the prediction length changes. However, in the PPrice dataset, the amount of data and related spatial variables are relatively increased, and DDTCN can achieve the best performance under the same input window and prediction length.

Table 2 Multivariate long time series forecasting performance of various models on the ETT dataset

Data	Method	96			192			336			720		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ETTh1	DDTCN	0.394	0.614	8.536	0.425	0.654	8.426	0.451	0.687	9.416	0.490	0.705	12.511
	TimesNet(2023)	0.421	0.632	9.642	0.489	0.720	10.689	0.488	0.722	11.572	0.597	0.844	14.581
	DLinear(2023)	<u>0.395</u>	<u>0.619</u>	<u>8.759</u>	<u>0.442</u>	<u>0.669</u>	<u>8.457</u>	<u>0.453</u>	<u>0.691</u>	8.442	<u>0.510</u>	<u>0.717</u>	8.913
	LightTS(2022)	0.541	0.797	11.169	0.573	0.836	11.399	0.590	0.849	11.076	0.636	0.874	<u>10.730</u>
	FEDformer(2022)	0.424	0.627	11.146	0.433	0.686	11.146	0.487	0.713	12.740	0.517	0.730	<u>13.563</u>
	Autoformer(2021)	0.468	0.702	11.771	0.505	0.745	12.617	0.550	0.810	14.430	0.578	0.826	14.051
	Informers(2021)	0.826	1.063	11,792	0.791	1.021	13,627	0.818	1.046	15,303	0.884	1.103	19,849
	Transformer(2017)	0.811	1.044	18,871	0.698	0.919	12,478	0.743	0.969	9,918	0.819	1.021	19,877
	LSTM(1997)	0.772	1.013	15,595	0.777	1.018	15,404	0.774	1.013	15,729	0.780	1.014	15,277
ETTm1	DDTCN	0.374	0.581	2.081	0.389	0.612	<u>2.059</u>	0.409	0.636	<u>2.066</u>	0.443	0.678	<u>2.155</u>
	TimesNet(2023)	0.376	0.581	2.377	0.403	0.625	2.498	0.443	0.690	2.722	0.468	0.714	2.724
	DLinear(2023)	0.424	0.638	<u>2.22</u>	<u>0.395</u>	<u>0.618</u>	1.980	<u>0.413</u>	<u>0.643</u>	2.028	<u>0.455</u>	<u>0.689</u>	2.111
	LightTS(2022)	0.421	0.623	2.672	0.446	0.659	2.704	0.475	0.699	2.733	0.509	0.735	2.666
	FEDformer(2022)	0.417	0.629	2.183	0.435	0.656	2.214	0.457	0.681	2.199	0.499	0.729	2.326
	Autoformer(2021)	0.485	0.705	3.202	0.519	0.772	3.903	0.561	0.833	3.023	0.527	0.773	2.810
	Informers(2021)	0.585	0.842	2.883	0.634	0.878	3.625	0.709	0.968	3.193	0.844	1.129	3.938
	Transformer(2017)	0.527	0.730	2.528	0.568	0.776	2.477	0.639	0.843	2.668	0.763	0.962	1.448
	LSTM(1997)	0.726	0.942	2.961	0.730	0.949	2.900	0.742	0.961	2.819	0.755	0.975	2.672

Table 3 Multivariate long time series forecasting performance of various models on the Exchange_rate dataset (EX_r), Electricity dataset (ELE), Traffic dataset and Weather dataset

Data	Method	96			192			336			720		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
EX_r	DDTCN	0.204	0.287	1.057	0.303	0.410	1.699	0.407	0.534	2.355	0.617	0.823	2.736
	TimesNet(2023)	<u>0.242</u>	<u>0.338</u>	1.532	<u>0.344</u>	0.469	2.062	0.457	0.625	3.255	0.859	1.112	6.842
	DLinear(2023)	0.257	0.340	<u>1.373</u>	0.349	<u>0.455</u>	<u>1.829</u>	<u>0.444</u>	<u>0.571</u>	2.504	<u>0.647</u>	<u>0.833</u>	<u>2.824</u>
	LightTS(2022)	0.327	0.434	1.566	0.434	0.574	2.034	0.666	1.500	<u>2.185</u>	0.784	1.021	2.480
	FEDformer(2022)	0.291	0.398	1.872	0.389	0.523	2.902	0.490	0.664	3.459	0.857	1.117	7.198
	Autoformer(2021)	0.305	0.422	1.791	0.369	0.506	2.460	0.474	0.640	3.352	0.853	1.091	6.935
	Informers(2021)	0.680	0.838	2.499	0.757	0.918	4.082	0.890	1.062	4.804	1.140	1.408	2.396
	Transformer(2017)	0.753	0.906	3.426	0.842	1.040	3.274	0.985	1.193	4.118	1.375	1.654	2.911
	LSTM(1997)	1.295	1.534	2.246	1.293	1.532	2.348	1.396	1.672	2.563	1.335	1.586	<u>1.982</u>
ELE	DDTCN	0.276	0.434	2.695	0.282	<u>0.435</u>	<u>2.733</u>	0.299	0.455	2.643	0.332	0.488	2.897
	TimesNet(2023)	0.282	0.422	2.471	<u>0.289</u>	0.434	2.745	0.354	0.522	3.097	0.371	0.547	3.117
	DLinear(2023)	0.347	0.502	2.641	0.353	0.505	2.697	0.364	0.517	<u>2.648</u>	0.391	0.548	2.746
	LightTS(2022)	0.340	0.490	2.886	0.328	0.477	2.784	0.348	0.499	2.885	0.368	0.527	3.049
	FEDformer(2022)	0.327	0.462	3.095	0.342	0.481	3.235	0.359	0.502	3.365	0.384	0.534	3.708
	Autoformer(2021)	0.323	0.456	3.220	0.323	0.460	3.069	0.337	0.474	3.300	0.357	0.501	3.591
	Informers(2021)	0.429	0.595	4.146	0.443	0.604	4.160	0.458	0.620	4.589	0.472	0.646	4.198
	Transformer(2017)	0.430	0.603	3.703	0.424	0.592	4.107	0.473	0.635	5.867	0.464	0.631	5.601
	LSTM(1997)	0.437	0.609	3.424	0.416	0.581	3.516	0.408	0.570	3.749	0.408	0.574	3.453
Traffic	DDTCN	0.391	<u>0.782</u>	4.999	<u>0.380</u>	0.767	4.755	<u>0.385</u>	0.774	<u>4.612</u>	<u>0.409</u>	0.802	<u>4.540</u>
	TimesNet(2023)	<u>0.370</u>	0.811	4.126	0.336	0.790	3.566	0.425	0.872	5.474	0.447	0.914	5.312

Table 3 continued

Data	Method	96			192			336			720		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
weather	DLlinear(2023)	0.490	0.899	5.713	0.475	5.537	5.537	0.477	0.882	5.289	0.492	0.904	5.072
	LightTS(2022)	0.535	0.945	6.923	0.546	0.953	6.981	0.554	0.965	6.920	0.506	0.913	5.997
	FEDformer(2022)	0.379	0.770	4.356	0.587	0.781	4.715	0.395	0.800	4.534	0.412	0.820	4.474
	Autoformer(2021)	0.479	0.872	6.014	0.509	0.892	6.228	0.469	0.856	6.109	0.584	0.987	7.236
	Informer(2021)	0.630	0.896	20.444	0.785	1.060	19.972	0.884	1.188	21.162	1.022	1.347	14.101
	Transformer(2017)	0.465	0.915	5.938	0.438	0.887	4.803	0.381	0.839	3.989	0.387	0.837	4.038
	LSTM(1997)	0.891	0.943	5.845	0.539	0.986	6.186	0.531	0.984	6.055	0.482	0.880	5.360
	DDTCN	0.247	0.409	8.788	0.286	0.463	10.110	0.326	0.515	9.051	0.378	0.581	8.160
	TimesNet(2023)	0.220	0.416	10.337	0.266	0.475	11.394	0.306	0.535	12.870	0.356	0.601	11.794
	DLlinear(2023)	0.347	0.502	12.641	0.297	0.488	12.539	0.331	0.531	11.800	0.381	0.588	10.717
energy	LightTS(2022)	0.315	0.483	12.556	0.329	0.524	18.104	0.399	0.598	19.944	0.422	0.641	17.035
	FEDformer(2022)	0.243	0.438	13.680	0.281	0.482	13.050	0.317	0.527	12.069	0.366	0.583	10.369
	Autoformer(2021)	0.358	0.539	12.124	0.312	0.516	18.879	0.346	0.564	17.760	0.407	0.633	18.234
	Informer(2021)	0.630	0.896	20.444	0.785	1.06	19.972	0.884	1.188	21.162	1.022	1.347	14.101
	Transformer(2017)	0.431	0.605	14.329	0.579	0.774	13.883	0.537	0.733	12.818	0.612	0.830	22.541
	LSTM(1997)	0.323	0.506	12.253	0.347	0.535	12.679	0.370	0.567	12.510	0.402	0.608	10.083

4.4 Analysis of the model

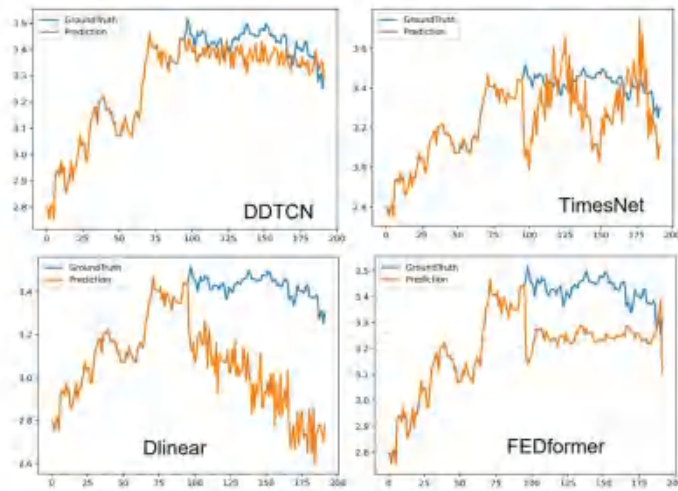
Figures 6, 7 and 8 show some prediction cases in which the DDTCN is compared with several models that perform well. The DDTCN can mine hidden regularity information from

inputs with complex time patterns because it can better capture variable correlation, obtain long-range dependencies in the time series of the same variable, and obtain useful information between variables. For example, in Figs. 6 and 8, it is obvious that when the inflection point of the input time

Table 4 Multivariate long time series forecasting performance of various models on the ILL dataset and PPrice dataset

Data	Method	24			36			48			60		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ILL	DDTCN	0.971	1.539	3.120	1.076	1.615	2.874	1.082	1.627	2.603	1.076	1.625	2.645
	TimesNet(2023)	0.947	1.706	4.319	0.877	1.353	3.080	0.966	1.520	4.249	0.902	1.440	3.110
	DLlinear(2023)	1.314	1.746	3.268	1.181	1.714	2.889	1.181	1.711	2.599	1.507	2.077	4.761
	LightTS(2022)	2.260	2.998	5.360	2.160	2.877	4.888	2.129	2.893	4.575	2.180	2.938	4.537
	FEDformer(2022)	1.014	1.526	3.33	1.084	1.641	3.173	1.108	1.706	2.947	1.071	1.648	2.878
	Autoformer(2021)	1.150	1.685	4.304	1.124	1.697	3.899	1.143	1.676	4.213	1.110	1.715	2.683
	Informer(2021)	1.601	2.314	2.163	1.396	2.190	2.164	4.657	2.158	1.997	1.523	2.218	2.052
	Transformer(2017)	1.400	2.120	1.829	1.504	2.245	1.798	1.392	2.343	2.109	1.551	2.278	1.628
LSTM(1997)	1.874	2.634	1.327	1.882	2.645	1.264	1.916	2.685	1.330	1.948	2.717	1.320	
PPrice	DDTCN	0.256	0.522	1.055	0.335	0.632	1.300	0.397	0.706	1.484	0.443	0.762	1.563
	TimesNet(2023)	0.243	0.620	1.613	0.390	0.722	1.815	0.245	0.298	2.068	0.452	0.836	2.045
	DLlinear(2023)	0.407	0.694	1.633	0.442	0.761	1.905	0.470	0.801	2.018	0.518	0.831	1.979
	LightTS(2022)	0.998	1.787	1.792	0.945	1.669	1.815	0.930	1.608	1.957	0.923	1.625	1.991
	FEDformer(2022)	0.396	0.716	1.758	0.437	0.778	1.999	0.466	0.817	2.109	0.495	0.853	2.185
	Autoformer(2021)	0.399	0.715	1.793	0.437	0.773	2.000	0.467	0.815	2.070	0.487	0.845	2.117
	Informer(2021)	1.601	2.314	2.163	1.496	2.190	2.164	4.657	2.158	1.997	1.523	2.218	2.052
	Transformer(2017)	0.932	1.404	3.332	0.982	1.479	3.514	1.023	1.534	3.679	1.045	1.567	3.688
LSTM(1997)	1.099	1.647	3.395	1.279	1.990	1.624	1.282	1.991	1.577	1.287	1.994	1.607	

Fig. 6 Prediction cases from the ETTh1 dataset under the input-96-predict-96 setting. The blue lines are the ground truths, and the orange lines are the model predictions. The first part, with a length of 96, is the input



series information is encountered in the steep rising trend, most other models will interfere with it and predict a downward trend. The DDTCN model can obtain complex variable information and time series information to ensure that the predicted values are closer to the real values.

Figure 9 shows a comparison of training and corresponding testing loss curves between the DDTCN and several other

models that performed well in terms of the metrics listed in the table above. Figure 9 shows that the DDTCN converges faster than the other models, and the testing loss remains at a lower level than that of the other models. As shown in Fig. 10, the DDTCN not only converges faster but also consumes less training time per epoch. Although the DDTCN is not the lightest in terms of the model parameter size, it

Fig. 7 Prediction cases from the ETTh1 dataset under the input-96-predict-96 setting. The blue lines are the ground truths, and the orange lines are the model predictions. The first part, with a length of 96, is the input

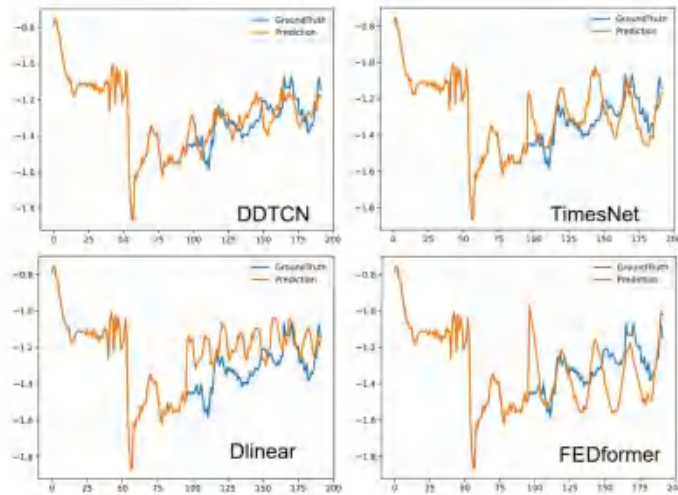
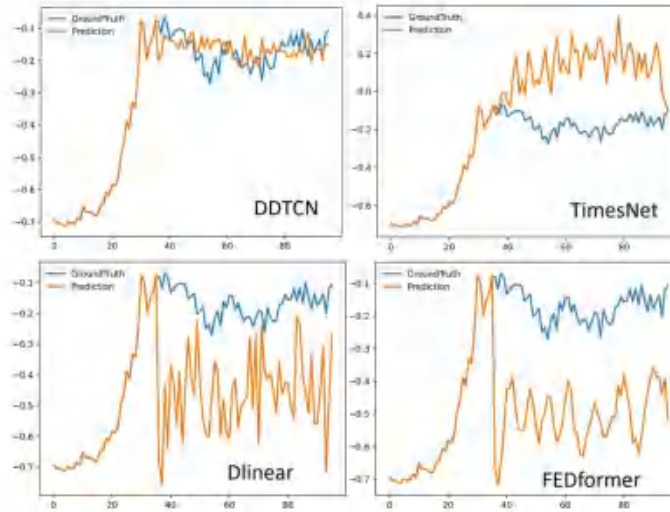


Fig. 8 Prediction cases from the PPrice dataset under the input-36-predict-60 setting. The blue lines are the ground truths, and the orange lines are the model predictions. The first part, with a length of 36, is the input



possesses greater flexibility to optimize the parameter size further according to requirements and the data.

4.5 Ablation studies

In this section, ablation studies are conducted, and the effectiveness of key components of the DDTCN is investigated.

First, the effectiveness of each module in the DDTCN model is explored. To explore this issue, we build three models based on the original DDTCN model. TCN_D is a model with a time trend decomposition module, but it is a convolution model along the time dimension. This model can

be regarded as a traditional TCN model with a trend seasonal decomposition module. A TCN is a convolution model along the time dimension without a decomposition module, that is, the traditional TCN model. The DTCN is a model without a time trend decomposition module, and the convolution parts slide along the spatial variable dimension.

In Fig. 11, the models utilizing spatial variable sliding convolution (DTCN and DDTCN) exhibit a maximum reduction of 90.16% in error levels when compared to those of network models employing temporal convolution (TCN_D and TCN). The minimum reduction in error is 11.90%. The proposed spatial variable sliding convolution model effectively captures more variable correlation information while

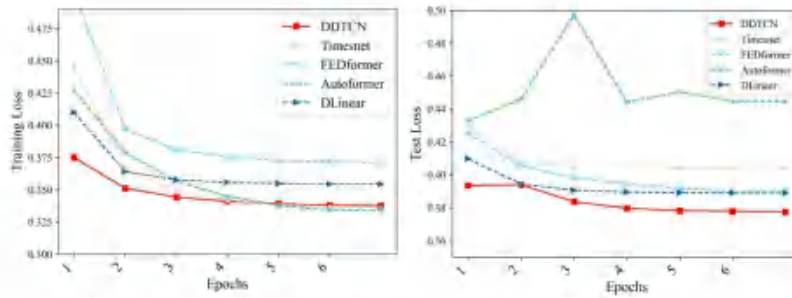


Fig. 9 The training loss of other models on the ETTh1 dataset with their corresponding testing loss curves

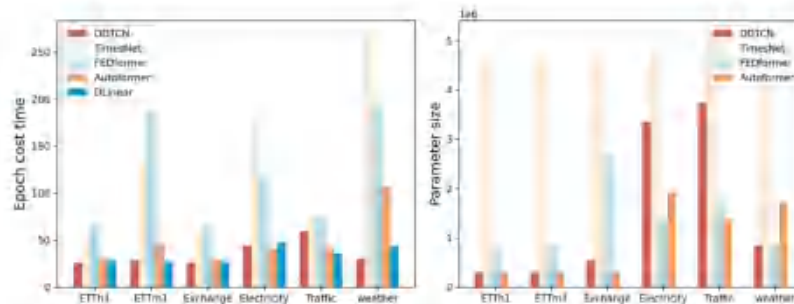


Fig. 10 Comparison of training time per epoch and model parameter size for each epoch. DLinear is a simple linear model based on an MLP, with a much smaller model parameter size than that of the other models, and it remains unchanged regardless of data variations

retaining temporal dependencies, leading to an improvement in the predictive accuracy.

Additionally, it can be observed that the TCN_D with the decomposition module exhibits higher error levels compared to those of the TCN in most cases. This is because the convolution kernels of the TCN model slide along the temporal dimension to capture temporal dependencies. However, the addition of the decomposition module introduces interference and leads to the loss of useful information. In contrast, the decomposition module in the spatial variable sliding convolution models provides global temporal features instead of introducing interference to the spatial convolution, thereby enhancing the model's prediction accuracy.

The variable correlation information obtained by the spatial variable convolution models is indeed impressive in terms of its use to improve the model performance. When examining the decomposition module in the spatial variable convolution models, models with the decomposition module are noteworthy. The DDTCN can achieve a maximum error reduction of up to 38.33% compared to that of the DTCN. However, in certain scenarios, the introduction of a decom-

position module may not lead to a significant improvement in the prediction accuracy. The impact of temporal dependency and global features on the prediction accuracy is relatively subtle compared to the effects of the spatial convolution capabilities of the DTCN module.

The cycle trend component data and the seasonal component data outputted by the decomposition module are processed separately by the tDTCN and sDTCN, respectively. The respective roles and importance of these modules are further explored. To investigate this issue, two models are constructed. The DDTCN-t model retains only the tDTCN module, while the DDTCN-s model retains only the sDTCN module. This also implies that the DDTCN-t model loses seasonal component data, while the DDTCN-s model loses cycle trend component data. A comparison of the prediction errors among the DDTCN-s, DDTCN-t, and DDTCN is shown in Table 5. The results show that the prediction error of the DDTCN-t is smaller than that of the DDTCN-s because the cycle trend data processed by the tDTCN better preserve the patterns and characteristics of the time series data over time, leading to a more accurate prediction. It is evident that

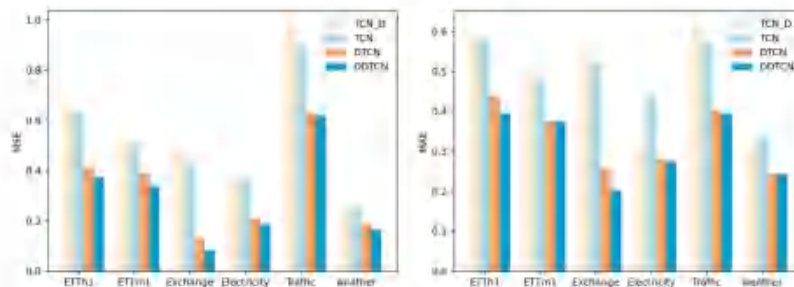


Fig. 11 Multivariate prediction results with a prediction length of 96 for the TCN_D, TCN, DTCN, and DDTCN models on several real datasets

Table 5 Comparison of the prediction errors among the DDTCN-s, DDTCN-t, and DDTCN

		ETTh1	ETTm1	Exchange_rate	Electricity	Traffic	Weather
MAE	DDTCN-s	0.691	0.688	1.414	0.440	0.454	0.597
	DDTCN-t	0.501	0.396	0.231	0.315	0.442	0.247
	DDTCN	0.392	0.374	0.204	0.277	0.403	0.247
MSE	DDTCN-s	0.790	0.779	2.869	0.376	0.703	0.617
	DDTCN-t	0.516	0.351	0.103	0.219	0.686	0.170
	DDTCN	0.377	0.338	0.082	0.186	0.632	0.168

The size of the input historical time window for the experiment is 96, and the prediction length is 96

the DTCN module plays a crucial role, contributing significantly more to the prediction accuracy compared to that of the sDTCN. Even on the Traffic dataset, retaining only the DTCN in the DDTCN-t results in a prediction error that is only 3.02% lower than that of the DDTCN. According to the characteristics of the dataset and the experimental results, as the number of time series variables and the data volume in the dataset increase, the ability of the DTCN module to capture time series correlations improves. Consequently, the gaps among the DDTCN-s, DDTCN-t, and DDTCN gradually diminish. On the ETTh1 dataset with fewer time series variables, the seasonal and trend data prove to be crucial for the prediction accuracy. Both the DTCN and sDTCN are indispensable, and if either of them is removed, the error will increase by at least 37.02%.

There are two approaches for changing the spatial receptive field in the DDTCN. The model that uses the approach of changing the size of the convolution kernel is named DDTCN-k, while another model named DDTCN-d uses dilated convolution. In the following section, the impact of these two methods on the model performance is explored. The parameter quantity in a model is typically employed to delineate the model complexity and the extent of space it occupies. Comparisons of the prediction errors and parameter quantities between the DDTCN-k and DDTCN-d are shown in Figs. 12 and 13. On the ETT dataset with fewer variables, the model parameter quantity of the DDTCN-k does not differ significantly from that of DDTCN-d, but there is a

noticeable difference in the prediction error levels. However, on the Electricity dataset with more variables, the parameter quantity of the DDTCN-k is 6.65 times greater than that of the DDTCN-d, while the prediction errors remain nearly the same. The influence of intervariable information on the model predictions is more significant because there are only 7 spatial variables in the ETT dataset. In contrast, the impact of intervariable information on the model predictions is relatively smaller for the Electricity dataset with 321 spatial variables. Additionally, there may be redundancy in the hidden information among variables with respect to their impact on predictions for datasets with a large number of variables. The model using dilated convolutions in the DTCN module can reduce this redundancy, while the model that expands the receptive field by changing the kernel size in the DTCN can adequately capture variable correlations. Therefore, the DDTCN-k is more suitable for datasets with fewer spatial variables, while the DDTCN-d is more suitable for datasets with a large number of spatial variables.

In summary, the DDTCN's strong predictive capabilities primarily stem from its ability to effectively capture spatial variable correlations and acquire information that is valuable for predictions. Decomposing time series into trends and seasonality has been proven to be effective; for the DDTCN, the ability of the decomposition module to capture global features is useful yet limited. The next steps may include focusing on enhancing the model's ability to capture long-term dependencies in time series from the perspective of the

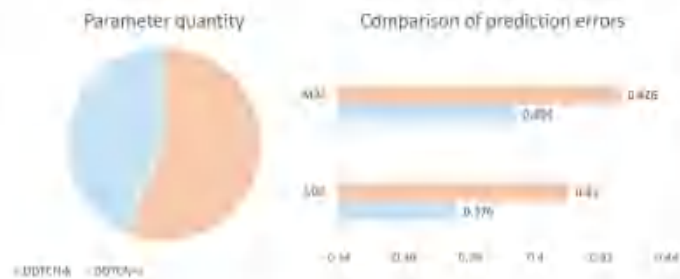
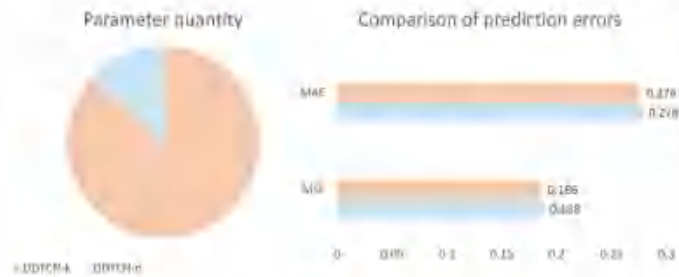
Fig. 12 Comparison of the prediction error levels and parameter quantities between the DDTCN-k and DDTCN-d on the ETTm1 dataset

Fig. 13 Comparison of the prediction error levels and parameter quantities between the DDTCN-k and DDTCN-d on the Electricity dataset



decomposition module, further improving the model's performance.

Nevertheless, when exploring the effect of time window size on the prediction accuracy, the DDTCN model shows the ability to capture the long-term dependence of time series. As shown in Fig. 14, although an increase in the time window will reduce the number of data batches input to the model in the dataset, the DDTCN can obtain more relevant time series information from a larger window when obtaining variable correlation information, including long-term dependence and global characteristics of the time series.

5 Conclusion and future work

A novel methodology named the DDTCN was proposed in this research. This convolution-based neural network model utilized a time trend decomposition module to preserve global features in time series data and captures variable correlations along spatial dimensions using convolutional neural

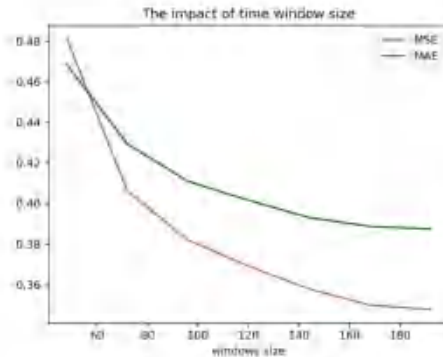


Fig. 14 Exploring the impact of the time window on the prediction errors on the ETT dataset when the prediction length is 192

networks. The DDTCN model had achieved significant performance improvements across various time series datasets in multiple domains and state-of-the-art results in most prediction tasks. In this study, a real-world dataset named PPrice was collected and organized, and experiments were conducted on multidimensional long time series forecasting using this dataset. The DDTCN model achieved the best performance in the experiments, demonstrating its remarkable generalizability. Furthermore, this study demonstrated through extensive experimentation the robust capability of the proposed DTCN module in capturing variable correlations and its contribution to the prediction accuracy. Through both experimentation and theory, it is evident that the DDTCN exhibited strong flexibility. This model can be adjusted based on the scale of time series data and the number of variables, thereby enhancing the model efficiency. Our work provided a powerful solution for multivariate long time series forecasting tasks.

While the DDTCN excels in extracting intervariable correlations, there is still room for improvement in the modules responsible for capturing global features and long-range dependencies. Future work on model enhancements could involve exploring alternative decomposition methods or employing deeper networks with a layer-by-layer decomposition approach. Considering the future applications of this model and given the DDTCN's powerful ability to capture spatial variable correlations, this model can be applied to real-world datasets with highly correlated multivariate variables, especially economic data.

Acknowledgements This work is supported in part by the Scientific Research Platforms and Projects of Guangdong Provincial Education Department (2021ZDZX1078; 2023ZDZX4002) and the Guangzhou Key Laboratory of Intelligent Agriculture (201902010081). The authors thank all the editors and reviewers for their suggestions and comments.

Author Contributions Conceptualization, Kaifeng Zheng and Jinfeng Wang; methodology, Kaifeng Zheng and Jinfeng Wang; software, Kaifeng Zheng; validation, Kaifeng Zheng and Jinfeng Wang; formal analysis, Jinfeng Wang and Wenzhong Wang; investigation, Kaifeng Zheng; resources, Jinfeng Wang and Rongjin Jiang and Wenzhong Wang; data curation, Kaifeng Zheng and Yunqiang Chen; writing—original draft preparation, Kaifeng Zheng; writing—review and editing,

Jinfeng Wang: visualization, Kaihong Zheng: Supervisor, Jinfeng Wang: project administration, Jinfeng Wang: funding acquisition, Jinfeng Wang: All authors have read and agreed to the published version of the manuscript.

Data availability and access: We use real-world datasets collected by [17, 48]. The other representative datasets and codes can be found at <https://github.com/1d3j/DDTCN>.

Declarations

Competing interests: The authors have no relevant financial or non-financial interests to disclose.

Ethical and informed consent for data used: All data used in this paper conform to the ethical and informed consent specifications.

References

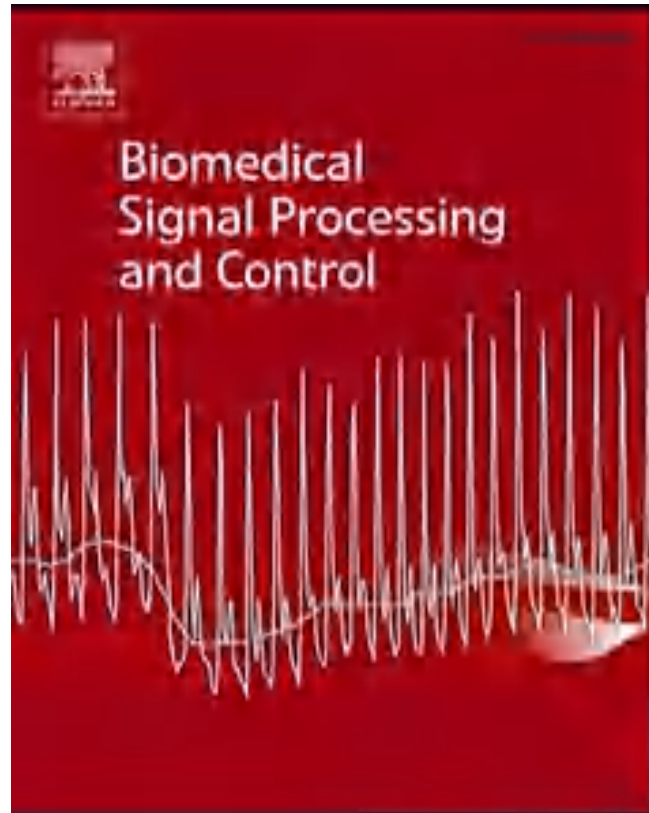
- Giaparis A, Luković S, Alippi C (2022) Deep learning for time series forecasting: The electric load case. *CAAI Trans Intell Technol* 7(1):1–25. <https://doi.org/10.1049/czt.12006>
- Peng Y, Gong D, Deng C, Li H, Cai H, Zhang H (2022) An automatic hyperparameter optimization deep model for precipitation prediction. *Appl Intell* 52(3):2703–2719. <https://doi.org/10.1007/s10489-021-02507-z>
- Xu C, Zhang A, Xu C, Chen Y (2022) Traffic speed prediction: spatiotemporal convolution network based on long-term, short-term and spatial features. *Appl Intell* 52(2):2224–2242. <https://doi.org/10.1007/s10489-021-02461-9>
- Sriyastava T, Mallick I, Bedi J (2024) Association mining based deep learning approach for financial time-series forecasting. *Appl Soft Comput* 111469. <https://doi.org/10.1016/j.asoc.2024.111469>
- Banerjee S, Lian Y (2022) Data driven covid-19 spread prediction based on mobility and mask mandate information. *Appl Intell* 52(2):1969–1978. <https://doi.org/10.1007/s10489-021-02381-4>
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2014) Attention is all you need. *Adv Neural Inf Process Syst* 30. <https://doi.org/10.48550/arXiv.1706.03762>
- Yang C, Wang Y, Yang B, Chen J (2024) Grotformer: A gated residual attention transformer for multivariate time series forecasting. *Neurocomputing* 127466. <https://doi.org/10.1016/j.neucom.2024.127466>
- Li Z, Cao Y, Xu H, Huang Y, He Q, Chen X, Tang X, Liu X (2024) Hidfomer: Hierarchical dual-sower transformer using multi-scale merge for long-term time series forecasting. *Expert Syst Appl* 239:122412. <https://doi.org/10.1016/j.eswa.2023.122412>
- Kitsyey N, Kaiser Ł, Leybaeva A (2020) Reformers: The efficient transformer. In: *8th International conference on learning representations*, ICLR 2020, Addis Ababa, Ethiopia, April 26–31, 2020
- Brockwerfer S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780. <https://doi.org/10.1162/nea-1997-08-1735>
- Sheng K, Chen Z, Liu Z, Song L, Zheng W, Yang B, Liu S, Yin L (2021) Haze prediction model using deep recurrent neural network. *Atmosphere* 12(12):1625. <https://doi.org/10.3390/atmos12121625>
- Lai G, Chang W-C, Yang Y, Liu H (2018) Modeling long-and short-term temporal patterns with deep neural networks, 95–104. <https://doi.org/10.1145/3209878.3210009>
- Shen L, Li Z, Kwok JT (2020) Timeseries anomaly detection using temporal hierarchical one-class network. In: Larochelle H, Razouvi M, Hadsell R, Balcan M, Lin H (eds.) *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*, December 6–12, 2020, Virtual
- Franceschi I, Dieuleveur A, Jaggi M (2019) Unsupervised scalable representation learning for multivariate time series. In: Wallach HM, Larochelle H, Beygelzimer A, d'Alelle-Buis F, Fox EB, Garnett R (eds.) *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019*, December 6–14, 2019, Vancouver, BC, Canada, pp 4652–4663. <https://doi.org/10.48550/arXiv.1901.10736>
- Zhang Z, Tian J, Huang W, Yin L, Zheng W, Liu S (2021) A haze prediction method based on one-dimensional convolutional neural network. *Atmosphere* 12(10):1327. <https://doi.org/10.3390/atmos12101327>
- He Y, Zhao L (2019) Temporal convolutional networks for anomaly detection in time series. *J Phys: Conference Series* 1213(A):042030. <https://doi.org/10.1088/1742-6596/1213/4/042030>
- Wu H, Xu J, Wang J, Long M (2023) Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In: Ranzato M, Beygelzimer A, Dauphin VN, Liang P, Vaughan JW (eds.) *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021*, December 6–14, 2021, Virtual, pp 22419–22430. <https://doi.org/10.48550/arXiv.2106.13008>
- Cleveland RB, Cleveland WS, McRae JL, Terpenning F (1990) Sit-A seasonal-trend decomposition. *J Off Stat* 6(1):5–73
- Yao M, Zhang R, Teng T, Meng J, Pihon E (2024) Recursive-performance prediction of automotive fuel cell based on conditional time series forecasting with convolutional neural network. In: *J Hydrogen Energy* 56:348–358. <https://doi.org/10.1016/j.jhydene.2023.12.168>
- Bai S, Kolter JZ, Kotturi V (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv:1803.01271*. <https://doi.org/10.48550/arXiv.1803.01271>
- Sensel H, Çakar Türk, Binnaz F, Gürgün E (2024) Improved methane production from ultrasonically-pretreated secondary sedimentation tank sludge and new model proposal: Time series (arima). *Bioresour Technol* 391:129866. <https://doi.org/10.1016/j.biortech.2023.129866>
- Hamilton JD (2020) *Time series analysis*. Princeton University Press
- Tsay RS (2014) *Multivariate time series analysis: With R and financial applications*. John Wiley & Sons, Inc., Hoboken
- Drachal K (2021) Forecasting crude oil real prices with averaging time-varying var models. *Resources Policy* 74:102244. <https://doi.org/10.1016/j.respol.2021.102244>
- Lütkepohl H (2006) Structural vector autoregressive analysis for cointegrated variables. *AdhenaesavStatistische Anzeiger* 90:75–88. <https://doi.org/10.1007/s10182-006-0022-4>
- Yang H, Pan Z, Tao Q, Qiu J (2018) Online learning for vector autoregressive moving-average time series prediction. *Neurocomputing* 315:9–17. <https://doi.org/10.1016/j.neucom.2018.04.011>
- Hertz JA (2018) *Introduction to the theory of neural computation*. Csu Press
- Shu W, Cai K, Xiang NN (2021) A short-term traffic flow prediction model based on an improved gate recurrent unit neural network. *IEEE Trans Intell Transportation Syst* 23(9):16654–16665. <https://doi.org/10.1109/ITITS.2021.9996659>
- Safinas D, Flunkert V, Garthaus J, Januschowski T (2020) Deepar: Probabilistic forecasting with autoregressive recurrent nei-

- works. *Int J Forecasting* 36(3):1161–1191. <https://doi.org/10.1016/j.ijforecast.2019.07.001>
30. Shih S-Y, Sun F-K, Lee H-y (2019) Temporal pattern attention for multivariate time series forecasting. *Machine Learn* 108:1421–1441. <https://doi.org/10.1007/s10994-019-05815-0>
 31. Song H, Rajan D, Thiagarajan L, Spatiaris A (2018) Attend and diagnose: Clinical time series analysis using attention models. In: *Proceedings of the AAAI conference on artificial intelligence*, vol. 32. <https://doi.org/10.1609/aaai.v32i1.31635>
 32. Qiu Y, Song D, Chen H, Cheng W, Jiang C, Cozzell GW (2017) A dual-stage attention-based recurrent neural network for time series prediction. In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp 2627–2633. <https://doi.org/10.24963/ijcai.2017/306>
 33. Changxia G, Ning Z, Youni L, Yan L, Hanyu W (2023) Multi-scale adaptive attention-based time-varying neural networks for multi-step time series forecasting. *Appl Intell* 1–20. <https://doi.org/10.1007/s10489-023-05057-7>
 34. Gao C, Zhang N, Li Y, Lin Y, Wan H (2023) Adversarial self-attention time-varying neural networks for multi-step time series forecasting. *Expert Syst Appl* 120722. <https://doi.org/10.1016/j.eswa.2023.120722>
 35. Chulita C, Oliveira KG, Oreshkin BN, Ramirez FI, Casero MM, Dubrawski A (2023) Nhts: Neural hierarchical interpolation for time series forecasting. *Proceedings of the AAAI conference on artificial intelligence* 37:6089–6097. <https://doi.org/10.1609/aaai.v37i6.25854>
 36. Zeng A, Chen M, Zhang L, Xu Q (2023) Are transformers effective for time series forecasting? *Proceedings of the AAAI conference on artificial intelligence* 37:11121–11128. <https://doi.org/10.1609/aaai.v37i9.26317>
 37. Zhang T, Zhang Y, Cao W, Bian J, Yi X, Zhong S, Li J (2022) Less is more: Fast multivariate time series forecasting with light sampling-oriented nlp structures. *arXiv:2207.01186*. <https://doi.org/10.48550/arXiv.2207.01186>
 38. Yin L, Wang L, Huang W, Liu S, Ying B, Zheng W (2021) Spatiotemporal analysis of haze in beijing based on the multi-convolution model. *Atmosphere* 12(11):1408. <https://doi.org/10.3390/atmos12111408>
 39. Lian J, Ren W, Li L, Zhou Y, Zhou B (2023) Psp-stgen: pedestrian trajectory prediction based on a spatio-temporal graph convolutional neural network. *Appl Intell* 53(7):2862–2878. <https://doi.org/10.1007/s10489-022-04524-1>
 40. Zhou T, Ma Z, Wu Q, Wang X, Sun L, Jin R (2022) Pedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In: *Chaunhuri K, Jegelka S, Song L, Szepesvári C, Niu D, Sabato S (eds.) International Conference on Machine Learning: IJML 2022, 17–23 July 2022, Baltimore, Maryland, USA. Proceedings of Machine Learning Research*, vol. 162, pp 27268–27280
 41. Zhang Y, Yan J (2023) Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In: *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1–5, 2023*
 42. Petropoulos F, Apiletti D, Assimakopoulos V, Babai MZ, Barrow DK, Ben Taieb S, Bergmeir C, Bessa RJ, Bjalk J, Boyan JE, Brownell J, Camesyale C, Casale JL, Cifra F, Clements MP, Condeiro C, Cyrino Oliveira FT, De Baets S, Dokumentov A, Elisson J, Frisöder P, Franses PH, Frazzer DT, Gilliland M, Gömlö MS, Gordon P, Gross L, Grushka-Cockayne Y, Guidolin M, Guidolin M, Günter U, Guo X, Gusev R, Harvey N, Hendry DF, Holtzman R, Januschowski T, Jeon J, José VRR, Kaufe Y, Koehler AB, Kralosa S, Kourentzes N, Levti S, Li F, Litsini K, Makridakis S, Marin GM, Martínez AB, Meeras S, Modis T, Nikolopoulos K, Onkal D, Pacagnini A, Panagiotelis A, Pampakidis I, Pavlou M, Peñín M, Pedregal DJ, Piram P, Ramos P, Rapach DE, Reade JJ, Roussis Tahar B, Rubaszek M, Sermpidis G, Shuang HL, Spiliotis E, Syntetos AA, Talagala PD, Talagala TS, Tashman L, Thomakos D, Thorarinnsson T, Tostini E, Trapero Arenas JR, Wang X, Winkler RL, Yusupova A, Ziel F (2023) Forecasting: theory and practice. *Int J Forecasting* 38(3):705–871. <https://doi.org/10.1016/j.ijforecast.2021.11.001>
 43. Taylor SJ, Letham B (2018) Forecasting at scale. *American Statistician* 72(1):37–45. <https://doi.org/10.1080/00031305.2017.1380990>
 44. Oreshkin BN, Carpov D, Chapalain N, Renjón Y (2020) N-BEATS: neural basis expansion analysis for interpretable time series forecasting. In: *8th International conference on learning representations, ICLR 2020, Addis Ababa, Ethiopia, April 26–30, 2020*
 45. Seo R, Yu H-R, Ohlim IS (2019) Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting. In: *Wallach H, Larochelle H, Beygelzimer A, Alché-Buc F, Fox E, Garnel R (eds.) Advances in neural information processing systems*, vol. 32
 46. Chen J, Yuan C, Dong S, Feng J, Wang H (2023) A novel spatiotemporal multigraph convolutional network for air pollution prediction. *Appl Intell* 1–14. <https://doi.org/10.1007/s10489-022-04418-y>
 47. García-Duarte L, Cifuentes J, Marillado G (2023) Short-term spatio-temporal forecasting of air temperatures using deep graph convolutional neural networks. *Stochastic Environ Res Risk Assessment* 37(5):1649–1667. <https://doi.org/10.1007/s00477-022-02358-0>
 48. Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiang H, Zhang W (2021) Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI conference on artificial intelligence* 35:11106–11115. <https://doi.org/10.1609/aaai.v35i12.17325>
 49. Wu H, Hu T, Liu Y, Zhou H, Wang J, Long M (2023) Timenet: Temporal 2d-variation modeling for general time series analysis. In: *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1–5, 2023*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

3.4. A simple self-supervised learning framework with patch-based data augmentation in diagnosis of Alzheimer's disease



Biomedical Signal Processing and Control | Supports open access | Submit your article »

Articles & Issues ▾ About ▾ Publish ▾ Order journal > 🔍 Search in this journal | Guide for authors

Actions for selected articles
Select all / Deselect all

- Download PDFs
- Email Notifications
- Show all article previews

Research article • Full text access
A simple self-supervised learning framework with patch-based data augmentation in diagnosis of Alzheimer's disease
Haoqiang Gong, Zhiwen Wang, Shuaihui Huang, Jinfeng Wang
Article 106572
[View PDF](#) [Article preview ▾](#)

Research article • Full text access
Denoising of 3D Magnetic resonance images based on balanced low-rank tensor and nonlocal self-similarity
Xiaotang Liu, Jingfei He, Peng Gao, Boudi Abdelmounim, Fan Lam
Article 106588
[View PDF](#) [Article preview ▾](#)

Research article • Full text access
Cascade dual-domain swin-conv-unet for MRI reconstruction
Jinhua Sheng, Xiaofan Yang, Qiao Zhang, Pu Huang, ... Haodi Zhu
Article 106623
[View PDF](#) [Article preview ▾](#)



A simple self-supervised learning framework with patch-based data augmentation in diagnosis of Alzheimer's disease

Haoqiang Gong, Zhiwen Wang, Shuaihui Huang, Jinfeng Wang*

College of Mathematics and Informatics, Anhui Chong Agricultural University, Guangshan, 310944, Guangdong, China

ARTICLE INFO

Keywords:
Alzheimer's disease
Self-supervised learning
MRI brain patch
Attention mechanism
Data augmentation

ABSTRACT

Alzheimer's disease (AD) stands as a prominent age-related disorder with significant global impact. Utilizing computer-aided diagnosis aids in the timely identification of mild cognitive impairment, facilitating early intervention and management. Self-supervised learning models have attracted much attention due to their advantages of no manual labeling, so they are very suitable for solving the problems of difficulty in data acquisition and high cost of manual labeling in medical image processing. However, the existing self-supervised algorithms applied to medical images often have poor diagnostic effects and consume large resources, as a consequence, the model's ability to learn meaningful representations from medical images is hampered, leading to suboptimal performance. Subsequently, diverse patches extracted from the same brain are amalgamated to create contrast views, and attention weights are then employed to enhance the fitting and generalization capacity of the model. Experimental results on the ADNI dataset with 1365 subjects show that PD-SIM has been improved in the diagnosis of different diseases (such as the classification ACC of AD and CM reached 0.797, and the classification ACC of early cognitive impairment reached 0.7036), and also alleviates the problem of large consumption of computer resources, downstream tasks require only 52ms per image. The proposed method performs well in atrophic structure identification and AD diagnosis. Therefore, PD-SIM has a wide range of application prospects and is of great practical significance within the realm of medical image analysis. The data and code can be found at <http://github.com/11306295575>.

1. Introduction

A common neurodegenerative disease, Alzheimer's Disease (AD), is characterized by phenomenon such as memory impairment and decline in imagination, and alter daily behavior [1]. Alzheimer's disease, which is caused by structural changes in the brain and is the cause of most of all dementia, is not easy to diagnose in its early stages and is clinically characterized by cognitive impairment. Based on the severity of the disease, it is primarily categorized into three main stages: normal cognition (NC), mild cognitive impairment (MCI), and Alzheimer's Disease (AD). MCI can be further subdivided into early mild cognitive impairment (EMCI) and late mild cognitive impairment (LMCI). The present diagnostic approach predominantly depends on the ability of doctors and participant intelligence examination, leading to a notable rate of misdiagnosis and subjective. Therefore, it is of practical significance to explore a simple and accurate clinical auxiliary diagnosis model for the detection of AD.

In recent years, the application of machine learning technology in the field of medical data processing has attracted much attention [2–4]. In fact, machine learning has demonstrated its ability to learn

underlying features from medical data and has significantly improved the diagnostic performance of diseases in many research efforts [5–11]. For example, in the final stage of disease diagnosis, recurrent neural network Long short-term memory (RNN-LSTM) architecture is used to create an AI-based classification model, which improves the classification performance [12]. However, because annotation of medical image datasets requires specialized knowledge, obtaining large amounts of annotated data is a daunting task. Some researchers have tried to solve this problem using unsupervised or semi-supervised learning methods. Among them, autoencoders and Generative adversarial networks (GAN) are representatives. The former can be used for feature extraction and dimensionality reduction [13], and the latter can be used to increase the diversity of data and synthesize new images [14]. In addition to using different learning methods, some researchers have proposed different network structures and loss functions to further improve the classification performance. For example, some researchers have proposed spatial or channel-attention approaches to better detect regions associated with disease [15].

* Corresponding author.
E-mail address: hwgong@ahcau.edu.cn (J. Wang).

<https://doi.org/10.1016/j.bspc.2024.100972>

Received 30 July 2024; Received in revised form 13 May 2024; Accepted 23 June 2024.

1746-8094/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

In addition, multi-task learning methods are used to combine multiple model results (such as disease diagnosis results and target segmentation results) for enhancing the accuracy and generalization ability of the overall framework [19]. Recently, SIMCLR (Simple Framework for Contrastive Learning of Representations) model has shown exciting performance in natural image classification [20], which is a representative learning-based model. A feature representation can be learned by maximizing the similarity between different transformations of the same image. This approach has been proven to be very effective in natural image classification tasks. Some researchers applied this method to medical tasks [18, 21], and achieved some success. Although SIMCLR model performs well in natural image classification tasks and has been applied in the field of medical image processing, it still has some shortcomings for medical diagnosis. One of the biggest issues is the limitation of data dimensionality and data augmentation. The existing self-supervised models take the whole graph as the object and perform data augmentation operations on it, which not only makes the model time-consuming, but also consumes a lot of computing resources.

Structural MRIs, typically grayscale images, lack diversity, contain small areas associated with disease pathology, and random cropping can cause disease-causing areas to disappear. Therefore, the application of self-supervised method in medical images has the following problems:

(1) Medical images (e.g., $256 \times 256 \times 256$) MRI usually have higher dimensions than natural images (224×224 , ImageNet [22]). SIMCLR models increase the diversity of data through data amplification, but at the same time increase the training cost and time of the model.

(2) Applied to natural images, these transformations may alter the anatomical structure or diseased areas of medical images, which may cause the model to learn incorrect feature representations. Therefore, data augmentation in medical image processing needs to keep the anatomical structure and lesion area of the image unchanged.

To deal with the problems mentioned above, a self-supervised learning framework Patch-based data augmentation (PD-SIM) is proposed, which can drastically reduce data dimensionality and generate appropriate data-augmented representations, and detect areas of brain structure that are abnormally shrunken. PD-SIM has three main modules, i.e. the data augmentation module, the multi-patch spatial attention combination mechanism, and the classical resnet18 feature extraction framework [23]. The proposed method is validated on the following two datasets: the ADNI public dataset and the Cingris [24] dataset, in certain real-world clinical diagnostic tasks (e.g., AD VS. NCI). PD-SIM method outperforms several self-supervised methods in multiple metrics and generalization performance. This paper makes the following contributions:

(1) A method for screening critical brains is proposed to detect focal microstructures from brain scans and group them into global features. It can reduce the size of input images and computational resources (e.g., reduce the number of features from 16,777,216 to 250,047), speed up diagnosis, and improve performance.

(2) A patch-based data augmentation method is proposed to generate two contrasting views from multiple key local regions which help the representation of the entire brain structure. It can avoid the loss of key brain region feature information caused by random clipping, stretching, blurring and other operations.

(3) The spatial attention mechanism is applied to the contrast view to make the lesion area focused better, so as to improve the image representation learning ability of the whole model.

The rest of the article is structured as follows: Section 2 describes the previous research. Section 3 introduces the involved materials and the proposed model. The experimental settings and comparison results of different AD classification tasks are given in Section 4. Section 5 discusses the effectiveness of multi-patch data augmentation approach identifying pathological locations and limitations. In the final section, a conclusion and future work are provided.

2. Related work

2.1. Diagnose early AD using MRI

Many researchers have made the diagnosis of AD efficient more and more and explored different types of biomarkers in different imaging modalities which include structural Magnetic Resonance Imaging (sMRI) [25], Functional Magnetic Resonance Imaging (fMRI) [26], Fluorodeoxyglucose Positron Emission Tomography (FDG-PET) [27], or Diffusion Tensor Imaging (DTI) [28]. sMRI does not use any radiation, so there is no risk to the human body. sMRI plays an important role in the medical diagnosis of AD. It can provide detailed information about the structure and morphology of the brain, including brain volume, cortical thickness, brain atrophy [29], etc. These structural changes are often present in patients with AD and are associated with disease development and clinical presentation. Common indicators include hippocampus volume and whole brain volume. The hippocampus is one of the early atrophic areas of AD, and its reduced volume is associated with the development of AD and cognitive decline.

MRI can provide very detailed images which are capable of showing fine structures of the nervous system, brain and joints [30]. MRI uses magnetic resonance phenomenon to image human tissues, which can provide high definition, multi-directional and multi-planar images, and can distinguish and quantitatively analyze different tissue types [31]. Hence, it serves as a crucial component of the standard diagnostic process for Alzheimer's Disease (AD) in clinical practice.

According to the division of the ROI of sMRI scans, previous studies on the diagnosis of brain lesions can be categorized into several levels: voxel-level, region of interest (ROI) level and patch level.

2.1.1. Image Voxel-level method

Voxel-level methods aim to distinguish subtle structures relevant to the disease within the sMRI scans of subjects' brains. The voxel-wise approach typically assumes tissue (e.g., gray and white matter) density as a feature for classifier algorithms [32–34]. Indeed, relying solely on voxel-level features neglects the strong inter-voxel correlations. Moreover, this voxel-level approach is prone to overfitting due to the high-dimensional feature representation compared to the limited training samples available.

Therefore, reducing feature dimensionality presents a notable challenge for voxel-level image methods seeking to improve AD classification performance. In Ref. [35], Voxel-based Morphometry (VBM) is used to convert MRI images into 3D image voxels and quantitatively analyze gray and white matter content of each Voxel. Brain structural changes in normal aging and Alzheimer's disease patients were compared. In Ref. [36], a method was proposed based on improved learning which can efficiently reduce the dimensionality of the subject images by removing high-frequency components from voxel-wise cortical thickness data.

2.1.2. Region-level methods

Compared to the previous ones, the region-level interest methods utilize the segmented ROI and can significantly reduce the feature dimension of the brain sMRI. For example, features were extracted from each pre-defined brain region using 93 ROIs labeled based on relevant expertise and an atlas warping algorithm. Then, the classical linear support vector machine (SVM) was used to diagnose AD. In addition, because atrophies in the hippocampus are often observed in the early stages of Alzheimer's disease, features of the hippocampus are often segmented and extracted from sMRI scans, and these resulting features are used in subsequent clinical tasks [36]. In Ref. [37], a method based on multi-core combined marginal Fisher analysis is proposed, which has the effect of sparse and dimensional ROI reduction. It can simultaneously explore various relationships between ROI information and AD. However, these pre-defined ROIs rely on medical experts' experience.

defining the disease and segmenting ROIs can be a time-consuming and labor-intensive process.

Moreover, the majority of such methods essentially a subset of hand-crafted features extracted from ROIs (e.g., only gray matter), which may ignore information that may be relevant to the AD.

2.1.3. Patch-level method

Visual-level methods can result in high dimensionality and small sample sizes, while region-level methods heavily rely on expert experience. Patch-level methods, positioned between these two approaches, proved to be more effective in capturing local brain lesions [15, 34, 77]. Weak classifiers were employed for classification by extracting patch information from brain scan images and utilizing the information obtained from different patches collectively for AD classification. Specifically, anatomical landmark detectors can be used to locate the central coordinates of these patches [34] or statistical methods [15]. Determining the center of the patch based on the anatomical landmark detector still relies on the prior knowledge of experts, which undoubtedly returns to the problem enumerated before. Determining the patch needs to consume huge resources. Therefore, it is necessary to propose a method that can automatically determine the patch center.

1.2. Self-supervised algorithm

SSL (Self-supervised Learning) is a method for training deep learning models without explicitly providing human annotations. The fundamental concept of self-supervised learning is to extract high-level features from unlabeled data, enabling these features to be effectively utilized in downstream tasks, such as classification and segmentation [36]. The principle behind SSL is to learn a meaningful representation of the data by grouping similar samples together and distinguishing them from dissimilar samples. The goal of contrast learning is to train the model by maximizing the similarity of similar pairs of samples and minimizing the similarity of dissimilar pairs of samples. The core of this approach is to compare multiple views of the same sample or pairs of samples from similar distributions to learn the feature representation of the sample.

SSL technology has been widely applied in image, speech, and natural language processing, yielding invaluable results in image segmentation, image classification, and other computer vision-related tasks. Self-supervised models include many types, such as Auto Encoder [77], Generative Adversarial Models [80], Contrastive Learning [80], etc. These methods can learn the intrinsic representations of unlabeled data without manual annotations. Among the various self-supervised learning methods, contrastive learning has been a hot topic of research in recent years. Contrastive Learning (CL) is a method that shortens the distance between similar objects or texts, making them closer, while pushing apart the distances of dissimilar data to enhance their discriminative properties [79]. Therefore, in contrastive learning, similarity assessment is used to measure the difference between two outputs, achieving this objective. In the context of image classification, the contrastive loss evaluates the underlying representation of images after passing through an encoder, followed by an MLP (Multi-layer Perceptron) layer for further feature extraction. For instance, after obtaining the sample data, common data augmentation techniques such as random cropping and adding Gaussian noise are applied to increase sample diversity. The augmented versions of the original images are considered positive samples, while other samples from the same batch are treated as negative samples. The similarity between samples is measured by calculating the similarity of positive sample pairs and the similarity of negative sample pairs. Through this process, the model can learn a meaningful representation of the data so that similar samples are closer together in the feature space and dissimilar samples are further apart. This representation capability can be useful in many computer vision tasks, such as image classification, object detection, and image generation. The advantage of contrast learning is that it

does not need annotated data, but only uses the internal structure and similarity information of the data itself for training, so it has strong adaptability and scalability.

Self-supervised Learning was widely used in the field of image. For example, Ref. [40] used the motion information and appearance features of dynamic objects for self-supervised learning and realized the feature representation learning of unlabeled image sequences. Furthermore, it achieved good performance in tasks such as object detection and image classification. At the same time, self-supervised learning also has important applications in the field of natural language processing. Ref. [41] was proposed by performing self-supervised learning on large-scale corpora, this paper presents a pre-trained language model based on transformers, which is an important milestone in the field of natural language processing.

(1) Autoencoder [77] - Autoencoder is a commonly used unsupervised deep learning model for data dimensionality reduction and feature extraction. The architecture of an autoencoder consists of an encoder and a decoder, which can learn the underlying representation of data by reconstructing the input signal. In recent years, autoencoders have been widely applied in various fields.

(2) MoCo [42] - MoCo (Momentum Contrast) is a self-supervised learning method for learning image representations. It learns image features by contrast learning tasks, where the model is trained by maximizing the similarity between positive sample pairs and minimizing the similarity between negative sample pairs. An image is selected from the input image as a query image and mapped to a query feature vector by the encoder. Then select a pair of positive sample images and their corresponding feature vectors (keys) from a buffer (Memory Bank). The buffer stores the feature vectors of the historical batches and is used to build positive sample pairs. Select a number of images from the input as negative sample images. These negative sample images share parameters with the query images in the encoder and generate corresponding feature vectors. The construction of negative sample pairs encourages the model to learn useful features by maximizing the difference between positive and negative sample pairs.

(3) SimSiam [43] - SimSiam is a simple self-supervised learning method that learns representations by predicting the output of a network after its input undergoes a series of transformations. Unlike MoCo, it does not require the use of negative samples and online networks. SimSiam has achieved comparable results to MoCo on many vision tasks.

(4) RotNet [44] - RotNet model can be used for image rotation prediction. Specifically, RotNet model works by rotating the input image by a certain angle before feeding into the network. In this way, the model can obtain and memorize the latent representation of the rotated image. During the inference phase, the model can accept input images at arbitrary angles and assess whether the image is rotated and provide the rotation angle. This ability of predicting rotations can be beneficial for tasks such as image classification, object detection, and image reconstruction, as it enhances the model's generalization and reliability.

(5) SIMCLR [17] - SIMCLR is a self-supervised learning model based on contrastive learning, which can improve the generalization ability of the model by maximizing the representation learning of different views of the same image. The model uses data augmentation and negative sample comparison techniques to train the model, which effectively improves the representation learning effect of the model. Experiments show that SIMCLR model outperforms many existing unsupervised learning methods on ImageNet, demonstrating its effectiveness in the aspect of contrastive learning.

2.3 Attention mechanism

Due to the potential for further improvement in image classification performance, the attention mechanism have been proposed and widely used to automatically find and highlight the regions with the highest

Table 1
ADNI and external datasets subject information statistics (MMSE: Mini-Mental State Examination).

Dataset	State	Quantity	Age	Gender	MMSE
ADNI	AD	312	75.9 ± 6.2(55–91)	162M/150F	21.2 ± 6.1(12–28)
	EMCI	912	76.1 ± 7.2(61–92)	106M/152F	26.2 ± 2.7(23–30)
	LMCI	271	76.5 ± 7.2(63–90)	153M/117F	28.2 ± 2.3(24–32)
	NC	450	76.4 ± 7.3(60–92)	242M/217F	29.0 ± 1.2(24–30)
Caltech	AD	137	76.0 ± 7.3(55–91)	67M/70F	23.1 ± 2.9(18–27)
	EMCI	79	74.8 ± 7.4(55–90)	43M/36F	26.5 ± 1.9(23–30)
	LMCI	134	74.5 ± 7.2(58–91)	68M/66F	27.2 ± 1.7(24–30)
	NC	163	76.2 ± 5.4(60–90)	76M/87F	28.1 ± 1.0(25–30)

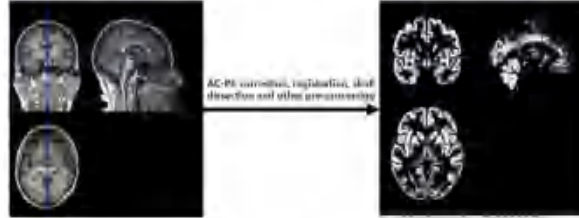


Fig. 2. On the left is the original MRI, and on the right is the pre-processed image after skull detection, registration, and head motion correction, etc.

3.1. Data and preprocessing

This study used a dataset from the public Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu/>). The dataset contains different T1-weighted (1.5T/3T) structural MRI scans of 1365 participants performed during screening visits at ADNI-1/3. Two data sets are used, one of which is the selected from ADNI and the other is from [30] with the same distribution. Based on standard clinical assessments such as Mini-Mental State Examination (MMSE) scores, ADNI dataset was divided into 4 groups, i.e. 332 AD patients, 271 LMCI patients, 312 EMCI patients, and 450 NC subjects. The demographic information of ADNI participants and the dataset of [30] are described in detail in Table 1.

As shown in Fig. 2, the original structural MRI data downloaded from ADNI have undergone pre-processing to facilitate the improved feature learning and classification in subsequent analyses. First, the obtained Nifti (3D Neuroimaging Informatics Technology Initiative) format was corrected by AC-PC (Anterior Commissure-Posterior Commissure) [31], and intensity correction by B1 non-uniformity correction [32]. B1 non-uniformity correction methods are usually based on signal strength differences during magnetic resonance imaging to calculate and correct the non-uniformity of the magnetic field. This process can be achieved by performing specific scanning and processing before and after imaging. The purpose of intensity correction by B1 non-uniformity correction is to eliminate the intensity variation caused by B1 non-uniformity, so that the image has a more uniform intensity distribution throughout the imaging region. This helps improve image quality, reduce noise and artifacts, and provide more accurate anatomical and pathological information. Secondly, a linear registration was performed to the Colin27 template [33] to remove global linear differences, such as differences in head movement, size, and rotation were followed by skull stripping on sMRI scans of the brains of each participant. SPM [34] is a tool used together with matlab to segment and extract various voxel features in brain scan images.

This study utilized the SPM12 toolbox and applied original parameters for pre-processing (<http://www.fil.ion.ucl.ac.uk/spm/>). GATK2 used the original parameters to segment the brain structure (<http://www.neuro.uni-jena.de/gat2/>), and only the voxels with in the brain were used for feature construction. Finally, MR images were normalized to the Colin27 standard space, resulting in a voxel size of $181 \times 217 \times 161$.

3.2. Patch-based data augmentation framework

In this study, the whole brain is no longer used for research because of data redundancy and high computational cost, so the features of the entire brain structure were represented by reassembled key MRI patches.

We regarded the analysis of brain morphometry patterns for AD diagnosis as a patch combination problem. And a simple contrast learning model based on patch-based data augmentation visual representation was constructed for the diagnosis of Alzheimer's disease.

Abnormal brain structure lesions occurred in a few local areas, especially in the early onset stage of AD [54,55]. Therefore, in this paper, the combination of patches in the patient's MRI was considered as a positive combination. Accordingly, the patch combinations in the cognitively normal group is considered to be negative. Finally, the selected patch represents the training data of the entire brain structure model.

The proposed model contained four key steps shown in Fig. 1, included a data augmentation layer, an attention layer applied to the patch combination, a feature extraction layer (namely resnet18), and a fully connected classification layer.

3.3. Proposed patch location selection

The patch correlation calculation layer is to reduce the data dimension and the computing resources and time. Alzheimer's disease does not have lesions in the whole brain structure, mainly concentrated in the hippocampus and temporal cortex [54,55]. The hippocampus is an important center for memory in the brain, while the temporal cortex is involved in a variety of functions such as language understanding, spatial perception, and emotional control.

Therefore, the model should focus on these affected areas in the early diagnosis and treatment of Alzheimer's disease. The PCC (Pearson Correlation Coefficient) is a statistical measure of the linear correlation between two variables, which can be calculated as Eq. (1).

$$PCC(i) = \frac{\sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y})}{\sqrt{\sum_{t=1}^n (X_t - \bar{X})^2} \sqrt{\sum_{t=1}^n (Y_t - \bar{Y})^2}} \quad (1)$$

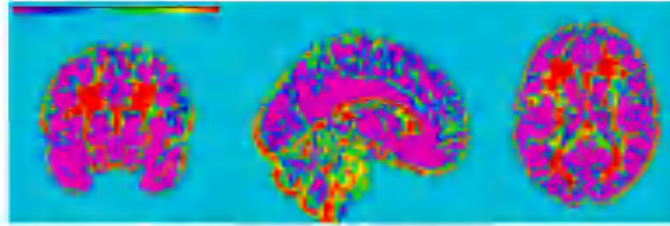


Fig. 3. The PCC correlation intensity maps of the brain.

Here, i represents a voxel, and X and Y represent the average lifted voxel features of different groups, respectively. X_i and Y_i represent different groups of voxels respectively.

PCC can measure the strength and direction of the linear relationship between two variables, with values ranging from -1 to 1 . In the feature selection process, PCC can be used to evaluate the correlation between each feature and the target variable. A high correlation coefficient means that the feature has a strong linear relationship with the target variable and may be an important feature. The difference between the AD and NC groups was explored in the correlation of brain patches. As shown in Fig. 3, PCC was adopted to assess the strength and direction of the linear relationship of patch mean voxel features between two groups. For each patch, the feature values in each mean voxel were calculated. Then, PCC was used to calculate the correlation among mean voxels. The correlation coefficient matrix was obtained for each patch and compared among different groups. In this way, PCC can eventually locate the center of the patch and measured the correlation of different sets of eigenvalues.

By this method, the significant difference of patch correlation between the AD group and the NC group is identified, thus providing strong support for the early diagnosis of Alzheimer's disease. As shown in Fig. 3, the dark purple region is negatively correlated, while the dark red region is positively correlated. The closer to the left of the reference bar in the diagram the color is, the more uncorrelated the eigenvalues are. In summary, the correlation coefficient of each feature is calculated as follows.

3.4. Patch-based data augmentation module

In order to evaluate each brain region, a patch screening network was designed. In this network, 5% of the images of the dataset were used for patch screening, specific details are shown in Fig. 5. As shown in Fig. 1, the network was mainly composed of the following three convolutional layers. Each convolution layer was followed by RELU layer and BN layer, and the Global Average Pooling layer successively. After the above convolution operation, the patch feature map (y) was obtained. These operations can be expressed as follows.

$$y = \text{BN}(\text{ReLU}(\text{Conv}(patch))) \quad (2)$$

where Conv represents the convolution operation.

In addition, after the first convolutional layer, another Global Average Pooling layer was used to obtain another Affect Score. This process can be expressed as:

$$\text{Affect Score} = \text{SoftMax} [GAP(y_1) \otimes GAP(y_2)] \quad (3)$$

in which y_1 and y_2 are the results of the first convolution and the third convolution, respectively. \otimes represents tensor multiplication, and GAP is Global Average Pooling. Finally, their tensor product was calculated through SoftMax layer, and the effective patch fraction was obtained. The network used the model to score the features of each brain region

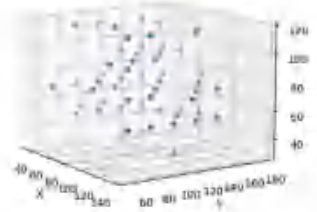


Fig. 4. The center positions of the selected patches.

and filtered according to the value. More accurate multiple brain regions was selected by scoring. Eventually, patches from different brain regions were recombined to represent the global features of the entire brain.

In this study, 54 patches were selected and randomly divided into two groups equally. This process was regarded as Patch-based Data augmentation. The center positions of the selected patches for these two groups are shown in Fig. 4.

The two groups of patches represented different data augmentation results of the same global representation of the brain, so as to generate two contrastive views as the data input of the contrastive learning framework. In this way, the change and noise of the data can be more accurately used for data augmentation, so as to enable the model to focus on learning the key regions of the images.

And the resulted contrast view can be expressed as follows:

$$PATCH = \{patch_i\}_{i=1}^N \quad (4)$$

where i represents the stack of each patch, N represents the recombined patches, and i is one of the patches within it, and PATCH represents the combination of multiple patches.

More reliable training samples were provided for the overall framework, thereby improving the classification accuracy and robustness of the model. This method of generating contrastive views can improve the model's perception of local information, increase the diversity of data, and improve diagnostic capability of the model. The above process is shown in Fig. 5.

The coordinates of the selected brain regions and the positioning information in the brain are shown as Fig. 6, which can assist doctors in medical diagnosis.

3.5. Feature extraction block with attention

The feature extraction layer in this paper was based on the classical resnet18 network [21] with appropriate modifications and optimization. To adapt to gray matter medical images, the first convolution layer was modified during the experiment (channel = 1, kernel size =

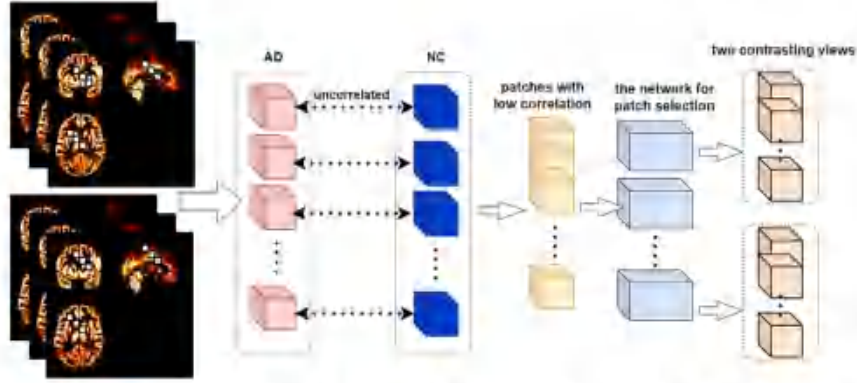


Fig. 5. The process is divided into three steps. Firstly, the PCC of different regions of AD and NC images is calculated. Secondly, the selected patch is input into the patch screening network. Finally, two contrastive views are generated as input data for subsequent models.



Fig. 6. Mapping of the selected patch three-dimensional space coordinate center position on brain MRI.

3, stride = 3, padding = 1). In addition, to strengthen the detection ability of the model for the key brain lesions of the patch combination, an additional attention mechanism was added at the beginning of the framework for improving the accuracy and robustness of the model. The spatial attention mechanism can reveal those parts emphasized or suppressed in the feature representation within the contrastive views.

By multiplying the feature maps of the contrastive views with the computed spatial attention map, a final representation with local spatial attention perception was generated. Subsequently, several equations are adopted to elucidate the attention mechanism.

The representation of the attention block's output is denoted as P , and the formula for maximum pooling along the channel axis is defined as follows.

$$P_{max} = Channel\ Max\ Pooling(P) \quad (5)$$

To represent average pooling over each channel's features, it can be expressed as:

$$P_{avg} = Channel\ Average\ Pooling(P) \quad (6)$$

Here,

$$P = P \otimes A_{spatial} \quad (7)$$

where \otimes is element level multiplication. The spatial attention map $A_{spatial}$ is calculated by connecting two feature maps as follows.

$$A_{spatial} = \sigma(W([P_{max}, P_{avg}])) \quad (8)$$

where σ is S-type activation, W is the weight of the convolution layer, and $[\cdot]$ is the connection.

3.6. Loss function

A loss function called Normalized Temperature Scaled Cross Entropy Loss (NT-Xent) was used to learn representation, which is a contrastive loss function used to measure the similarity between two representations.

Given a pair of positive samples (x_i, x_j) , where x_i and x_j are both data-augmented samples, the NT-Xent loss function is calculated as follows:

$$l_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j) / \tau)}{\sum_{k \neq i} \mathbb{1}_{(x_i, x_k)} \exp(\text{sim}(z_i, z_k) / \tau)} \quad (9)$$

where z_i and z_j are the encoded representations, $\text{sim}(\cdot, \cdot)$ is the similarity function, τ is a temperature parameter, and N is the batch size. $\mathbb{1}_{(x_i, x_k)}$ is an indicator function that is 1 when $k \neq i$ and 0 otherwise. The purpose of this indicator function is to exclude the possibility of itself being a negative sample. The NT-Xent loss function prompts the model to explore feature representations that can narrow the distance between positive and negative samples by calculating the similarity of positive samples in the same batch and normalizing negative samples. The similarity is regulated by the temperature parameter τ , which can narrow the range of similarity and alter the model's robustness.

To validate the model's ability to learn image representation, a single fully connected layer is adopted as the linear classifier. In this module, except for using cross-entropy as the loss function, no other operations were performed.

3.7. Implementation

The PD-SIM framework proposed in this study was implemented in Python using PyTorch software package. In PD module, the neural networks for each patch do not share weights, which can identify distinct patches although it increases the number of trainable parameters. It is worth noting that model's learning ability on key regions can be effectively enhanced by extracting patch combinations from different anatomical structures as network image inputs.

Two datasets (ADNI and Cuhgnet) were used to test the performance and generalizability of PD-SIM method. Specifically, ADNI data set was divided into training set and testing set. k -fold cross-validation was adopted to avoid overfitting during training process. In this study, k was set to 5. The data was divided into 5 equal parts through random partitioning. Then, 80% (i.e., four parts) of the data was used for

Table 2 Diagnosis of AD (AD vs. NC) and early diagnosis of AD (NC vs. EMCI) on the ADNI testing set.

Method	AD VS. NC					NC VS. EMCI				
	ACC	SEN	SPE	AUC	F1	ACC	SEN	SPE	AUC	F1
Autoencoder	0.6191	0.6079	0.5775	0.7144	0.6441	0.6239	0.6609	0.5585	0.6678	0.6798
MoCo	0.5922	0.5976	0.6111	0.7257	0.7116	0.6176	0.7281	0.6926	0.7610	0.7604
MoCo	0.6148	0.6235	0.6031	0.6567	0.6323	0.5689	0.6103	0.6490	0.6456	0.5959
SimSiam	0.6715	0.7057	0.6295	0.7074	0.7036	0.6635	0.6977	0.6176	0.7591	0.7673
PD-SIM	0.7970	0.7726	0.8148	0.8543	0.7736	0.7936	0.7103	0.6947	0.8974	0.8466

training process, and 20% (one part) was used for testing. This data splitting procedure was repeated 5 times, ensuring each data partition was involved in the testing phase.

By optimizing hyperparameters, a set of optimal model parameters were obtained and evaluated on the ADNI dataset. To verify the model's robustness and generalization ability, the model was tested on another dataset (Gingret). These steps ensure that model can be generalized to new datasets and has good performance.

4. Experiment

In this section, the experimental setup, model performance comparison, and generalizability of PD-SIM method will be introduced. The proposed model is compared with the Self-Supervised Learning methods on multiple AD-related diagnostic tasks.

4.1. Experimental setup

When training PD-SIM, by utilizing a group comparison approach, PCC values are computed on a 5% dataset, comprising AD group and NC group, covering the entire MRI images. These PCC values are then used to initialize the positions of input image patches. Then, patches are extracted from the selected positions in the MRI image and input to the corresponding PD-SIM network for training. The Adam optimizer with 100 epochs is recommended, running on NVIDIA RTX 2080Ti GPU with approximately 5.5 h, followed by evaluation on the resulting testing subset.

The network structure of PD-SIM (such as the number of channels) and hyperparameters (such as learning rate = $3e-6$, batch size = 64, patch size = $21 \times 21 \times 21$) are selected based on the average validation performance across all cross-validation folds.

Five metrics were used to evaluate the classification performance in the experiment, including accuracy (ACC), area under the curve (AUC), sensitivity (SEN), specificity (SPE), and F1 score (F1). TP (True Positive) represents correct predictions and FP (False Positive) represents incorrect predictions for real patients. Conversely, TN (True Negative) represents correct predictions and FN (False Negative) represents incorrect predictions for normal cases.

Except for AUC, the definitions of these metrics are shown as follows:

$$Accuracy = \frac{TP + TN}{FN + TN + TP + FP} \quad (10)$$

$$Sensitivity = \frac{FN + TP}{FN + FP} \quad (11)$$

$$Specificity = \frac{TN}{FN + FP} \quad (12)$$

$$Specificity = \frac{2 \times TP}{2 \times TP + FN + FP} \quad (13)$$

The calculation of AUC can be divided into two steps. Firstly, the ROC curve is plotted by sorting the prediction results of the classification model in descending order of confidence. By varying the classification threshold, the samples are classified into positives and negatives, and True Positive Rate (TPR) and False Positive Rate (FPR) are computed. Then, AUC is calculated, which represents the area under the ROC curve and indicates the overall performance of the model at different thresholds.

In this study, the proposed PD-SIM method is compared with four Self-Supervised Learning algorithms applied in medical image analysis, i.e., Autoencoder, MoCo, SimSiam and SIMCLR. All of these methods have been introduced in the previous section, and are trained. The same data partitioning approach is utilized ensuring that the training and testing sets used in different methods are identical.

4.2. Classification performance on ADNI dataset

The performance of PD-SIM method and competing methods on AD classification and early diagnosis of AD on the testing set is shown in Table 2. The PD-SIM method has better performance for AD diagnosis in most self-supervised scenarios. For example, in the diagnosis of AD and NC, PD-SIM method has achieved a leading position in all indicators, among which ACC reached 0.7970 and SEN reached 0.7726. These two indicators are relatively important in medical diagnosis and treatment. In addition, in the early diagnosis of AD tasks, the PD-SIM method obtained significantly better results than the other four methods, where ACC reached 0.7936, although lower than the experimental results described above, but this may be because the early lesions of AD are not obvious. The above experimental results may be attributed to the improvement of the model's ability to learn the structure of brain injury, which is due to the addition of the screening module of patch, which improves the model's learning ability.

Compared with the existing self-supervision methods applied to medical images, the patch-based data-enhanced self-supervision algorithm can achieve better results. This is mainly because patch-based self-supervision method can learn more accurate image representation, and at the same time alleviate the overfitting problem caused by small sample high-dimensional data, that is, greatly reduce the dimension of input images, and improve the feature recognition ability of key brain patches. As is shown in Table 3, the overall classification performance of the proposed method is better and faster under the same input. PD-SIM model can effectively learn different representations of data, and has a good effect on AD detection. Specifically, the PD-SIM method proposed in this study has achieved a good improvement in accuracy, which means that the PD-SIM method in the self-supervised algorithm has a higher diagnostic accuracy in AD classification. Experimental results show that PD-SIM method has a high ability to identify brain injury areas.

To evaluate the performance of PD-SIM in other classification tasks, additional experiments were conducted on a number of challenging classification groups.

Specifically, the classification of EMCI versus LMCI and NC versus LMCI is a challenging task because there are only slight structural differences in the brain during the transition phase of Alzheimer's disease. As shown in Table 4, the PD-SIM method shows superior performance in the above classification tasks. In terms of diagnostic outcomes related to MCI conversion, there are four indicators in the lead, among which ACC reached 0.7492, SEN reached 0.8451, and SPE reached 0.5911, the value of SPE is not the best. At the same time, it can be noted that the SPE performance of EMCI and LMCI, NC and LMCI in Table 4, and NC and EMCI in Table 5 is not ideal, which may be due to the imbalance of categories, which affects the calculation of specificity. If the sample size of a class is small, the model may be more likely to misclassify it as another class, possibly because of overlapping

Table 3

Comparison of prediction accuracy (AD vs. NC) and computational cost of the model used on ADNI			
Method	Input size	ACC	Downstream task: take time per image
Autoencoder	$101 \times 217 \times 101$	0.6181	76.37 ms
SIMCLR	$101 \times 217 \times 101$	0.6422	451.2 ms
MoCo	$101 \times 217 \times 101$	0.6148	222 ms
SimSiam	$101 \times 217 \times 101$	0.6715	195.7 ms
PD-SIM	$101 \times 217 \times 101$	0.7471	52.6 ms

Table 4

MCI conversion (EMCI vs. LMCI) and diagnosis task of distinguishing LMCI from NC (NC vs. LMCI) on The ADNI testing set.

Method	EMCI vs. LMCI					NC vs. LMCI				
	ACC	SEN	SPE	AUC	F1	ACC	SEN	SPE	AUC	F1
Autoencoder	0.6002	0.7997	0.4286	0.6183	0.7043	0.6705	0.7522	0.5232	0.5564	0.6457
SIMCLR	0.5352	0.3703	0.9677	0.7093	0.5333	0.6232	0.8995	0.7596	0.5456	0.7469
MoCo	0.5864	0.4010	0.7742	0.7012	0.4946	0.6284	0.5516	0.6743	0.6232	0.6360
SimSiam	0.6897	0.5816	0.8764	0.6973	0.5965	0.6573	0.7528	0.4259	0.6983	0.6007
PD-SIM	0.7492	0.8451	0.5911	0.7246	0.8169	0.7480	0.8168	0.6092	0.7437	0.8008

Table 5

Diagnosis of AD (AD vs. NC) and early diagnosis of AD (NC vs. EMCI) on the Cuingnet testing set.

Method	AD vs. NC					NC vs. EMCI				
	ACC	SEN	SPE	AUC	F1	ACC	SEN	SPE	AUC	F1
Autoencoder	0.6873	0.7462	0.656	0.7713	0.7005	0.6412	0.4895	0.7154	0.4326	0.4654
SIMCLR	0.6657	0.7072	0.6118	0.7065	0.5209	0.6023	0.2675	0.8483	0.5063	0.3365
MoCo	0.5541	0.6041	0.975	0.675	0.6002	0.5451	0.7265	0.6509	0.5001	0.3178
SimSiam	0.6253	0.5294	0.7183	0.7532	0.6573	0.6346	0.2579	0.7397	0.5582	0.3451
PD-SIM	0.7588	0.8612	0.672	0.7210	0.7655	0.6881	0.7042	0.6805	0.7432	0.5891

disease stages. For example, there may be a blurred boundary between EMCI and LMCI, which makes it more difficult to distinguish them, thus affecting the calculation of specificity. In general, PD-SIM method is superior to other methods in the diagnosis of LMCI and EMCI.

4.3. Generalization performance on Cuingnet

To validate the generalizability of the proposed method, PD-SIM method is further evaluated on the Cuingnet dataset. The experimental results for AD diagnosis and early diagnosis of AD are presented in Table 5.

It is obvious that the proposed PD-SIM method generally outperformed the other four competing methods (i.e. Autoencoder, MoCo, SimSiam, and SIMCLR) in most metrics for both AD-related diagnostic tasks. Specifically, PD-SIM achieved the best ACC (0.7588) for AD classification diagnosis. PD-SIM outperformed Autoencoder (0.6875), SIMCLR (0.6657), MoCo (0.5864), and SimSiam (0.6897) in the AD classification task on the Cuingnet dataset. For the diagnosis task of NC vs. EMCI, the proposed PD-SIM method outperformed the second-best method with more impressive results. These results demonstrate that the proposed PD-SIM method exhibits robustness across different datasets. More comparison results are shown in Tables 4 and 5. Notably, the specificity of results for AD and NC using the MoCo model was 0.975, while sensitivity was only 0.6041, meaning that the model was less capable of correctly identifying true positive samples (patients with AD) as positive, but was able to identify true negative samples (normal controls) as negative with a high degree of accuracy. It could be that there was an imbalance in the sample in the study, i.e. the number of AD patients was relatively small. This may result in poor performance and less sensitivity of the model in the identification of AD patients. This suggests that MoCo model may not be the best choice for the AD vs. NC classification task, or that the model parameter configuration used in the study may not be appropriate for the task. Other models or parameter settings may perform better in terms of sensitivity.

In most metrics for the AD task, the performance of PD-SIM had a significant decrease in sensitivity for the NC and EMCI tasks. One possible reason is that there were fewer EMCI samples in the Cuingnet

dataset, and a small number of misclassifications can significantly reduce sensitivity. Table 3 reports the computational costs, and it is clear that PD-SIM is significantly better than other models in terms of time consumption.

Overall, these results indicate that the proposed method demonstrates a pleasing level of generalization for the diagnosis of AD. These favorable results may be due to the fact that the proposed approach reduces the size of the input into the model, while improving the ability to acquire the pathogenic brain region and reducing the risk of overfitting.

5. Discussion

In this section, the effectiveness of data augmentation module (Patch-based Data augmentation framework) is evaluated and the impact of spatial attention in the proposed method. The potential for pathologic inaction and clinical transformation by this method is then mentioned. Finally, the limitations of the experiment are analyzed and the research direction is further explored.

5.1. Effectiveness of data augmentation module

In this study, further ablation experiments were conducted, specifically evaluating the effectiveness of the AttentionSIM module without patches. The purpose is to evaluate the effectiveness of the data augmentation module and attention module used in this study.

Both methods were evaluated through two AD-related diagnostic tasks (e.g. AD vs. NC and LMCI vs. EMCI), and the results are presented in Table 6, in which non-Attention is the module without spatial attention, non-augmentation is the use of conventional data augmentation without patch-based augmentation module, and non-both are SIMCLR.

Table 6 shows that the data augmentation module proposed in this study improved the performance of the model in various metrics. For example, PD-SIM method based on patch data augmentation module achieved higher accuracy than its counterpart, namely non-Augmentation, in AD diagnosis and MCI conversion diagnosis. The results indicate that using this module led to more promising diagnostic performance. The possible reason is that the data augmentation module

Table 6
Diagnosis of AD (AD VS. NC) and MCI conversion (EMCI VS. LMCI) on the ADNI dataset on.

Method	AD VS. NC					EMCI VS. LMCI				
	ACC	SEN	SPE	AUC	F1	ACC	SEN	SPE	AUC	F1
non-Attention	0.7409	0.5855	0.8579	0.8159	0.6587	0.6982	0.4815	0.8718	0.7128	0.5028
non-Augmentation	0.7548	0.5401	0.8912	0.7914	0.6210	0.6131	0.6981	0.9893	0.8074	0.7214
non-Both	0.6813	0.3925	0.8005	0.7203	0.6603	0.6388	0.3765	0.8077	0.7093	0.5002
PD-SIM	0.7970	0.7726	0.8148	0.6548	0.7736	0.7492	0.8452	0.9511	0.7246	0.8169

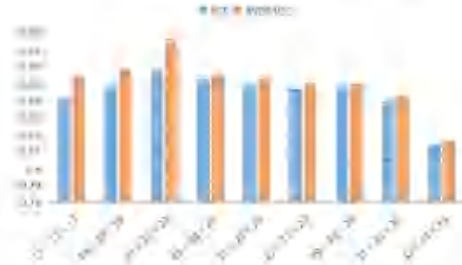


Fig. 7. The influence of patches with different sizes on AD classification.

can effectively screen the discriminative patch related to AD, while reducing the dimension of the input data, alleviating the dimensional curse brought by high latitude, and reducing the risk of overfitting. In this module, the most important thing should be the pre-screening of the discriminant brain region, which reduces the difficulty of the model to learn the underlying information of the image.

5.2. Influence of attention mechanisms

The spatial attention of the combined patches is further tested. Because the new patch combination clustered highly disease-related regions in the brain, it is possible that each region of the new global representation of the brain could not close the contribution gap for the diagnosis of AD, resulting in the reduction of the classification effect. Therefore, the spatial attention mechanism applied to the global representation of the brain can make the model pay more attention to the differences between the normal and atrophic brains. By placing more weight on key discriminative brain patches, the mechanism can improve the classification accuracy and provide more accurate auxiliary information for medical diagnosis. The model removing Attention mechanism is called PD-SIM without attention (non-Attention). The experimental results are shown in Table 5, which list the difference with the attention module and without the attention module. After removing this module, both AD and NC, and the difficult classification group (EMCI conversion group) showed a decline in the index. So, the attention module can improve the diagnostic ability of the model in the framework of this paper.

5.3. Influence of image patch size

Different sizes of patches may have varying impacts on the classification results, hence, brain patches with different sizes ($17 \times 17 \times 17$, $19 \times 19 \times 19$, $21 \times 21 \times 21$, $23 \times 23 \times 23$, $25 \times 25 \times 25$, $27 \times 27 \times 27$, $29 \times 29 \times 29$, $31 \times 31 \times 31$, $33 \times 33 \times 33$) are assembled to remove the training network in batches to train the network.

The corresponding AVERAGE, that is $(SEN+SPE+AUC)/3$, and ACC values are shown in Fig. 7. When a relatively small image patch (i.e. $17 \times 17 \times 17$) was input to the model, ACC (0.764) and AVERAGE (0.891) showed poorly, which indicates that a relatively small image patch could not provide sufficient brain structure information.

When the input size of the image is within the time interval ($21 \times 21 \times 21$, $23 \times 23 \times 23$), the satisfactory performance is obtained. It implies that these two sizes of image patches contain sufficient information about focal brain lesions. When the input image patch size is larger than $33 \times 33 \times 33$, the classification performance of PD-SIM decreases; it is possible that larger image patches contain more redundant information, which interferes with the model's ability to discern subtle changes in brain structure within the patches.

Finally, in order to include as much spatial information as possible while considering the computational overhead during training, patches of size $21 \times 21 \times 21$ were selected as the optimal input size for this study.

5.4. Comparison with previous works

For validating the generalization of our proposed model, several methods are selected to be compared. Since there are few MRI-based algorithms for correlated self-supervised Alzheimer's diagnosis, results were obtained using fMRI data multimodal data. Each method did not provide the source code and data, so we just extracted the corresponding results in original papers. Due to the difference in the number of research objects and image preprocessing steps, as well as the difference in the calculation methods of some indicators or the absence of some indicators, comparison can be referred roughly. As can be seen from the Table 7, except for the ACC values of INE and LSEI, most methods performed well in diagnosis of AD and NC. Among them, the accuracy of 3DCNN+GLw-BHb12 method reaches 0.804, which may be because the method used a large amount of data for pre-training. At the same time, the accuracy of the method proposed in this study reached 0.797, which has tiny gap without being pre-trained in advance with small consumption on time and resources. In addition, our proposed method in this study performs best on AUC values, and there is almost no difference between the first (3DCNN+GLw-BHb12) and second (ours) values on F1 values. In the diagnosis of MCI conversion, the results of this study are better than those provided in the table.

6. Conclusion

This study proposed a simple self-supervised learning framework based on patch-based data augmentation (PD-SIM) for computer-aided AD diagnosis, which consists of three main components: patch screening module, patch-based data augmentation module, and attention module. PD-SIM was evaluated for multiple AD-related diagnostic tasks including 2144 subjects from the ADNI and Cingrign dataset. The experimental results showed that the data augmentation method proposed in this study can not only identify different lesion locations in the brain, but also provide a data source for generating a comparative view. As a result, our model presented impressive classification performance. In general, the proposed method performed very well in most AD-related diagnosis scenarios. All indicators are in a leading position in the diagnosis classification of AD and NC, although the Specificity and Sensitivity were not always the most outstanding in the early diagnosis of AD, which may be related to the strategy of selecting patch.

Due to little data to determine the selected patch, this study only considered the brain MRI correlation between AD and NC. In the future work, we will explore the influence of the correlation between different categories and how adding more data to the PCC calculation would

- [29] M. Vounou, E. Janssens, R. Wolf, J.L. Stein, F.M. Thompson, D. Bruckner, G. Mistlari, Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in Alzheimer's disease, *NeuroImage* 60 (1) (2012) 700–716.
- [30] H. Coluţoi, E. Gerardo, J. Trastéras, G. Avants, S. Leblond, M.-D. Hoban, M. Chupin, H. Benali, O. Collin, Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of l₁ methods using the ADNI database, *NeuroImage* 36 (2) (2011) 766–781.
- [31] J. Barua, G. Chételat, B. Desgranges, G. Perleay, S. Landeau, V. de La Sayette, F. Bastard, In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease, *NeuroImage* 14 (2) (2001) 298–309.
- [32] H. Minoda, Voxel-based morphometry of brain MRI in normal aging and Alzheimer's disease, *Aging Dis* 4 (1) (2013) 29.
- [33] F. Gu, X. Liu, J. Yang, B. Zhao, M. Huang, J. Zhang, O. Zaïene, Nonlinearly-aware based dimensionality reduction and over-sampling for AD/MCI classification from MRI measures, *Comput. Biol. Med.* 91 (2017) 91–97.
- [34] J. Zhang, Y. Guo, Y. Guo, B. Russell, D. Shen, Detecting anatomical landmarks for late Alzheimer's disease diagnosis, *IEEE Trans. Med. Imaging* 35 (12) (2016) 2524–2533.
- [35] Q. Shuang, P.S. Joshi, M.J. Miller, X. Chouhan, Z. Xie, K. Gody, C.G. H. A.S. Joshi, D. Brigid, Z. Shaban, Development and validation of an interpretable deep learning framework for Alzheimer's disease classification, *Brain* 143 (5) (2020) 1920–1933.
- [36] L. Jing, Y. Tian, Self-supervised visual feature learning with deep neural networks: A survey, *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (11) (2020) 4037–4056.
- [37] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science* 313 (5786) (2006) 504–507.
- [38] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A. Bharath, Generative adversarial networks: An overview, *IEEE Signal Process. Mag.* 35 (1) (2018) 53–65.
- [39] A. Jaiswal, A.R. Babu, M.Z. Zaidi, D. Benerjee, H. Malhotra, A survey on contrastive self-supervised learning, *Technologies* 9 (1) (2020) 2.
- [40] X. Zhai, X. Fan, Z. Liu, D. Liu, C.C. Loy, Self-supervised learning via conditional action propagation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1881–1889.
- [41] J. Devlin, M. Chung, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018, *arXiv preprint arXiv:1810.04805*.
- [42] K. He, H. Fan, Y. Wu, S. Xie, B. Girshick, Momentum contrast for unsupervised visual representation learning, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9729–9738.
- [43] X. Chen, R. He, Exploring simple siamese representation learning, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15750–15758.
- [44] S. Gidaris, P. Singh, N. Komodakis, Unsupervised representation learning by predicting image rotations, 2018, *arXiv preprint arXiv:1804.07728*.
- [45] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, Y. Bengio, Show, attend and tell: Neural image caption generation with visual attention, in: *International Conference on Machine Learning*, PMLR, 2015, pp. 2048–2057.
- [46] S. Wias, J. Park, J. Lee, I.S. Kwon, Chant: Convolutional block attention module, in: *Proceedings of the European Conference on Computer Vision*, ECCV, 2018, pp. 3–19.
- [47] H. Wang, Y. Pan, Z. Wang, L. Jiao, B. Schiele, Parameter-free spatial attention network for person re-identification, 2018, *arXiv preprint arXiv:1811.12150*.
- [48] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.
- [49] R. Xu, Z. Liu, Y. Luo, H. Hu, L. Shan, B. Du, K. Huang, J. Yang, SQA: Towards 3D universal pulmonary nodule detection via slice grouped domain attention, *IEEE/ACM Trans. Comput. Biol. Bioinform.* (2023).
- [50] J. Jovicich, S. Calmon, D. Grise, E. Haley, A. van der Kouwe, R. Gollub, G. Kennedy, P. Schmitz, G. Brown, J. MacFall, Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data, *NeuroImage* 30 (2) (2006) 436–443.
- [51] P. Naruyana, W. Bey, M. Kulkarni, C. Sivasubbar, Compensation for surface coil sensitivity variation in magnetic resonance imaging, *Magn. Reson. Imaging* 6 (3) (1988) 271–274.
- [52] C. Hulst, B. Hege, I. Collins, K. Woods, A. Toga, A. Evans, Enhancement of MR images using registration for signal averaging, *J. Comput. Assist. Tomogr.* 22 (2) (1998) 324–333.
- [53] J. Polzella, H. Yoon, K. Tubelov, Structural adaptive segmentation for statistical parametric mapping, *NeuroImage* 52 (2) (2010) 515–523.
- [54] H. Braak, E. Braak, Neuropathological staging of Alzheimer-related changes, *Acta Neuropathol.* 82 (4) (1991) 239–250.
- [55] C. Jack, D. Knopman, W. Jagun, R. Petron, M. Weiner, P. Alami, L. Shaw, P. Vemuri, H. White, S. Wengand, Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers, *Lancet Neurol.* 12 (2) (2013) 207–216.
- [56] J. Dattay, Q. Zhou, E. Adet, G. Zaharchuk, K. Pohl, Self-supervised learning of neighborhood embedding for longitudinal MRI, *Med. Image Anal.* 82 (2022) 102571.
- [57] Q. Zhao, Z. Liu, E. Adet, K. Pohl, LSSL: Longitudinal self-supervised learning, *Med. Image Anal.* 71 (2021) 102051.
- [58] N. Dhanraj, S. Thiagarajan, P. Rajagopalan, D. Sriraj, J. Ambika, G. Ver Steeg, P. Thompson, Evaluation of transfer learning methods for detecting Alzheimer's disease with brain MRI, in: *18th International Symposium on Medical Information Processing and Analysis*, Vol. 12567, SPIE, 2023, pp. 504–513.
- [59] A. Petrov, E. Geerjard, L. Wu, T. Sylvain, T. DeRamus, M. Luck, M. Minnie, G. Mitterle, R. Helm, S.M. Pils, Self-supervised multimodal learning for group inference from MRI data: Discovering disorder-relevant brain regions and multimodal links, *NeuroImage* 285 (2024) 120485.

3.5. 基于聚类框架与局部感受野的实时人脸疲劳检测



- ◆ 英国《科学文摘》(INSPEC)来源期刊
- ◆ 俄罗斯《文摘杂志》(AJ)来源期刊
- ◆ 美国《乌利希期刊指南(网络版)》(Ulrichsweb)收录期刊
- ◆ 2017—2019年中国国际影响力优秀学术期刊(自然科学与工程)
- ◆ 第二届国家期刊奖百种重点科技期刊
- ◆ 中国科技核心期刊 ◆ 全国中文核心期刊
- ◆ 中国科技论文统计源期刊
- ◆ 中国学术期刊综合评价数据库来源期刊
- ◆ RCSE核心期刊 ◆ 中国期刊方阵双效期刊
- ◆ 《日本科学技术振兴机构数据库》(JST)来源期刊
- ◆ 美国《艾博思科学数据库》(EBSCO)全文来源期刊
- ◆ 美国《剑桥科学文摘(自然科学)》(CSA (NS))核心期刊
- ◆ 波兰《哥白尼索引》(IC)来源期刊
- ◆ 中国科学引文数据库(CSCD)来源期刊
- ◆ 《中文科技期刊数据库》来源期刊
- ◆ 《中国期刊网》《中国学术期刊(光盘版)》来源期刊
- ◆ 中国精品科技期刊顶尖学术论文(F5000)项目来源期刊
- ◆ 《电子科技文献数据库》来源期刊
- ◆ 《中国工程技术电子信息网》来源期刊

计算机应用研究

Jisuanji Yingyong Yanjiu

第 37 卷 第 12 期 2020 年 12 月

目 次

综述评论

- 高维点测测技术综述 梅 林, 张凤蓉, 高 强(3521)
- 面向大数据的 K-means 算法综述 任远航(3528)
- 基于 IAS 系统的位图轨迹隐私保护技术综述 张青云, 张 兴, 李万杰, 李隆会(3534)

算法研究探讨

- 面向大数据的图模式挖掘模型算法 姜雨薇, 李叶飞, 吕龙龙, 陈晋鹏, 魏林冲(3545)
- 科技社交网络中基于图神经网络分析的表级排序学习推荐方法研究 董 博, 王吉宏, 张耀斌, 王 刚(3552)
- 基于时间序列聚类的轨迹停留点检测算法 兰志辉, 陈 旭, 段治州(3557)
- 改进向不平衡图神经网络学习分类模型研究 傅忠明, 刘 鹏, 钱大高, 杨传杰, 梁 刚(3561)
- 融合协同过滤和三元模型的推荐系统推荐算法 高俊彪, 杨 雷(3565)
- 一种基于改进的巴氏系数的协同过滤推荐算法 王 伟, 周 涛(3569)
- 一种基于自注意力机制的推荐方法 刘宗瀚, 任洪凯, 魏林浩(3572)
- 改进图补法和多权重相似度相结合的推荐算法 邹 洋, 吴和成, 姜立志, 赵应丁(3578)
- 基于多流降维自编码器模型的 top-k 推荐算法 方义秋, 俞晨曦, 葛君伟(3582)
- 基于注意力聚集算法的大数据挖掘 田 平, 何 翼(3586)
- 改进神经网络神经网络在欺诈信息的应用研究 孙林娟, 贾月辉(3590)
- 基于语义依存分析的图网络文本分类模型 范国凤, 刘 刚, 韩绍文, 姜如阳(3594)
- 基于长短期记忆网络和社区演化函数 魏东乐, 刘原泉, 张 楠(3599)
- 融合动态社交关系的图两度推荐算法 葛君伟, 岑 刚, 方义秋(3603)
- 基于可变损失和稀疏正则化的生成对抗网络 丁 露, 吕 洋(3607)
- 混合策略改进正弦余弦算法 林 杰, 何 庆(3612)
- 一种改进的鲸鱼优化算法 武泽权, 李永敏(3618)
- 混合策略改进的鲸鱼优化算法 郝晓弘, 宋吉祥, 周 强, 马 雷(3622)
- 一种新的直交模糊集距离及其在决策中的应用 符昌林, 茂菊红(3627)
- 多 agent 分层强化学习在数据推荐中的应用研究 洪壮壮, 王梓保, 张 尊, 黄兆华(3635)
- 基于框架表示学习的双谱图神经网络 陈运强, 曹学飞, 崔 军, 王瑞波, 李济洪, 李 凯(3640)
- 双层规划下考虑环境噪声的垃圾分拣中心选址研究 张 晨, 刘鹏明, 叶青明, 李冠林(3645)
- 基于多目标重力规则-柔性 AC 算法的机器人路径规划 谢金亮, 任海青, 吴彪理, 蒋欣欣, 刘成凯(3650)
- 基于 GA-GRASP 的多智能体强化学习的协作策略研究 郭长杰, 郑峻波, 崔中雷(3655)

系统应用开发

- 基于改进遗传算法的工程协同仿真测试用例生成方法 李文轩, 周文利, 和晓斌, 陈春所, 曹 勇(3662)

CPU+MCU异构系统中动态请求处理模型研究	袁国华,刘 颖,孙 东(3667)
3D-ACC:基于3D集成电路的卷积神经网络加速结构研究	王吉军,赵宇宇,李宏亮(3671)
Billed 包络谱的谱特征在驱动个体识别中的应用	伍嘉颖,郑娜娥,蒋春蔚(3677)
数据失调下基于WGAN和GAPCNN的轴承故障诊断研究	薛振涛,满君丰,彭 斌,廖 珂(3681)

信息安全技术

基于区块链智能合约的代币系统	方燕妮,周剑明(3686)
基于区块链的游戏软件产品版权保护研究	贺彪霞,智路平,魏海蕊(3691)
基于PI演算的Android App权限提升攻击检测	王 涛,马 川(3699)
基于LSTM-CNN的容器内恶意软件静态检测	金逸灵,陈兴蜀,王玉尧(3704)
基于信任评估的量子区块链网络匿名选举协议	邢 涛,吕 燕,张仕斌(3708)
针对去中心化基于属性指定证实者签名	张 健,黄俊杰,陈群山(3712)
基于软硬件实现的安全轻量S盒的设计	李 凡,张文英,邢朝晖(3717)
基于ZooKeeper的全网统一信任模型研究	史博轩,章 峰,蒋文保(3722)
基于改进型Heron映射的快速图像加密算法	赵洪祥,谢淑琴,邵建中,吴 刚(3726)
任意二粒子纠缠态的受控量子通信	彭家寅(3731)

网络与通信技术

一种基于叠加编码的协作传输协议	肖佳丽,曹雪虹(3736)
边缘计算中基于博弈论的数据协作缓存策略研究	王 刚,高 岭,高全力,牛秀娟,马景超(3739)
基于拓扑稳定性定向网络链路预测方法	李治成,古立新,刘树新,曹黎明(3744)
基于机器学习数据流突发变更服务功能结构策略	赵季红,李文群,苗 博,赵建龙,王 珂,吴豆豆(3749)
任务序列感知的大规模集群服务器控制模型	蔡文伟,朱嘉璐,张会兵(3753)
一种新的无线传感器网络干扰攻击检测技术	陈 涛,刘树德,梁修梁(3757)
基于软件定义网络的触觉互联网端到端系统	魏 萌,吕廷勤(3760)
WSN中基于中继选择的DH-MSMR-CARQ时延与吞吐量分析	周末强,代 雷,王 波,黎镇平(3764)
基于覆盖力的LoRa浮标网络覆盖优化算法	刘 侠,郑建道,吴华萍,张倩楠(3768)
基于软件定义网络的边缘控制部署机制	路 军(3773)

图形图像技术

结合空洞卷积的CNN实时微表情识别算法	魏松志,陈人和,钱育群(3777)
基于动态密集条件随机场增量推理计算的多类别视频分割	张松旭,卢先领,周洪钧(3781)
用于图像超分辨率的密集跳跃注意力连接网络	吴梁贵,蒋 平(3788)
基于邻域回口的医学图像超分辨率重建	端木春江,沈碧娟(3792)
基于聚类框架与局部感受野的实时人脸疲劳检测	刘科扬,王金凤(3795)
基于字典学习的卫星图像压缩算法研究	卢光武,崔晓霞,李 坤(3799)
遗传序列关系保持二值编码	王 刚,孙祖强,张龙波,刘萍萍(3803)
硬件优化的高清视频实时去雾算法	陆 斌,严利民,路志和(3807)
基于卷积神经网络和语义相关的协同显著性检测	张华迪,樊 坤,黄 睿(3811)
基于ResNeXt和WGAN网络的单图像超分辨率重建	曹庆亮,南方哲,尚通雅,孙 华(3815)
基于SPTNet的视频运动放大方法	周 飞,林振龙(3820)
引入多尺度特征融合的人脸关键点检测网络	齐国强,姚剑敏,胡海龙,严 群,林志贤(3825)
学习小波超分辨率系数的人脸超分辨率	刘 颖,孙定华,公衍超(3830)
基于新的视觉沙漏网络的人脸对齐	邢 菲,邢耀明,高 榕,王章校(3836)

信息集萃

下期要目	(3670)
------------	--------

期刊基本参数:CN51-1196/TP=1984-m* A4+320*di+P* ¥50.00+5840=66+2020-12=11

本期责任编辑 包 帆,黄 莉

基于聚类框架与局部感受野的实时人脸疲劳检测*

刘君扬, 王金凤[†]

(华南农业大学 数学与信息学院, 广州 510642)

摘要: 针对在自然环境下人脸疲劳识别迟到的问题, 如人脸检测率不高、识别疲劳的特征过于单一、检测速度很慢, 提出了一种基于聚类框架与局部感受野的实时人脸疲劳检测方法。首先对人脸尺寸进行聚类分析, 根据聚类测试集检测各个簇设置检测框大小, 根据预测特征图的感受野与人脸尺寸匹配的原则设置网络层数, 最后通过最小化损失函数学习多种疲劳特征。实验证明, 在驾驶室等环境下基于聚类框架与局部感受野的方法在保证识别准确率的同时提高了检测速度, 使用 GPU GeForce GTX TITAN 能达到 125 fps, 满足了实时性要求。

关键词: 神经网络; 深度学习; 目标检测; 疲劳识别; 感受野; 聚类

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1001-3695(2020)12-057-09

doi:10.19734/j.issn.1001-3695.2019.07.0315

Real-time face fatigue detection based on receptive field and clustering

Liu Junyang, Wang Jinfeng[†]

(College of Mathematics & Information, South China Agricultural University, Guangzhou 510642, China)

Abstract: Faced with the problems encountered in fatigue detection under natural environment, such as low detection rate of the face, slow detection speed and single feature for judging, etc., this paper proposed a fatigue detection method combining receptive field with clustering algorithm. Firstly, it applied cluster on the face sizes and returned the cluster numbers which determined the number of detection layer. Then, it set the size of anchor boxes according to the face size. Secondly, the algorithm set the number of convolutional layers according to the principle that the receptive field should match the face size in the predicted feature map. Finally, new algorithm learnt a variety of fatigue characteristics by minimizing the loss function. Experiments show that this method based on clustering framework and local receptive field have improved the detection speed while maintaining the recognition accuracy. It can reach 125 fps by using GPU GeForce GTX TITAN, and satisfy the request of real time.

Key words: neural network; deep learning; object detection; fatigue identification; receptive field; cluster

疲劳识别在道路交通领域是一个重点研究方向,倘若能够及时检测到驾驶员的疲劳状态则可以避免很多交通事故。徐礼祺等人^[1]提出的基于瞬时心电信号的疲劳检测识别算法以及祝荣欣等人^[2]提出的基于心电图的融合驾驶员疲劳检测均表明人的疲劳状态可以从生理测量、心电信号等基于医学设备的方法获得,但是这种方法需要专业设备,在实际生活中布置的可能性较小。第二种是基于机器视觉的疲劳识别方法,传统的疲劳识别方法一般是通过 Haar-like 特征检测出人脸的位置,然后通过计算眼睛纵横比来描述眼睛的张开程度判断疲劳状态^[3]。王莉^[4]提出基于肤色分割算法检测人脸并降低算法复杂度,文献[5]用 MT-CNN^[6]进一步提高人脸检测效率。谢冰凌等人^[7]通过计算眼睑和瞳孔区域像素个数占眼部区域像素总个数的比值和嘴部区域的宽高比,分别判断眼睛和嘴巴的关闭状态,从而提取出 FEROS 特征,跟踪眼率和哈欠频率等面部疲劳特征,通过疲劳特征可以进一步判断驾驶员的疲劳状态,但算法没有将特征关联。洪志阳等人^[8]对眼睛和嘴巴进行定位和状态分析,在决策阶段采用信息融合的方法对疲劳状态进行判断,但实时性不足。综上,疲劳识别存在两个问题需要解决,即检测速度慢,单一特征的识别结果并不可靠。

近年来由于计算机硬件的进步以及卷积神经网络的发展,图像识别方面的研究取得了很大突破。在 2012 年的 ImageNet 图像识别竞赛中,Hisao 带领的小组采用深度学习模型

AlexNet^[9]摘得桂冠。至 2014 年,Facebook 基于深度学习网络的 DeepFace^[10]项目更是在人脸识别精度上达到了 97% 的精确率。与此同时,Gasbick 等人^[11]提出了基于卷积神经网络的目标检测模型 B-CNN,随后他本人对 B-CNN 进行改进,分别提出 Fast B-CNN^[12]和 Faster B-CNN^[13],加快了检测速度。针对单一特征识别疲劳不可靠,本文将疲劳识别转换为二分类问题,制作数据集时标注人脸部分,通过监督学习得到表征疲劳的各种特征。基于深度学习的目标检测算法和传统图像处理算法相比,其检测的召回率提高了许多,但是检测速度有所下降。针对检测速度慢,本文提出采用 one-stage 目标检测算法进行疲劳识别,并在此基础上结合感受野和 CW^[14]聚类算法,提出聚类检测框架 (detection with cluster)。经过在自建数据集 CASIA-FaceV5^[15]和 YawoID^[16]上的测试,识别准确率达到 96% 以上,使用 GPU GeForce TITAN 能达到 125 fps,满足了实时性的需求。

1 目标检测算法概述

目标检测算法分为 two-stage 和 one-stage 两种。two-stage 检测算法分为两个阶段,首先产生候选框 (region proposal), 然后对候选框分类,这类算法的典型代表是基于 region proposal 的 R-CNN 系列算法,如 R-CNN、Faster R-CNN 等。

传统的 two-stage 检测算法流程如图 1 所示。输入图片经过卷积层 (conv) 和池化层 (pooling) 得到特征图 (feature map);

收稿日期: 2019-07-19; **修回日期:** 2019-09-08 **基金项目:** 广东省科技计划项目 (2017A040406023); 广州市科技计划项目 (20180404053)

作者简介: 刘君扬 (1993-), 男, 广东佛山人, 硕士研究生, 主要研究方向为计算机视觉; 王金凤 (1974-), 女, 通信作者, 河北黄骅人, 硕士, 主要研究方向为机器学习, 邮箱祝荣欣 (wangjinfeng@163.com)。

然后通过 RPN^[1] 产生候选区域, 同时将区域分为背景和目标两种类别, 并且对目标的位置进行初步预测。ROI pooling^[12] 把候选区域从特征图对应位置裁剪出来统一大小, 然后经过一个全连接层 (fc) 得到特征向量; 最后通过分类和回归两个分支实现对候选目标类别和位置的判定, 而 one-stage 检测算法不需要 region proposal 阶段, 直接产生目标类别概率和位置坐标, 经过一次检测得到最终的检测结果。

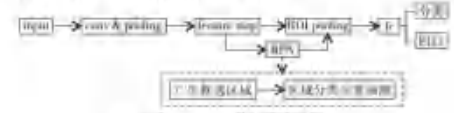


图1 two-stage检测算法流程
Fig.1 Two-stage detection algorithm flow

2 人脸定位与疲劳识别方法

2.1 基于感受野的特征提取网络

卷积神经网络在图像分类任务中取得很好的成绩, 主要得益于经过多个卷积层后网络能够提取丰富的特征。相比于多层感知机具有权值共享的特点, 卷积神经网络参数和计算量大大减少, 除此之外, 卷积层采用滑动窗口的方式对输入层进行扫描, 并没有将输入层进行向量化, 保留了输入层的几何特性, 得到的特征图也更有意义。

在疲劳识别场景中, 除了要求有较高的疲劳识别准确率, 还需要有较快的检测速度, 因为只有实时检测才能达到预警的作用。现有的大多卷积神经网络通过加深网络的层数或者一度来提高网络的识别准确率, 然而随着网络结构变得复杂, 网络的计算量也会增加。本文提出基于感受野设计特征提取网络, 能有效提取局部特征并且减少网络层数, 如图2所示。图中, conv3×3表示步长为1的3×3卷积层, conv3×3/2表示步长为2的3×3卷积层, 经过该层后特征图尺寸为原来的二分之一, conv1×1表示步长为1的1×1卷积层, 将特征图不变的计算称为一个block, 在每个block之间会进行下采样, 目的在于减少网络参数达到压缩网络的目的。本文没有采用传统的池化方法进行下采样, 因为使用卷积层进行下采样可以加强网络的学习能力, 同时利用感受野设计特征网络, 从而达到优化目的。

$$r' = n \times r \quad (1)$$

$$r' = n \times (k - 1) + k \quad (2)$$

式(1)和(2)分别用于计算网络最终特征图感受野和步长。特征提取网络经过5个block后特征图的尺寸为19×19×512。特征图在第一次卷积后的感受野为3×3, 由于底层卷积感受野很小, 网络只能学习到线条等低级特征。继续卷积操作, 每进行一次3×3的卷积, 特征图感受野就会增大。本文提出的特征提取网络共进行了10次3×3的卷积, 最终得到的特征图感受野为145×145, 总步长为32。网络结构应该适配识别的对象, 疲劳识别需要结合人脸的特征, 所以网络最终的感受野应该符合人脸大小, 如图3所示。

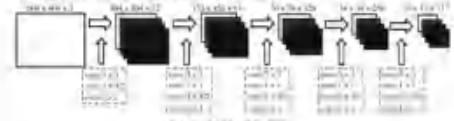


图2 特征提取网络
Fig.2 Feature extraction network

2.2 聚类检测框架

传统的人脸检测算法在复杂的场景下鲁棒性较差, 例如AdaBoost^[13], 而深度学习在这方面有较大的优势。复杂背景

会影响算法的准确率, 需要先对人脸进行裁剪然后再进行检测。DWC (detector with cluster) 指的是在设置先验框前先对数据集的人脸进行 CW (Chinese whisker) 聚类分析, 在得到人脸尺寸的分类数后根据检测对象选择检测层, 最后再设置先验框大小。CW 聚类算法可以对未知分类数量的样本查找类别个数并进行快速聚类。CW 算法在初始化阶段以每个节点为一个类别构建无向图, 不同节点之间计算相似度, 当相似度超过阈值就在两个节点间形成关联边, 权重为相似度。在迭代阶段首先随机选取一个节点 a , 在相邻的节点中选择权重最大的 b , 然后将 a 归为 b 的类别; 遍历所有节点, 重复迭代直到满足迭代次数。如图4所示, 标注信息包含人脸框的左上角坐标和长度, 将 a 和 b 组成二维向量 (a, b) 进行聚类。为了减少随机性, 本文对 (a, b) 进行了50次CW聚类, 聚类结果均为1, 训练数据的人脸尺寸相近, 只需要在一个特征值上进行功能。



图3 感受野示意图
Fig.3 Exemplar receptive field



图4 标注的样本
Fig.4 Annotated sample

本文借鉴了YOLO目标检测算法的思想, 并结合聚类检测框架来提高疲劳检测的实用性。YOLO检测算法没有 region proposal 阶段, 直接得到检测框的坐标, 检测框中含有检测对象的概率和检测对象类别的概率, 实现了检测和分类两个任务在同一个网络里完成, 最终的检测特征图维度数为

$$C = D \times (k_w + \text{conf} + c_b) \quad (3)$$

其中, k_w 取值为4, 表示用四个通道预测检测框位置, 四个通道预测的值分别是检测框的左上角坐标和长度; conf 取值为1, 表示用一个通道预测置信度, 这个通道用来预测检测框中含有人脸的概率; c_b 取值为2, 表示用两个通道预测类别, 在人脸疲劳识别中表示疲劳和非疲劳两类; D 为特征特征图中每个点所预测的检测框数目, detection with cluster 的编码网络结构如图5所示。

对于边界框的不同属性采取不同的损失函数, 将人脸定位任务和疲劳识别任务同时进行, 式(4)是网络总体损失。

$$\begin{aligned} \text{loss} = & 4 \times \sum_{i=1}^D \sum_{j=1}^C \left(\frac{1}{2} \left(|x_i - \hat{x}_j|^2 + |y_i - \hat{y}_j|^2 \right) + \right. \\ & 4 \times \sum_{i=1}^D \sum_{j=1}^C \left(\frac{1}{2} \left(|w_i - \hat{w}_j| + |h_i - \hat{h}_j| \right) + \right. \\ & \sum_{i=1}^D \sum_{j=1}^C \left(\frac{1}{2} \left(|p_i - \hat{p}_j| + |q_i - \hat{q}_j| \right) \right) + \\ & \sum_{i=1}^D \sum_{j=1}^C \left(\frac{1}{2} \left(|r_i - \hat{r}_j| + |s_i - \hat{s}_j| \right) \right) + \\ & \sum_{i=1}^D \sum_{j=1}^C \left(\frac{1}{2} \left(|t_i - \hat{t}_j| + |u_i - \hat{u}_j| \right) \right) + \\ & \sum_{i=1}^D \sum_{j=1}^C \left(\frac{1}{2} \left(|v_i - \hat{v}_j| + |w_i - \hat{w}_j| \right) \right) + \\ & \sum_{i=1}^D \sum_{j=1}^C \left(\frac{1}{2} \left(|x_i - \hat{x}_j|^2 + |y_i - \hat{y}_j|^2 \right) + \right. \\ & \left. \sum_{i=1}^D \sum_{j=1}^C \left(\frac{1}{2} \left(|w_i - \hat{w}_j| + |h_i - \hat{h}_j| \right) + \right. \right. \end{aligned} \quad (4)$$

其中, 第一和第二项计算人脸定位任务中的坐标误差; 第三和第四项计算的是物体置信度的损失; 最后一项计算疲劳识别任务的分类误差; 总损失特征图中的像素点称为 cell, 每个 cell 预测 N 个边界框; $\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i, \hat{p}_i, \hat{q}_i, \hat{r}_i, \hat{s}_i, \hat{t}_i, \hat{u}_i, \hat{v}_i, \hat{w}_i$ 分别对应于检测框中各个属性的估计, 由各项损失构成了最终的损失函数。

2.3 疲劳检测指标

类似于 FRCLUS, 本文提出 FOP (frequency of fatigue), 表示某段时间内检测到人脸疲劳的频数。

$$F_{OP} = \frac{n}{N} \times 100\% \quad (5)$$

其中, n 是某段时间内人脸疲劳的频数; N 是这段时间内的总帧数; L 值越大代表疲劳程度越大。当检测网络所确定一帧后就将其放到一个先进先出队列, 该队列长度是单位时间的总帧

数,每当队列添加一帧,队头就移除一帧,最后计算 ρ ,并与疲劳阈值比较,流程如图 6 所示。

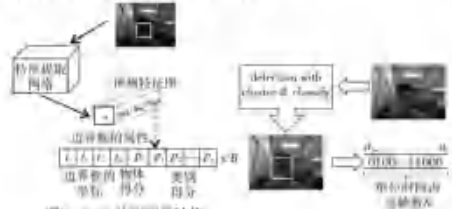


图5 DWL的端到端结构
Fig.5 End-to-end structure of detection with cluster

合以上信息,YOLOv3^[12]只保留了YOLOv3的大目标检测层。表2的结果表明YOLOv3^[12]虽然只有一个检测层,但是检测能力并没有下降,原因在于这个检测层是经过聚类分析得到的,检测层的anchor^[13]大小设置合理,从而验证了聚类检测框架的有效性。

表2 聚类检测对算法的影响
Tab.2 Influence of clustering detection on algorithm

算法	IoU 平均值	时间/ms
YOLOv3	0.881	55.937
YOLOv3 [*]	0.886	23.802

之后对 two-stage 和 one-stage 检测算法进行了对比实验,以验证本文方法的准确率和检测速度。其中 two-stage 算法选取了具有代表性的 Faster R-CNN 和 MTCNN + CNN,而 one-stage 算法包括 YOLOv3 和本文 DWL。

表3的结果表明,two-stage 的识别算法在分类准确率方面比 one-stage 检测算法高。Faster R-CNN 算法包含 RPN (region proposal network) 和 CNN 分类网络两部分。第一阶段 Faster R-CNN 先将数据送入 RPN (region proposal net),得到可能含有人脸的区域 (ROI),第二阶段再将 ROI 送入分类网络得出结果。MTCNN 是一个能快速检测人脸的算法,除了检测人脸之外还能检测眼睛和嘴巴。MTCNN + CNN 第一阶段用 MTCNN 检测人脸并裁剪,第二阶段将裁剪后的人脸输入 CNN 识别,这种方法能将人脸和背景分离,在几种对比算法中获得最高识别准确率。YOLOv3 属于 one-stage 算法,DarkNet53 是 YOLOv3 的特征提取网络,含有 53 个卷积层,一个大目标检测层,一个中目标检测层和一个小目标检测层,检测速度比 Faster R-CNN 快了三倍。本文算法在特征提取阶段基于感受野优化网络,在检测阶段基于 CW 聚类减少冗余,识别精度与 YOLOv3 相同,比分阶段的识别算法略低,检测速度比 YOLOv3 快 4.5 倍。原因在本文算法中网络的感受野更合理,通过聚类得到的检测层效率更高。

表3 one-stage 和 two-stage 算法的 mAP 和运行时间对比
Tab.3 Comparison of mAP and running time between one-stage and two-stage algorithm

算法	mAP/%	时间/s	算法	mAP/%	时间/s
Faster R-CNN	97.0	297.368	YOLOv3	96.2	55.937
MTCNN+CNN	99.8	87.321	DWC	96.2	8.101

最后,将本文算法与现有的几个根据五官特征判断疲劳的算法在自建数据集 CASIA-FACEVS 和 YauDD 数据集上进行对比,测试样本数均为 200 张图像。图 9 分别是这三个数据集的图像示例。其中,基于 ASM 的疲劳识别算法结合 Haar 级联特征实现对眼睛疲劳状态的识别;MTCNN + ROI + random ferns, MTCNN + MultiHOG + SVM, MTCNN + MSP-Net 三种算法均是先用 MTCNN 检测出眼睛和嘴巴,再根据眼睛和嘴巴的闭合判断疲劳状态;而本文算法则是综合了人脸的全部信息。



图9 测试数据集示例
Fig.9 Examples of test dataset

三个数据集上的对比结果如表 4~6 所示。结果表明本文算法具有更高的准确率和效率。在 CASIA-FACEVS 数据集中大多是非疲劳状态的样本,各个算法的精度并无太大波动。在

3 实验与分析

本文算法在训练阶段需要用到人脸疲劳状态的图片和正常状态的图片。训练数据由网上搜集和制作两部分组成。制作部分是从视频中获取的各种疲劳形态图片和非疲劳状态图片共 800 张,另外从网上搜集 200 张符合要求的图片,总共 1 000 张图片。实验采用的处理器为英特尔 Core i5-4460 @ 3.20 GHz,内存 RAM 16 GB,显存 12 GB,显卡 GeForce GTX TITAN,输入图像分辨率为 640 × 480,输入网络前会先将图片尺寸统一为 608 × 608。初始学习率设置为 0.100,迭代 30 000 次,在每次迭代开始前会通过改变图片的饱和度、曝光、色调产生新的训练图片。为了让算法学习到丰富的疲劳表达形式,训练集包含各种表征疲劳的行为,其中包括佩戴口罩和不佩戴口罩的正负状态以及在疲劳时会出现的打哈欠和打吨等面部表情。

实验先用 VGG2007^[14]和 VGG2012^[15]对特征提取网络预训练,再用训练集训练整个检测网络。表1的结果表明在训练数据不充分的情况下训练网络会造成严重的过拟合。因此,本文采用迁移学习^[16]可以加快网络收敛。图7展示的是经过迁移学习后训练的 loss 曲线变化图,可以看出初始的 loss 很低,迁移学习不仅加强了网络的泛化能力,而且加快了网络的收敛速度。

表1 迁移学习的作用
Tab.1 Effect of transfer learning

VGG 预训练	训练准确率	训练时间/s
是	98.0	13.5
否	98.2	280.0

本文为了满足疲劳识别的实用性,减少网络冗余,提出了聚类检测框架,根据对人脸尺寸聚类结果决定检测层数,分别对特征提取网络部分和目标检测部分进行了优化。图8是本文算法训练的 IoU 曲线,IoU 曲线表征算法的定位准确率。经过聚类的检测层定位准确率很高,IoU 基本在 0.8 以上。

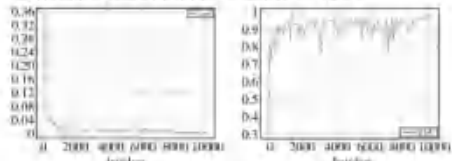


图7 迁移学习后的 loss 曲线
Fig.7 Curve of loss after transfer learning

图8 本文算法训练的 IoU 曲线
Fig.8 Curve of IoU trained by this algorithm

为了验证本文聚类检测框架的有效性,将 YOLOv3 与结合了聚类检测框架的 YOLOv3^{*} 作对比。

对训练样本中的人脸宽度 (m、k) 聚为 1 类后,分别对 m 和 k 求均值, m 的均值为 138.8 像素, k 的均值为 143.3 像素。数据集中 m 的最小值为 119 像素, k 的最小值为 119 像素。综

YaoDD 数据集中有说话和唱歌的样本,导致根据嘴巴闭合程度判断疲劳状态的算法精度有所下降,而李文算法仍然保持比较平稳的精度。

表 4 与其他疲劳识别算法在自建测试集上的对比

算法	mAP/%	时间/ms
基于 ASM 的疲劳识别 ^[4)]	92.5	45.03
MTCNN + IOU + random time ^[2)]	94.1	14.20
MTCNN + MultiDPG + SVM ^[17)]	94.4	48.97
MTCNN + MSP+S ₁ ^[15)]	95.3	11.002
DWL	95.2	8.101

表 5 与其他疲劳识别算法在 CASIA-FACEVS 数据集上的对比

算法	mAP/%	时间/ms
基于 ASM 的疲劳识别	92.2	45.76
MTCNN + IOU + random time	95.2	14.63
MTCNN + MultiDPG + SVM	94.4	48.32
MTCNN + MSP+S ₁	95.2	12.648
DWL	96.8	8.055

表 6 与其他疲劳识别算法在 YaoDD 数据集上的对比

算法	mAP/%	时间/ms
基于 ASM 的疲劳识别	91.5	46.09
MTCNN + IOU + random time	93.1	15.38
MTCNN + MultiDPG + SVM	93.3	49.77
MTCNN + MSP+S ₁	94.7	12.714
DWL	96.3	8.254

4 结束语

本文针对现阶段人脸疲劳识别算法的不足展开研究,相较于多数疲劳识别算法利用单特征判断疲劳,本文提出利用人脸综合信息进行判断,具体做法是将疲劳识别视为二分类问题,搭建深度学习网络学习人脸特征;6)经过分析疲劳识别数据的特点,提出基于感受野优化特征提取网络,基于 CW 聚类减少检测层的冗余,在保持识别准确率的同时降低了检测速度,满足了疲劳识别的实时性要求。本文算法综合整个人脸的特征进行疲劳识别,对于脸部有遮挡的情况识别度不高,例如检测的人脸带着墨镜或口罩。此外,由于每个人都会出现疲劳,疲劳的表现形式每个人是否都不一样,这些问题都值得进一步研究。

参考文献

[1] 李孔旺,任朝刚,陈宇行. 基于眼动心电信号的疲劳驾驶检测研究[J]. 东北大学学报:自然科学版,2019,40(7):937-941. (Li Kongwang, Ren Chaogang, Chen Yuxing. Fatigue driving detection algorithm based on electrooculogram EOG signals[J]. Journal of Northeastern University: Natural Science, 2019, 40(7):937-941.)

[2] 祝维欣,王全成. 基于心电比的融合人脸驾驶人疲劳检测研究[J]. 农机化研究,2020,42(12):8-14,43. (Zhu Weixin, Wang Quancheng. Research on driver fatigue detection of combine harvester based on myocardial electricity[J]. Journal of Agricultural Mechanization Research, 2020, 42(12):8-14, 43.)

[3] 谢志飞,任伊庭,徐星明. 基于人脸特征点分析的疲劳驾驶实时检测算法[J]. 微机技术,2018,42(12):27-30,55. (Xie Zhifei, Ren Yiting, Xu Xingming. Fatigue driving real-time detection method based on face feature point analysis[J]. Video Engineering, 2018, 42(12):27-30, 55.)

[4] 王舒. 基于人工智能的疲劳驾驶检测中人脸检测技术研究[J]. 科技视界, 2018(25):49-50. (Wang Shu. Research on face detection technology in fatigue driving detection based on artificial intelligence[J]. Science and Technology View, 2018(25):49-50.)

[5] Ji Yingen, Wang Shiguang, Zhao Yan, et al. Fatigue-state detection based on multi-scale feature and time recognition network[J]. IEEE

Access, 2019, 7:64136-64147.

[6] Zhang Kaiqiang, Zhang Zhanpeng, Li Zhifeng, et al. Joint face detection and alignment using multi-task cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10):1499-1503.

[7] 靳永波, 乐爽, 陈子银. 基于视觉的高级驾驶员特征提取方法[J]. 计算机技术与发展, 2018, 28(11):193-197. (Jin Yongbo, Le Shuang, Chen Ziyin. Vision-based driver fatigue feature extraction method[J]. Computer Technology and Development, 2018, 28(11):193-197.)

[8] 谢志川, 王瑞飞, 徐志雄, 等. 基于机器视觉的驾驶员疲劳检测算法[J]. 物联网技术, 2018, 8(7):78-79. (Xie Zhichuan, Wang Ruifei, Xu Zhixiong, et al. Machine vision based driver fatigue detection method[J]. Internet of Things Technologies, 2018, 8(7):78-79.)

[9] Kihiravsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communication of the ACM, 2012, 60(2):84-90.

[10] Tangman Y, Yang Ming, Kanazaki M, et al. DeepFace: closing the gap to human-level performance in face verification[C]. Proc of IEEE Conference on Computer Vision Pattern Recognition, Piscataway, NJ: IEEE Press, 2014:1701-1708.

[11] Girdhar R, Dunham J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proc of IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, NJ: IEEE Press, 2014:580-587.

[12] Girdhar R. Fast R-CNN[C]. Proc. of IEEE International Conference on Computer Vision, Piscataway, NJ: IEEE Press, 2015:1440-1448.

[13] Ren Shuangping, He Kaiyang, Girdhar R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137-1149.

[14] Benam C. Chinese whispers: an efficient graph clustering algorithm and its application to natural language processing problems[C]. Proc of the 1st Workshop on Graph Based Methods for Natural Language Processing, Stroudsburg, PA: Association for Computational Linguistics, 2003:73-80.

[15] Song Feiyu, Fan Xiuming, Liu Xun, et al. Eye/closedness detection from still images with multi-scale histograms of grayscale oriented gradients[J]. Pattern Recognition, 2014, 47(9):2825-2838.

[16] Alaluf S, Omryguez M, Shimulomnani S, et al. YaoDD: a yawning detection dataset[C]. Proc of ACM Multimedia Systems Conference, New York: ACM Press, 2014:19-21.

[17] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]. Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Piscataway, NJ: IEEE Press, 2001:511-518.

[18] Everingham M, Winn J. The PASCAL visual object classes challenge 2007 (VOC2007) development kit[J]. International Journal of Computer Vision, 2006, 111(1):98-136.

[19] Yosinski J, Clune J, Bang S, et al. How transferable are features in deep neural networks? [EB/OL]. (2014-11-06). <https://arxiv.org/abs/1411.1792>.

[20] 司哲, 姚旭光, 蔡杨, 等. 基于 ASM 的驾驶员面部疲劳状态识别方法[J]. 计算机工程与设计, 2018, 39(10):3240-3245. (Si Zhe, Yao Xuguang, Cai Yang, et al. ASM-based driver facial fatigue state recognition method[J]. Computer Engineering and Design, 2018, 39(10):3240-3245.)

[21] Dora Yascha, Zhang Yan, Yao Jiaqiang, et al. Comparison of random forest, random ferns and support vector machine for eye state classification[J]. Multimedia Tools and Applications, 2016, 75(19):11763-11783.

[22] Bodden J, Divvala S, Girdhar R, et al. You only look once: unified, real-time object detection[C]. Proc of IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, NJ: IEEE Press, 2016:779-788.

[23] Lin Y Y, Maire M, Belongie S, et al. Microsoft coco: common objects in context[C]. Proc of European Conference on Computer Vision, Berlin: Springer, 2014:740-755.

3.6. 融入平滑组稀疏化的脑部 MRI 图像分类



中国图象
图形学报

2022
03
VOL.27

ISSN1006-8961
CN11-3758/TB



中国图象图形学报

刊名题字: 宋健 月刊 (1996年创刊)



第27卷第3期 / 总第211期
2022年3月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿,均视为同意在本刊网站及CNKI等全文数据库出版,所刊论文已获得著作权人的授权。本刊所有图片均为非商业目的使用,所有内容,未经许可,不得复制或其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjg.cn

广告发布登记号 京朝工商广登字20170218号

总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱:北京399信箱 邮编:100049)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong

Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics

Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China

Zip code 100190

E-mail jig@aircas.ac.cn

Telephone 010-58887035

Website www.cjg.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

ISSN 1006-8961

CODEN ZTTXFZ

国外发行代号 M1406

国内邮发代号 82-831

国内定价 60.00元

中图分类号: TP181 文献标识码: A 文章编号: 1006-8961(2022)03-0885-13
论文引用格式: Huang S H and Wang J F. 2022. Brain MRI image classification incorporating smooth group sparseness. *Journal of Image and Graphics*, 27(03): 0885-0897. (黄帅辉, 王金凤. 2022. 融入平滑组稀疏化的脑部 MRI 图像分类. *中国图象图形学报*, 27(03): 0885-0897. [DOI: 10.11834/jcg. 210367])

融入平滑组稀疏化的脑部 MRI 图像分类

黄帅辉, 王金凤*

华南农业大学数学与信息学院, 广州 510642

摘要: 目的 阿尔茨海默症 (Alzheimer's disease, AD) 是主要的老年病之一, 并正向年轻化发展。早期通过核磁共振 (magnetic resonance imaging, MRI) 图像识别 AD 的发病阶段, 有助于在 AD 初期及时采取相关干预措施和治疗手段, 控制和延缓 AD 疾病恶化。为此, 提出了基于平滑函数的组 L1/2 稀疏正则化 (smooth group L1/2, SGL1/2) 方法。方法 通过引入平滑组 L1/2 正则化实现组内稀疏, 并将原先组 L1/2 方法中含有的非平滑的绝对值函数向平滑函数逼近, 解决了组 L1/2 方法中数值计算振荡和收敛难的缺点。SGL1/2 方法能够在保持分类精度的前提下, 加速对模型的求解。同时在分类方法中, 引入一个校准 hinge 函数 (calibrated hinge, Clange) 代替标准支持向量机 (support vector machine, SVM) 中的 hinge 函数, 形成校准 SVM (calibrated SVM, C-SVM) 用于疾病的分类, 使处于分类平面附近的样本更倾向于分类的正确一侧, 对一些难以区分的样本能够进行更好的分类。结果 与其他组级正则化的正则化方法相比, SGL1/2 与校准支持向量机结合的分类模型对 AD 的识别具有更高的分类性能, 分类准确率高达 94.70%。结论 本文提出的组稀疏分类模型, 实现了组间稀疏和组内稀疏的优点, 为未来 AD 的自动诊断提供了客观参照。

关键词: 阿尔茨海默症 (AD); 组 L1/2 稀疏正则化; 校准支持向量机 (C-SVM); 结构化磁共振; 组间稀疏; 组内稀疏

Brain MRI image classification incorporating smooth group sparseness

Huang Shuaihui, Wang Jinfeng*

College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China

Abstract: Objective Alzheimer's disease (AD) is one of the most common irreversible neurodegenerative diseases. The main clinical manifestation is cognitive impairment, which usually occurs in the middle-aged and elderly stage. The diagnosis and recognition of AD is constrained of the inefficient experience and mis-diagnosis of clinicians. If the onset stage of AD can be targeted in early diagnosis, it will yield relevant interference and treatment measures to the early stage of AD. A high accuracy automatic detection algorithm will help clinicians to prevent and treat AD. **Method** Structured magnetic resonance imaging (S-MRI) is used to extract gray matter density characteristics of each subject after pre-treatment. A group L1/2 sparsity method based on smoothing function (SGL1/2) is illustrated. This smooth grouping sparsity method is based

收稿日期: 2021-06-01; 修回日期: 2021-09-26; 预印本日期: 2021-10-08

* 通信作者: 王金凤, wangjinfeng@scau.edu.cn

基金项目: 广州市智慧农业重点实验室建设项目 (201902010081); 广东省科技计划项目 (2017A040408023); 广州市科技计划项目 (201804010353)

Supported by: Guangzhou Key Laboratory of Intelligent Agriculture (201902010081); Science and Technology Planning Project of Guangzhou Province (2017A040408023); Science and Technology Planning Project of Guangzhou City (201804010353)

on the common group $L1/2$ (GL1/2) method. GL1/2 method can find Group Lasso method to select high correlation feature group without sparse the selected feature group, so it can remove some redundant features in redundant group and surviving group simultaneously. However, this group sparsity method has the difficulties of oscillation and convergence because it contains non-smooth absolute value function. The selected SGL1/2 method uses a smooth function to approximate the non-smooth absolute value function of the traditional GL1/2 method. The sparse model becomes a smooth model, which can quickly converge to the optimal value. Meanwhile, these regularization benchmark need to group features in advance at the group scale. In order to obtain scientific grouping of human brain features, a registered anatomical automatic labeling (AAL) template, which registers the original AAL template into the standard template, so that the obtained AAL template has the same spatial distribution as each pre-processed targeting image. Based on this grouping template, each region is regarded as a group. All voxels in each region are targeted as the features of each group for group sparsity. In respect of classification method, a calibrated hinge is opted to substitute the hinge loss function in SVM (calibrated SVM, C-SVM). C-SVM can make the points near the classification plane more inclined to the correct side of the classification. We select four classification methods, which are support vector machine (SVM), calibrated support vector machine (C-SVM), linear classification and logistic regression. C-SVM uses a calibrated hinge loss function to replace the hinge loss function of traditional support vector machine, and then classifies the features of AD and cognitive normal (CN) control group via a simple classification function. The analyzed results illustrates that the classification accuracy of C-SVM reaches 91.06%. We select C-SVM as the benchmark, and the SGL1/2 group sparse method is integrated based on registered AAL template to form the main classification method. **Result** The key demonstration is selected from the "ADNI: Complete 2Yr 1.5T" dataset of Alzheimer's disease neuroimaging initiative (ADNI) for training and testing, and compared the classification performances with the classification model combined with GL1/2 and SGL1/2 group level regularization methods. The validation support vector machine classification model has a good classification effect on the recognition of AD based on SGL1/2 group sparsity method. In the control group of AD and CN control group, the classification accuracy reaches 94.70%. In order to show the generalization of the model, we also selected Golang data sets from relevant for classification literatures. The results also confirmed that the illustrated model had qualified classification effect. In the control group of AD and CN, the accuracy rate reached 91.97%. In terms of the features of each group segmented by the registered AAL group template, the C-SVM model is used to classify, and the classification effect of each group region is obtained. We select six regions with the best classification effect, and compare the results of the targeted illustrated model. **Conclusion** The demonstrated SGL1/2 group sparse classification model is based on AAL grouping template proposed in this research. The priorities of inter group sparse and intra group sparse are involved. The high correlation features of high correlation group are selected in all human brain regions, and realize the accurate location of high correlation brain regions with AD, which provides a benchmark for the automatic diagnosis of AD further.

Key words: Alzheimer's disease (AD); smooth group $L1/2$; calibrated support vector machine (C-SVM); structural magnetic resonance imaging; inter group sparsity; intra group sparsity

0 引言

阿尔茨海默症 (Alzheimer's disease, AD) (Mattison, 2004) 是一种由于大脑神经细胞死亡造成的不可逆的神经退行性疾病, 在临床上主要体现为智能损害。针对受试者是否患病或病情程度的不同, 主要分为认知正常 (cognitive normal, CN), 轻度认知功能障碍 (mild cognitive impairment, MCI) 和 AD 3 种。患者在 AD 的治疗过程中使用的一些治疗 AD 的药物只能在某种程度上延缓病情的恶化。因此, 实现

AD 早期诊断, 对 AD 的诊断与预防都有着至关重要的作用。现阶段由于没有系统化的诊断方式, 医生临床对 AD 诊断和后续治疗经常采用观察法, 但受人为影响较大, 不仅效率低下, 还容易造成误判 (Sakkalis, 2011)。因此, 智能化区分 AD 发展阶段是帮助临床医生预防和治疗 AD 的重要步骤。

核磁共振成像 (magnetic resonance imaging, MRI) (Frisoni 等, 2010; Feng 等, 2021; Lee 等, 2019) 是一种常见的生物标志物, 具有无创性、可用性以及对疾病发作后大脑变化的高度敏感性, 并具有高空间分辨率的脑形态可视化能力等一系列优点 (Liu,

2000), 是研究 AD 引起的各种脑结构和形态学变化的理想工具, 通常用于诊断 AD 的标准临床评估(陈弘扬等, 2021)。MRI 分为结构 MRI (structural MRI, s-MRI) 和功能 MRI (functional MR, f-MRI)。本文研究 s-MRI。对 MRI 而言, 能够获取的图像数量通常远小于图像的特征维度, 并且图像中的许多特征与学习任务并不相关, 即具有小样本高维度的特性。对此, 如果先验知识不足, 可以考虑对大脑全体结构(de Magalhães-Oliveira等, 2010; Liu等, 2013)的特征进行分析, 这种方法通过将每个结构的描述压缩成标量或低维表示, 并忽略其结构内部的详细信息, 从而获取关于整体的特征。此外如果可以结合一些标准的脑部模板进行配准, 将每个受试者图像映射到同一标准空间中, 再分别提取目标特征, 将大幅提升结构内的相关性, 以获取更为合理的特征分布。这些标准大脑模板通常将大脑分为若干个区域, 如 116 个区域的自动解剖标准模板(anatomical automatic labeling, AAL)(Tsarou-Mazoyer等, 2002)。

选择所有体素特征进行分类不是一个理想的选择。通过特征选择步骤(Fung和Stoeckel, 2007)或稀疏诱导范数(时中荣等, 2017)降低特征数量和维度, 不仅能够提高分类精度, 还能产生更具临床意义的结果。

组级上的范数针对特征进行分组, 以特征组为单位进行操作, 进行特征组间的稀疏和选取相关特征组。传统组稀疏正则化方法(group lasso, GL)是以组为级别的一种稀疏方法(Yuan和Liu, 2006; Meier等, 2008), 可以针对每个解剖组进行稀疏, 排除一些冗余的组, 但不能对组内特征进行稀疏, 因此无法去除组内一些冗余特征。针对此问题, 组 L1/2 稀疏正则化(group L1/2, GL1/2)(Liu等, 2016)通过将 GL 中组内的非稀疏平方范数换成能实现稀疏效果的 L1/2 范数进行解决, 组 L1/2 稀疏正则化不仅可以去除冗余的组, 还可以对选取组的组内特征再次进行稀疏, 从而选取更为准确的特征, 以提升预测准确性。但这种方法仍有局限性, 由于包含一个非平滑的绝对值函数, 在数值计算中将会引起振荡, 很难收敛。

本文使用平滑的组 L1/2 稀疏正则化(smooth group L1/2, SGL1/2)(Li等, 2018; Alemu等, 2019; Fan等, 2020)方法, 将原先组 L1/2 方法中含有的非平滑的绝对值函数向平滑函数逼近, 解决组 L1/2 方

法中数值计算振荡和收敛难的缺点。在本研究中, 组级别使用经过与标准模板配准过的 AAL116 模板的各结构区域作为分组模板, 其配准流程如图 1 所示, 然后提取每个结构区域的体素作为各自的组, 在此基础上实现对组级上的正则化效果。本文将配准后的 AAL 模板分成的每个区域各当做一组, 每组中含有该组中所有的体素特征。同时通过 SGL1/2 结合 C-SVM 对每组的特征进行选择 and 分类, 将原先 GL1/2 方法中含有的非平滑的绝对值函数向平滑函数逼近, 解决 GL1/2 方法中数值计算振荡和收敛难的缺点。

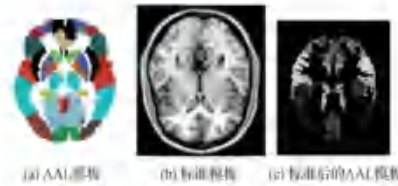


图1 AAL模板标准化流程图

Fig.1 AAL template standardization process

(a) AAL template map; (b) standard template map; (c) standard AAL template map

1 基于组稀疏方法的分类模型

本文对 GL, GL1/2 和 SGL1/2 这 3 种组级上的正则化方法进行研究。

1.1 基于组稀疏化方法的特征提取

GL 主要处理的是组间的稀疏关系, 而对组内没有稀疏关系。GL 可以用于去除一些不相关组, 以达到减少数据特征维度, 提升分类性能的目的。GL 的函数定义为

$$GL(\mathbf{x}) = \sum_{g=1}^G \beta_g \|\mathbf{x}_g\|_1 \quad (1)$$

式中, g 表示第 g 个组, β_g 表示第 g 个组的组权重, \mathbf{x}_g 表示第 g 个组的特征。可以看到 GL 可以用于对组间的稀疏, 但不能对幸存组的组内特征进行稀疏。同时, 它的导函数可以近似为

$$\rho_{GL} = \max\left(0, \left(1 - \frac{\lambda \beta_g \alpha}{\|\mathbf{x}_g\|_1}\right)\lambda\right) \quad (2)$$

式中, λ 是正则化系数, β 是组权重, α 为近似梯度的步长。

GLI/2 范数有利于解决 GL 只能对组间进行稀疏而不能对组内特征稀疏的问题,它将 GL 中组间的结合方式从 L1 方法改变为一个 LI/2 方法,而组内从 L2 方法改变为一个 LI 方法。从而对组内特征能够进行稀疏选择且兼顾防止过拟合的效果。因此 GLI/2 既能去除冗余的特征组,又能够对幸存组的组内特征进行稀疏选择,以得到强相关组的强相关特征,最后经过分类模型训练,得到更好的分类精度。GLI/2 的函数定义为

$$GLI_2(x) = \sum_{\sigma=1}^G \beta_{\sigma} \sqrt{\|x_{\sigma}\|_1} \quad (3)$$

其导函数可近似为

$$\rho_{GLI_2} = \max\left(0, \left(1 - \frac{\lambda \beta_{\sigma} \alpha}{2 \|x_{\sigma}\|_1^2}\right) \text{sign}(x_{\sigma})\right) \quad (4)$$

式中, β_{σ} 表示不同组之间的权重,用于权衡每个组对分类的影响, sign 为符号函数。

SGLI/2 范数是对 GLI/2 的一种平滑逼近,虽然 GLI/2 不仅能去除冗余的特征组,还能够对幸存组的组内特征进行稀疏选择,但它与 L1/2 正则化类似,其定义为

$$\|x\|_{\frac{1}{2}} = \sum_{\sigma=1}^G \sqrt{|x_{\sigma}|} \quad (5)$$

由此可见,GLI/2 的缺点是涉及一个非光滑的绝对值函数,导致数值计算出现振荡,收敛分析困难。SGLI/2 提出对 GLI/2 中非光滑的绝对值函数进行平滑逼近,使得能够更快收敛和计算平稳。

SGLI/2 的函数定义为

$$SGLI_2(x) = \sum_{\sigma=1}^G \beta_{\sigma} \sqrt{f(|x_{\sigma}|)} \quad (6)$$

SGLI/2 的导函数变为

$$\rho_{SGLI_2} = \max\left(0, \left(1 - \frac{\lambda \beta_{\sigma} \alpha}{2 f(|x_{\sigma}|)}\right) f'(|x_{\sigma}|)\right) \quad (7)$$

式中, $f'(|x_{\sigma}|)$ 为 $f(|x_{\sigma}|)$ 的导函数。SGLI/2 提出将 GLI/2 中的 $|x|$ 用一个平滑函数 $f(x)$ 替代,其定义为

$$f(x) = \begin{cases} |x| & |x| \geq c \\ -\frac{1}{8c^2}x^3 + \frac{3}{4c}x^2 + \frac{3}{8}c & |x| < c \end{cases} \quad (8)$$

式中, c 为一个很小的正数, $f(x) \in \left[\frac{3}{8}c, +\infty\right]$, $f'(x) \in [-1, 1]$, $f''(x) \in \left[0, \frac{3}{2c}\right]$ 。

本文中特征提取均根据配准后 AAL 模板划分的脑区组进行提取,将获取各组的灰质密度体素作为各组的特征,如图 2 所示,受试者的大脑图像经过预处理后获得的灰质图像也与标准模板进行配准,以保证 AAL 分组模板和受试者大脑存在相同空间分布。将同样空间分布的 AAL 模板和受试者灰质图像进行定位,对所有受试者图像都在相同定位区域提取各脑区特征,这种方法可以很好地降低受试者个体的差异性,保证研究中提取的特征实用可靠。

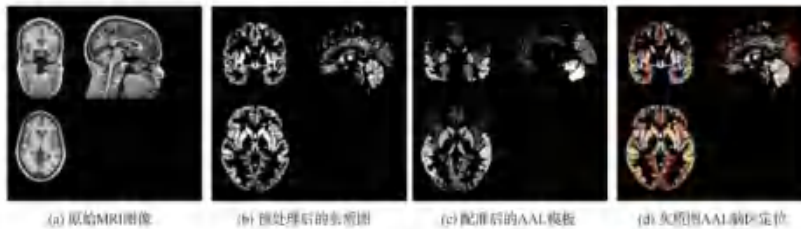


图2 图像预处理和各脑区组特征提取过程图

Fig. 2 Image preprocessing and feature extraction of brain regions (a) original MRI map; (b) preprocessed gray matter map; (c) registered AAL template; (d) localization of AAL brain regions on gray matter map

1.2 分类模型建立

支持向量机 (support vector machine, SVM) (Cortes 和 Vapnik, 1995) 是一种通过监督学习方式

对数据进行二元分类的广义线性分类器,其决策边界是对学习样本求解的最大边距超平面。支持向量机采用损失函数与最大边距项进行结合,通过求解

最小化问题来计算最优权重 w_{opt} 和偏置参数 b_{opt} , 具体为

$$\{w_{opt}, b_{opt}\} = \operatorname{argmin} \frac{1}{n} \sum_{i=1}^n \mathcal{L}_{\text{hinge}}(u)^c + \frac{1}{2} \lambda_1 \|w\|_2^2 \quad (9)$$

$$u = y_i (\langle w, x_i \rangle + b) \quad (10)$$

式中, $\|w\|_2^2$ 表示最大边距惩罚项, 它能够更好地进行分类的划分, 大边距惩罚项假设每个特征是独立的, λ_1 是一个非负的参数, 表示最大边距惩罚。权重 w 的维数与特征向量 x_i 的维数相同, y_i 是该疾病真实标签, b 为偏差。传统的基于铰链损失函数的 SVM 模型具有良好的分类性能, 但对于两类间边界较模糊的情况, 仍有待提高。

采用基于一种校准 hinge (calibrated hinge, Calinge) 损失函数的 SVM (calibrated SVM, C-SVM) 作为终端分类器, 在特征提取模块的基础上完成 AD 的分类识别。模型构建流程如图 3 所示, C-SVM 采用 Calinge 作为损失函数, 具体为

$$\mathcal{L}_{\text{calinge}}(u) = \mathcal{L}_{\text{hinge}}(u) = 1 - \frac{1 + \max(0, u)}{2 + \operatorname{abs}(u)} \quad (11)$$



图3 主要分类流程图

Fig. 3 Main flow-process diagram

Clinge 函数分布如图 4 所示, 当 $u > 0$ 时, u 越大, Clinge 损失越小; 当 $u \leq 0$ 时, u 越小, Clinge 损失越大。因此这个校准损失函数可以给分类超平面处于预测正确的一侧更小的损失, 给分类超平面处于预测错误的一侧更大的损失, 它从侧面向分类平面输出更大的损失。从而确保靠近超平面的样本的预测结果倾向于分布在正确的一侧。

此外, 对于单纯 SVM 模型来说, 权重是一个系数向量, 其中大部分是非零项 (Sun 等, 2018)。本研究选取的数据集是一个高维小样本数据, 因此选择

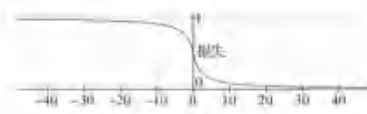


图4 Clinge 损失函数分布图

Fig. 4 The distribution chart of Clinge loss function

所有体素特征进行分类是一个不太理想的选择, 特征数量过多, 维度过大对程序运行有极大影响, 而且对分类效果没有多大提升, 甚至可能降低。因此, 为了减少样本特征数量, 降低运算成本, 本文采用基于校准的 AAL116 模板作为分组模板, 以大脑区域的灰质密度体素作为特征, 结合 SGL1/2 正则化的 C-SVM 分类模型, 这种模型的总代价函数定义为

$$f_{\text{cal}} = \frac{1}{n} \mathcal{L}_{\text{calinge}}(y_i (\langle w, x_i \rangle + b)) + \frac{1}{2} \lambda_1 \|w\|_2^2 + \lambda_2 \sum_{i=1}^c B_i \|f(w_i)\|_1 \quad (12)$$

该模型结合了校准 SVM 和 SGL1/2 的各自优点, 能够在保证分类效果的同时, 找到大脑特征关键区域中的关键体素, 具有很好的可解释性。对校准支持向量机函数定义和 SGL1/2 正则化模型分析可知, f_{cal} 模型是可微的, 能够使用梯度下降的方法进行极小化, 可以大幅减小模型求解的复杂度。

2 实现及分析

将 AAL116 模板作为分组模板, 以大脑区域的灰质密度体素作为特征, 结合 SGL1/2 正则化的 C-SVM 分类模型, 对经过预处理的 ADNI (Alzheimer's disease neuroimaging initiative) 数据集通过交叉验证进行训练和测试, 获取测试的分类效果, 并用该分类模型对其他数据集进行重新训练与测试, 对比分类效果以验证模型的泛化性能。

2.1 数据预处理

使用的数据均来自阿尔茨海默病神经成像计划 (ADNI) 数据集 (<http://adni.loni.usc.edu>)。ADNI 的研究对象来自美国和加拿大的 50 多个地点, 主要目的是测试磁共振成像 (MRI)、正电子发射断层扫描 (positron emission tomography, PET)、其他生物标志物以及临床和神经心理学评估相结合是否可以应用于度量轻度认知损害 (MCI) 和早期阿尔茨海默病

(AD)的识别。

实验选取 ADNI 中公开的 ADNI: Complete 2Yr 1.5 T 的部分数据集(Wyman 等,2013),fMRI 图像为 T1 加权的 1.5 T MR 图像,共选择 511 例来自不同受试者的数据,包括 AD 患者 145 名, MCI 患者 209

名和正常对照(control normal, CN)组成员 157 名。在性别成分中,AD 男性 65 名,女性 80 名; MCI 男性 111 名,女性 98 名; CN 男性 65 名,女性 92 名。所选受试者详细情况如表 1 所示,其中 MMSE 表示简易精神状态检查(后同)。

表 1 ADNI 数据集的人员分布情况统计
Table 1 Distribution of ADNI datasets

类型	数量	年龄分布(均值±方差[范围])	性别分布(M:男,F:女)	MMSE(均值±方差[范围])
AD	145	77.4±7.8[56-91]	65M/80F	21.1±5.1[2-28]
MCI	209	76.1±6.3[55-91]	111M/98F	26.4±3.1[8-30]
CN	157	77.7±6.1[60-92]	65M/92F	29.2±1.0[24-30]

注:[]中括号内的值分别代表数据集实际年龄分布范围和 MMSE 值分布范围。

采用 SPM(statistical parametric mapping)(杨雄雄等,2003)对 fMRI 图像提取的灰质体素特征进行计算,全过程使用 SPM 的默认参数。

另外,本研究选取 Cuingnet 等人(2013)的 Cuingnet 数据集进行测试,以比较模型的泛化性能。

在这个数据集中,509 名受试者都来自 ADNI,包括 162 例 CN,137 例 AD,76 例将在 18 个月内转为 AD 的 MCI(MCIe)和 134 例保持稳定的 MCI(MCIc)。Cuingnet 数据集的特征描述如表 2 所示。

表 2 Cuingnet 数据集的人员分布情况统计
Table 2 Distribution of Cuingnet dataset

类型	数量	年龄分布(均值±方差[范围])	性别分布(M:男,F:女)	MMSE(均值±方差[范围])
AD	137	76.0±7.3[55-91]	67M/70F	23.2±2.0[18-27]
MCIc	76	74.8±7.4[55-88]	43M/33F	26.5±1.9[23-30]
MCIe	134	74.5±7.2[58-91]	84M/50F	27.2±1.7[24-30]
CN	162	76.3±5.4[60-90]	76M/86F	29.2±1.0[25-30]

注:[]中括号内的值分别代表数据集实际年龄分布范围和 MMSE 值分布范围。

由于测试的不同脑部结构大小不一,在对图像进行特征提取前需进行预处理。本研究使用 SPM 软件在 MATLAB 平台对受试者的 sMRI 图像数据进行标准的数据预处理操作,获得不同大脑区域的灰度体积作为训练模型的输入特征。对每个 sMRI 图像数据的预处理流程如图 5 所示。首先对图像进行 AC-PC 校准(Eskandar 和 Tournoux,1988),AC-PC 为前连合(anterior commissure,AC)后缘中点至后连合(posterior commissure,PC)前缘中点的连线,又称 AC-PC 线,在脑立体定位层解剖研究中多以此线为基线。在这步操作中,先将图像的原点设置为中心,再将其注册到 MNI(Montreal Neurological Institute)空间中。之后对校正后的脑图进行去颅骨和分割处理,获取去除非脑组织的图像。然后对图像进行配准,配准使用的模板是剥离组织的 Colin27 模

板(Aubert-Broche 等,2006),以获取更加清晰,高分辨率的空间模板。配准后进行分割获取灰质图像,并对灰质图像结构进行归一化等一系列操作。

2.2 实验设置

实验环境在单机系统搭建,处理器为 Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHz(2201 MHz),32 GB 内存,通过 MATLAB2019a 进行相关实验的运行。在实验中,将预处理过的 ADNI 数据集分成 3 个对比组,分别为 AD 与 CN,AD 与 MCI 以及 CN 与 MCI。同时为了避免训练过程中的过拟合,使用 K-折交叉验证法(Wu 等,2010;James 等,2013),并在数值训练中选择 K 值为 10,即 10 折交叉验证。采用准确度(accuracy, ACC),接受者工作特征曲线(receiver operating characteristics curve, ROC)下的面积(area under curve, AUC),灵敏度(sensitivity,

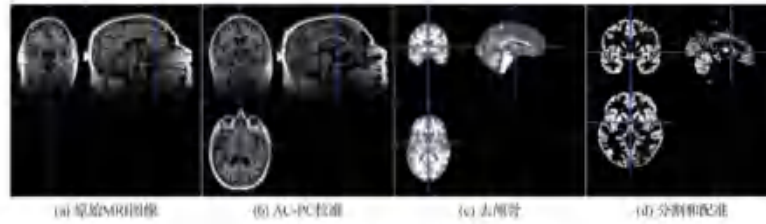


图5 图像预处理过程

Fig.5 Image preprocessing process

(a) original MRI map; (b) AC-PC calibration; (c) skull removal; (d) segmentation and registration

SEN), 特异性 (specificity, SPE)、精确率 (precision)、召回率 (recall)、几何平均值 (geometric mean, Gmean) 和 F1 分数 (F1-score) 等指标对分类性能进行评价。在实际 AD 分类中一般会出现 4 种情况: 若一个样本是 AD 类 (以下称为正类) 且预测为正类, 即为真正类 (true positive, TP); 若一个实例是非 AD 类 (以下称为负类) 但预测为负类, 即为假负类 (false negative, FN); 若一个实例是负类但预测为正类, 即为假正类 (false positive, FP); 若一个实例是负类且预测为负类, 即为真负类 (true negative, TN)。根据这 4 种分类, 使用的评价指标的具体定义为

$$f_{acc} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (13)$$

$$f_{acc} = \frac{\sum rank_i - \frac{(TP + TN) \times (TP + TN + 1)}{2}}{(TP + TN) \times (FP + FN)} \quad (14)$$

式中, $\sum rank_i$ 表示所有正样本的序号值之和。

$$f_{mat} = f_{se} = \frac{TP}{TP + FN} \times 100\% \quad (15)$$

$$f_{sp} = \frac{TN}{TN + FP} \times 100\% \quad (16)$$

$$f_{precision} = \frac{TP}{TP + FP} \times 100\% \quad (17)$$

$$f_{gmean} = \sqrt{f_{mat} \times f_{sp}} \times 100\% \quad (18)$$

$$f_{f1} = \frac{2 \times f_{precision} \times f_{mat}}{f_{precision} + f_{mat}} \times 100\% \quad (19)$$

2.3 实施及结果分析

首先验证 SVM 算法采用不同损失函数对数据分类性能的影响, 以获取对数据集分类效果较好的模型; 之后对比在组级别上的 GL、GLI/2 范数和

SGLI/2 范数结合稀疏范数选择特征的分布和分类性能的效果; 而后使用结合 SGLI/2 范数的校准支持向量机 (C-SVM) 模型, 对更困难的对照组进行分类, 并获取其分类效果; 最后对 Cuingnet 数据集进行测试。

2.3.1 校准铰链损失函数对分类模型的性能影响

选取线性损失函数、逻辑损失函数、hinge 损失函数和 Chinge 损失函数进行对比, 验证基于不同损失函数的 SVM 模型对测试数据的分类性能。在数据集中对 AD 与 CN 对照组进行比较, 结果详见表 3。可以看出, Chinge 函数比其他损失函数具有更好的分类效果。C-SVM 的准确率高达 91.06%, 与 hinge 作为损失函数的标准 SVM 相比, 分类准确率提高了 2.65%, 同时高于线性分类的 90.40% 和逻辑回归的 83.77%。依据此结果, 本实验选择以 Chinge 为损失函数的 C-SVM 作为分类模型, 在接下来的测试中结合其他模型进行训练和测试。

表3 不同模型的分美效果

Table 3 Classification effect of different models				
分类方法	ACC	AUC	SPE	SEN
C-SVM (Chinge)	0.910 6	0.964 6	0.937 9	0.885 4
SVM(hinge)	0.884 1	0.951 8	0.910 3	0.859 9
线性分类	0.904 0	0.959 0	0.937 9	0.872 6
逻辑回归	0.837 7	0.914 1	0.875 9	0.802 5

注: 加粗字体为各列最优结果, 分类方法的括号内为损失函数。

2.3.2 组级别的稀疏化方法比较

根据组稀疏范数的定义, 在组级别的稀疏范数的分类模型中, 由于组稀疏的约束作用, 将倾向选择

与 AD 分类相关性较大的部分脑区域,使得选中的每个区域内的体素相对集中,从而更好地展示与 AD 疾病有关的大脑区域。为了验证组稀疏范数的分类效果,选取 GL、GLI/2 和 SGLI/2 3 种组稀疏范数进行对比。图 6 展示了基于 AAL 模板的 GL 和 SGLI/2 校准支持向量机模型在大脑区域选取的重要分类区域,其中 λ_2 是用于控制组稀疏范数的超参数。可以看出,海马、海马旁回、舌回和梭状回等区域对于 AD 的分类非常重要,这些组方法基本上都选取了这些区域。不同的是,SGLI/2 方法对选中区域内的一些体素进行了稀疏,去除了一些相关度较低的体素,选中的脑部区域如表 4 所示。图 6(a)~(c)是 C-SVM+SGLI/2 模型,能够选取高相关脑区的高相关体素,在加大正则化项惩罚时,仅会去除相对相关度低的脑区和相关度高的脑区中相对相关度低的体素,因此红色标注区域会近似对称;图 6

(d)~(f)是 C-SVM+Group Lasso(GL)模型,仅针对组间进行稀疏,一旦认为某些脑区相对于其他脑区相关性低时,将去除整个脑区中的所有体素。因此在加大正则化项惩罚时,会依次去除相对相关性低的完整脑区,导致惩罚较大时红色标注区域不完全对称的视觉效果。组级别范数的 C-SVM 在 AD 与 CN 上的分类性能如表 5 所示,图 7 是基于 AAL 模板的组方法模型的性能比较。结合表 5 和图 7 可以看出,以 AAL 为分组模板的 SGLI/2 校准支持向量机有一个更好的分类效果,平均运行时间最短,而 GLI/2+C-SVM 平均运行时间最长,这是由于 GLI/2 的不稳定性引起的。GL 的稀疏效果没有 SGLI/2 强,因此存在更慢的迭代求解速度。SGLI/2 范数是对 GLI/2 的一种平滑逼近,二者的组稀疏效果大致相同,但是改进的平滑 GLI/2 范数 SGLI/2 能够使函数更快收敛,防止数据发生振荡。

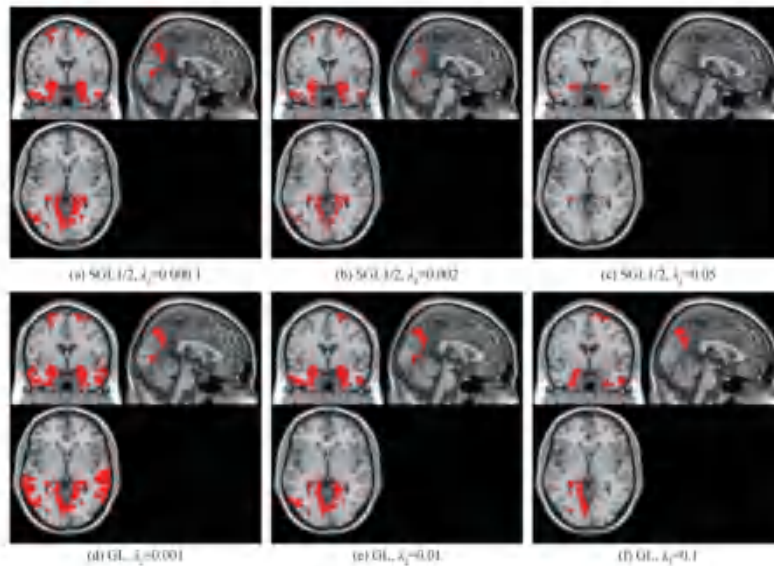


图 6 基于 AAL 模板分组的组稀疏范数模型选中的组特征区域

Fig. 6 AAL template grouping based group feature regions selected by different group sparsity models

((a) SGLI/2 and $\lambda_2 = 0.0001$; (b) SGLI/2 and $\lambda_2 = 0.002$; (c) SGLI/2 and $\lambda_2 = 0.05$;

(d) GL and $\lambda_2 = 0.001$; (e) GL and $\lambda_2 = 0.01$; (f) GL and $\lambda_2 = 0.1$)

表4 基于SGL1/2的C-SVM模型(本文模型)在ADNI数据集上选择的脑区

Table 4 Selected brain regions of C-SVM model based on SGL1/2 (proposed model) in ADNI dataset

脑区编号	中文名称	脑区名称
3	左额外侧额上回	Frontal_Sup_L
37	左海马	Hippocampus_L
38	右海马	Hippocampus_R
39	左海马旁回	ParaHippocampal_L
40	右海马旁回	ParaHippocampal_R
41	左杏仁核	Amygdala_L
42	右杏仁核	Amygdala_R
43	左颞叶周围皮层	Calcarine_L
48	右舌回	lingual_R
55	左梭状回	Fusiform_L
56	右梭状回	Fusiform_R
83	左颞极;颞上回	Temporal_Pole_Sup_L
89	左颞下回	Temporal_Inf_L

注:按 AAL 模板编号从小到大排序。

表5 组级范数的C-SVM在AD与CN上的分类性能

Table 5 Classification performance of group level norm C-SVM on AD and CN

组方法	ACC	AUC	SEN	SPE	平均时间/s
GL	0.940 4	0.978 6	0.937 9	0.942 7	758.11
GL1/2	0.947 0	0.977 4	0.951 7	0.942 7	103.4.02
SGL1/2	0.947 0	0.978 8	0.944 8	0.949 0	446.77

注:加粗字体为各列最优结果。

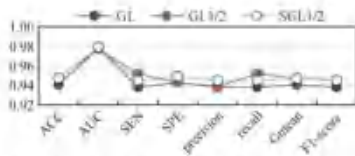


图7 基于AAL模板的组方法模型的性能比较

Fig. 7 Classification performance comparison of group method models based on AAL template

2.3.3 对较难分辨的对照组的分类效果测试

本文结果表明平滑稀疏化在组级上是有效

的,使得特征选择更加聚集,获取与疾病高度相关的局部区域,有更好的分类效果。前文比较了脑部形态差异最大的AD与CN组,在实际中,很多病患通常处在介于AD和CN之间的MCI状况,临床表现通常没有AD患者严重,这种状态与其他两种状态也更容易进行区分,分类边界较为模糊。为了进一步验证模型性能,对AD组与MCI组,CN组与MCI组进行分类测试,分组模板仍使用AAL模板,对比GL和SGL1/2两种组稀疏化方法,结果如表6所示,表6中时间为平均求解时间。可以看出,在AD与MCI对照组中,SGL1/2模型分类准确率为71.47%,高于GL模型的70.62%,平均运行时间较GL减少269.34s,效率提升33.53%;在MCI与CN对照组中,SGL1/2模型准确率为78.81%,比GL模型高1.97%,时间效率比GL提高38.30%。

AD与MCI,CN与MCI的分类效果分别如图8和图9所示。可以看出,基于SGL1/2范数的分类模型的性能都明显高于基于GL范数的分类模型。由此可见,基于SGL1/2的C-SVM模型在更困难的对照组中也能体现良好的分类效果,这对模型的可靠性是一个强有力的支持。

2.3.4 在Cuignet数据集上的分类性能

为了验证本文模型的泛化能力,使用Cuignet数据集对本文模型进行训练和测试。根据不同疾病状态分为不同对照组,即CN与MCI,AD与MCI以及MCIc与MCIe。其中,MCI组由MCIc和MCIe合并而成。通过调整正则化项的超参数,在训练数据集上进行重新优化,其他参数和其他模型一致,完成训练后再进行测试。本文模型与无稀疏化和GL稀疏化两种方法的对比结果如表7所示,表7中时间为平均求解时间。可以看出,在所有对照组中,SGL1/2的分类性能在大部分情况都高于其他模型的性能,仅在CN与MCI组呈较为接近状态,求解时间均大幅降低,在AD与MCI组和CN与MCI组中,在保持高于或近乎持平的分类性能前提下,时间效率较GL方法提升了近60%,优势明显。同时,基于组范数的C-SVM的各项分类性能都优于无范数的C-SVM模型,同时SGL1/2略优于GL,基于SGL1/2的C-SVM模型在Cuignet数据集上也能相较于其他正则化范数达到更好的效果,这充分说明了该模型应对其他数据集时具有较强的泛化能力。

表 6 基于 SGLI/2 的 C-SVM 模型在困难对照组上的分类结果
Table 6 Classification performance of C-SVM model based on SGLI/2 in more difficult control group

对照组	稀疏方法	ACC	AUC	SEN	SPE	时间/s
AD 与 MCI	GL	0.706 2	0.801 3	0.724 1	0.693 8	803.32
	SGLI/2	0.714 7	0.803 3	0.737 9	0.698 6	533.98
CN 与 MCI	GL	0.768 4	0.831 7	0.793 1	0.751 2	784.50
	SGLI/2	0.788 1	0.859 9	0.827 6	0.760 8	484.05

注:加粗字体为对比组中各项性能的最优结果。

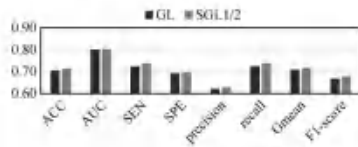


图 8 基于 SGLI/2 的 C-SVM 模型在 AD 与 MCI 的分类效果

Fig. 8 Classification effect of C-SVM model based on SGLI/2 in AD and MCI

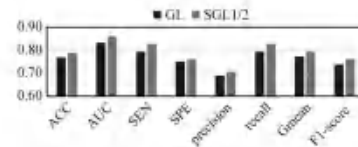


图 9 基于 SGLI/2 的 C-SVM 模型在 CN 与 MCI 的分类效果

Fig. 9 Classification effect of C-SVM model based on SGLI/2 in CN and MCI

表 7 基于 SGLI/2 的 C-SVM 模型在 Cuingnet 数据集上的分类性能
Table 7 Classification performance of C-SVM model based on SGLI/2 on Cuingnet dataset

对照组	稀疏方法	ACC	AUC	SEN	SPE	时间/s
AD 与 CN	-	0.899 7	0.955 4	0.897 8	0.901 2	678.69
	GL	0.916 4	0.967 2	0.912 4	0.919 8	730.65
	SGLI/2	0.919 7	0.963 3	0.912 4	0.925 9	428.31
AD 与 MCI	-	0.639 8	0.686 1	0.678 8	0.614 3	1 361.82
	GL	0.648 4	0.702 3	0.678 8	0.628 6	1 698.31
	SGLI/2	0.651 3	0.713 5	0.686 1	0.628 6	520.52
CN 与 MCI	-	0.680 1	0.772 2	0.734 6	0.638 1	1 177.65
	GL	0.685 5	0.791 3	0.746 9	0.638 1	1 448.78
	SGLI/2	0.682 8	0.778 9	0.740 7	0.638 1	499.44
MCI _o 与 MCI _e	-	0.628 6	0.667 2	0.684 2	0.597 0	468.97
	GL	0.661 9	0.679 8	0.723 7	0.626 9	649.78
	SGLI/2	0.661 9	0.687 6	0.710 5	0.634 3	359.18

注:加粗字体为对比组中各项性能的最优结果;“-”表示未使用组稀疏方法。

3 基于 AAL 模板的各特征组独立分类效果比较

为进一步了解与 AD 高度相关的单独脑部区域

对区分 AD 的效果,以 AAL 分组模板对进行分组的各组特征使用单独分类函数进行分类,对左右脑平均分类排名前 27 的脑部各区域及同一区域左右脑区的分类效果进行对比,结果如图 10 所示。其中,分类效果最好的右舌回(Lingual_R)、右海马旁回

(ParaHippocampal_R)、右海马(Hippocampus_R)、右梭状回(Fusiform_R)、左颞下回(Temporal_Inf_L)和左背外侧额上回(Frontal_Sup_L)6个脑区的具体分类效果如图11所示。

将高相关的单独组区域得到的最佳结果与基于SGL1/2正则化的C-SVM模型进行比较,如表4和图11所示,可以发现这种组稀疏方法也基本选

中了本文模型获取的分类准确率排名前6的脑区,证明了该组方法的可靠性。同时,本文模型在基于AAL模板中获得的分类准确率为94.70%,比高相关单独组经过非组稀疏的最佳分类准确率高2.32%,这一结果同样很好地证明了多个脑区之间存在一定的相关性,它们共同影响AD的发病。

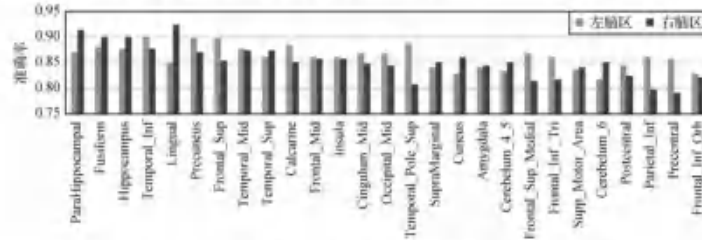


图10 单独分类模型在AAL模板中左右脑区域分类准确率比较

Fig. 10 Comparison of classification accuracy between left and right brain in AAL template

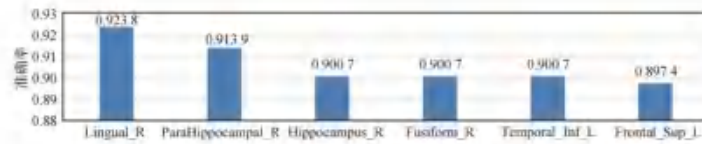


图11 分类准确率排名前6的脑区

Fig. 11 Brain regions with top 6 classification accuracy

4 与现有方法比较

为了进一步验证本文基于SGL1/2组稀疏方法的C-SVM分类模型性能优越性,与多核学习方法(multiple kernel learning, MKL) (Ben Ahmed等, 2017)、3维卷积神经网络(3D convolutional neural networks, 3D-CNN) (Khvostikov等, 2018)、SVM + GL + SAR (spatial anatomical regularization) (Sun等, 2018)、deep CNNs (convolutional neural networks) (Aderghal等, 2018)、3维卷积神经网络与长短期记忆网络(3D convolutional neural networks—long short-term memory, C3d-LSTM) (Li等, 2020)和SVM + ROICSE (ROI-based contourlet subband energy (Feng等, 2021)等AD诊断方法在ADNI数据库上进行对

比,以诊断准确率(ACC)为评判指标,结果如表8所示。可以看出,本文模型具有更好的分类效果。

表8 不同方法分类性能比较

Table 8 Comparison of classification performance of different methods

模型	ACC
MKL (Ben Ahmed等, 2017)	0.902 0
3D-CNN (Khvostikov等, 2018)	0.829 0
SVM + GL + SAR (Sun等, 2018)	0.893 0
deep CNNs (Aderghal等, 2018)	0.925 0
C3d-LSTM (Li等, 2020)	0.927 8
SVM + ROICSE (Feng等, 2021)	0.935 7
本文	0.947 0

注:加粗字体表示最优结果。

5 结论

本研究提出基于解剖平滑分组的 $L_{1/2}$ 正则化方法,并结合校准支持向量机组合成分类框架,实现对阿尔茨海默症(AD)的分类。这种平滑解剖分组稀疏的方法以流行的 Group Lasso 为原型,针对性地解决了 Group Lasso 不能对每个解剖组内的特征进行稀疏的缺点,达到可以同时去除冗余组和相关组内一些冗余特征的效果。此外,针对组级别上的正则化范数提出了经过配准后的 AAL 模板,将每个区域作为一组,每个区域内所有体素作为每组的特征进行组稀疏。然后在提出的 C-SVM 框架中以组为级别进行训练,最后预测出受试者是否患有 AD。实验结果表明,本文模型结果比其他稀疏范数有更好的结果,分类准确率最高达 94.70%。同时,选中的权重分布较为集中,能够更好地发现与疾病高度相关的区域,增强了可解释性。根据权重图的分布情况可以将选定脑区定位到 AD 相关的脑区,并且选择的区域也会较为集中。根据当前的知识可知,海马旁回、舌回和梭状回是与 AD 高度相关区域,权重图在这些区域也展现了较高的相关性,这为实验结果的合理性提供了有效支持。

参考文献(References)

- Adedajal K, Klyasnikov A, Krylov A, Benois-Pinaud J, Allet K and Catheline G. 2018. Classification of Alzheimer disease on imaging modalities with deep CNNs using cross-modal transfer learning//Proceedings of the 31st IEEE International Symposium on Computer-Based Medical Systems, Karlstad, Sweden: IEEE: 345-350 [DOI: 10.1109/CIBMS.2018.00067]
- Aleno H Z, Zhao J H, Li F and Wu W. 2019. Group $L_{1/2}$ regularization for pruning hidden layer nodes of feed forward neural networks. *IEEE Access*, 7: 9540-9557 [DOI: 10.1109/ACCESS.2019.2890740]
- Aidun-Binudil B, Evans A C and Collins L. 2006. A new improved version of the realistic digital brain phantom. *NeuroImage*, 32(1): 178-185 [DOI: 10.1016/j.neuroimage.2006.03.052]
- Ben Ahmed O, Benois-Pinaud J, Allard M, Catheline G and Ben Amar E. 2017. Recognition of Alzheimer's disease and mild cognitive impairment with multimodal image-derived biomarkers and multiple kernel learning. *Neurocomputing*, 220: 98-110 [DOI: 10.1016/j.neucom.2016.08.041]
- Chen H Y, Gao J Y, Zhao D, Wang H Z, Song H and Su Q H. 2021. Review of the research progress in deep learning and biomedical image analysis till 2020. *Journal of Image and Graphics*, 26(7): 475-486(邵江涛,高静霞,赵敏,江红志,宋红,苏庆华. 2021. 深度学习与生物医学图像分析 2020 年综述. 中国图象图形学报, 26(7): 475-486)
- Geis C and Vapnik V. 1995. Support-vector networks. *Machine Learning*, 20(3): 273-297 [DOI: 10.1023/A:1022627411411]
- Güngör R, Ghaemi J A, Ghossein M, Benali H and Galifian O. 2013. Spatial and anatomical regularization of SVM: a general framework for neuroimaging data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3): 682-696 [DOI: 10.1109/TPAMI.2012.142]
- De Magalhães Oliveira F P B, Nitrini R, Busatto C, Buchpiguel G, Sato J R and Amar E Jr. 2010. Use of SVM methods with surface-based cortical and volumetric subcortical measurements to detect Alzheimer's disease. *Journal of Alzheimer's Disease*, 19(4): 1263-1272 [DOI: 10.3233/JAD-2010-1322]
- Fan Q W, Xu L and Kang Q. 2020. Regression and multiclass classification using sparse convex learning machine (in smoothing $L_{2,1}$ regularization). *IEEE Access*, 8: 191482-191494 [DOI: 10.1109/ACCESS.2020.3031647]
- Feng J W, Zhang S W and Chen J N. 2021. Extracting ROI-based cortical-subcortical young features from the sMRI image for Alzheimer's disease classification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, online, 1st [DOI: 10.1109/TCBB.2021.3081177]
- Filippi G B, Fox N C, Jack C R, Scheltens P and Thompson P M. 2010. The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology*, 6(2): 67-77 [DOI: 10.1038/nrn2609.215]
- Fing G and Shao G J. 2007. SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information. *Knowledge and Information Systems*, 11(2): 243-258 [DOI: 10.1007/s101150064043-5]
- James G, Witten D, Hastie T and Tibshirani R. 2013. *An Introduction to Statistical Learning, with Applications in R*. New York, USA: Springer
- Klyasnikov A, Adedajal K, Benois-Pinaud J, Krylov A and Catheline G. 2018. 3D CNN-based classification using sMRI and MD-OT images for Alzheimer disease studies (EB-01). [2021-05-01]. <https://arxiv.org/pdf/1801.05968v1.pdf>
- Lee H, Elahi W and Choi J Y. 2019. Using deep CNN with data permutation scheme for classification of Alzheimer's disease in structural magnetic resonance imaging (sMRI). *IEEE Transactions on Information and Systems*, E012-19(7): 1384-1395 [DOI: 10.1587/transinf.2018EDP7393]
- Li F, Zoude J M and Wu W. 2018. Smooth group $L_{1/2}$ regularization for deep layer of feedforward neural networks. *Neurocomputing*, 314:

- 109-119 [DOI: 10.1016/j.neuroim.2018.06.046]
- Li W., Liu X. F. and Chen X. 2020. Detecting Alzheimer's disease based on 4D fMRI: an exploration under deep learning framework. *Neurocomputing*, 388: 280-287 [DOI: 10.1016/j.neuroim.2020.01.053]
- Liu W. J. 2000. Principles of magnetic resonance imaging: a signal processing perspective [book review]. *IEEE Engineering in Medicine and Biology Magazine*, 19(5): 129-130 [DOI: 10.1109/EMMB.2000.870245]
- Liu M. H., Suk H. I. and Shen D. G. 2013. Multi-task sparse classifier for diagnosis of MCI conversion to AD with longitudinal MR images // *The 4th International Workshop on Machine Learning in Medical Imaging*. Nagoya, Japan; Springer: 243-250 [DOI: 10.1007/978-3-319-02267-3_31]
- Liu S. C., Zhang J. S., Liu J. M. and Yin Q. Y. 2016. $l_{1/2,1}$ group sparse regularization for compressive sensing. *Signal, Image and Video Processing*, 10(5): 861-868 [DOI: 10.1007/s11760-015-0829-6]
- Liu Y. X., Tang Y. Y., Wu J. L. and Song H. S. 2003. Introduction of brain functional imaging analysis software SPM. *Chinese Journal of Medical Imaging Technology*, 19(7): 926-928 (杨晓星, 唐一雷, 伍律林, 宋鹤山. 2003. 脑功能成像分析软件 SPM 使用介绍. *中国医学影像技术*, 19(7): 926-928) [DOI: 10.3321/j.issn:1003-3289.2003.07.057]
- Mattson M. P. 2004. Pathways towards and away from Alzheimer's disease. *Nature*, 430 (7000): 631-639 [DOI: 10.1038/nature02621]
- Meier I., Geer S. V. D. and Bühlmann P. 2008. The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1): 53-71 [DOI: 10.1111/j.1467-9868.2007.00627.x]
- Sakkalis V. 2011. Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Computers in Biology and Medicine*, 41 (12): 1110-1117 [DOI: 10.1016/j.compbiomed.2011.06.020]
- Shi Z. R., Wang S. and Liu C. C. 2017. Sparse representation via $l_{2,p}$ norms for image classification. *Journal of Nanjing University of Science and Technology*, 41 (1): 80-89 (时中荣, 王胜, 刘传才. 2017. 基于 $l_{2,p}$ 范数稀疏表示的图像分类方法. *南京理工大学学报*, 41 (1): 80-89) [DOI: 10.14177/j.cnki.32-1397n.2017.41.01.011]
- Sun Z., Qiao Y. C., Leditschke B. P. F. and Staring M. 2018. Integrating spatial-anatomical regularization and structure sparsity into SVM: improving interpretation of Alzheimer's disease classification. *NeuroImage*, 178: 445-460 [DOI: 10.1016/j.neuroimage.2018.05.051]
- Talairach J. and Tournoux L. 1988. *Co-Planar Stereotaxic Atlas of the Human Brain; 3-Dimensional Proportional System; an Approach to Cerebral Imaging*. New York, USA; Thieme Medical Publishers
- Tournois-Mazoyer N., Landeau B., Papathanassiou D., Crivello F., Etard O., Delcroix N., Mazoyer B. and Joliet M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1): 273-289 [DOI: 10.1006/nimg.2001.0978]
- Wu W., Li L., Yang J. and Liu Y. 2010. A modified gradient-based non-fuzzy learning algorithm and its convergence. *Information Sciences*, 180(9): 1630-1642 [DOI: 10.1016/j.ins.2009.12.030]
- Wyman B. T., Harvey D. J., Crawford K., Bernstein M. A., Carmichael O., Cole P. F., Crane P. K., DeCarli C., Fox N. C., Gatter J. L., Hill D., Killiany R. J., Parfitt C., Schwarz A. J., Seluff N., Seung M. L., Suly J., Thompson P. M., Weiner M., Jack C. R. Jr. and Alzheimer's Disease Neuroimaging Initiative. 2013. Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimer's and Dementia*, 9(3): 332-337 [DOI: 10.1016/j.jalz.2012.06.004]
- Yuan M. and Lin Y. 2006. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68 (1): 49-67 [DOI: 10.1111/j.1467-9868.2005.00532.x]

作者简介

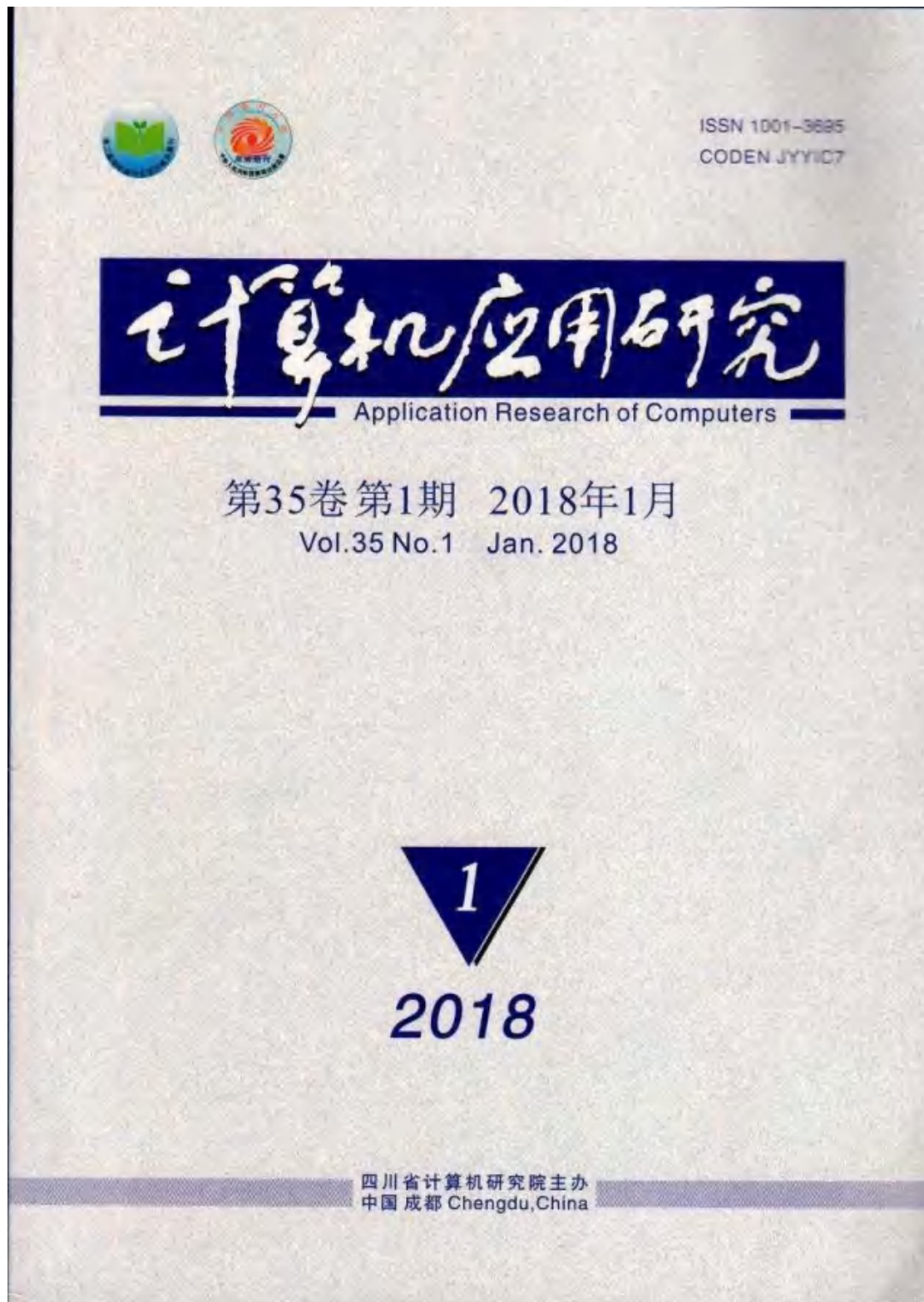


黄坤辉, 1997年生, 男, 硕士研究生, 主要研究方向为医学图像识别。
E-mail: 13979445473@163.com



王金凤, 通信作者, 女, 副教授, 主要研究方向为机器学习和图像识别。
E-mail: wangjinfeng@scnu.edu.cn

3.7. 并行计算和稀疏存储在模糊积分上的应用



系统应用开发

多目标粒子群算法在混装线再平衡中的应用 戴金州, 吴永明, 李少波, 罗利飞 (145)

变邻域遗传算法在多目标柔性车间调度中的应用 曹 磊, 叶春明, 黄 霞 (150)

Doc2vec 在新水预测中的应用研究 潘 博, 张育川, 于重豪, 曹 帅 (155)

贸易信用下考虑零售违约风险的供应链协调问题研究 马中华, 郑周均 (158)

改进型神经网络在室内三维定位中的应用研究 刘 敏, 黄文斌, 徐晋水, 陈 涛 (162)

并行计算和稀疏存储在模糊部分上的应用 靳润健, 王金凤 (166)

基于 KPCA 和改进级联学习机的煤与瓦斯突出危险性测试 李 非, 胡海杰 (172)

代教连通性在社会网络影响力传播最大化中的应用研究 赵富强, 杨贵军, 王双琳, 何 杰 (177)

基于 Steam 的 DMS 交互系统客户端的设计与实现 刘路永, 谈 杰, 姜一本, 王成传 (182)

基于自组织神经网络和模糊推理的校园无线网络用户学习兴趣度行为分析 王法玉, 姜 妍 (186)

网络与通信技术

基于前态优先级数据中心网络拥塞感知 TCP 协议 刘 洪, 伊 鹏, 胡宇翔 (190)

基于射频识别和无接触传感融合技术的仓储定位方法研究 王胡丹, 王宜祥, 刘 颖 (195)

大列重负载态度的 QC-LDPC 码构造 赵 辉, 郭凯勇, 郭海英 (199)

基于环路分类的图长至少为 10 的 QC-LDPC 码呈式构造方法 谷 静, 范裕子, 张国华 (204)

网络控制系统时变采样周期的建模与切换控制 杨文才, 刘 斌 (208)

一种前馈神经网络和遗传算法结合的寻优方法 洪 珊, 张 莹, 赵艳杰 (213)

面向云搜索集合网中节点资源有限的移动数据库研究 李银亮, 谭大禹, 朱永强, 董红娟, 杨大禹 (217)

音频水印提取算法中降低信号方差的方法 吕 伟, 陈 怡, 高 光, 刘 影 (222)

基于网络编码和多边网点对称的视频多播方案 廖金明, 冯光耀 (226)

区域级调分支界定混合 P2P 云存储数据编码优化 安 进, 李 木 (232)

信息安全技术

基于混沌的视觉光学加密技术研究 王洪伟, 李 军, 刘 帆, 冷志光 (237)

背包公钥密码安全新方案 廖向东, 潘 芳, 潘 航 (240)

m 子序列的密码学性质研究 孙全玲, 吕 虹, 陈万里, 葛 鹏 (245)

基于密文策略属性基加密的多授权中心访问控制方案 雷丽敏, 李 勇 (248)

改进型 NIST 测试对 ZUC 算法随机序列的可视化检测 吴若雪, 梁发祥, 郑智斌 (253)

WSNs 中基于可变长度动态路由的跟踪隐私保护方案 刘 亚, 许拥磊, 宋 梁 (257)

帧式自编码的态替代划分算法研究 罗世亮, 田生伟, 孙 华, 冯 亮 (261)

基于改进遗传算法的大容量可逆数据隐藏 何文广, 潘顺德, 周 珂, 蔡 浩 (266)

S-HSSE 算法 李梦杰, 蔡坤儒, 郭玉芳 (272)

基于 Modbus 功能的指纹识别算法研究 张仁斌, 李恩刚, 刘 飞, 许辅英 (277)

图形图像技术

一种新的多角度人脸表情识别方法 何 俊, 何志文, 蔡建峰, 胡灵芝 (282)

基于 PT(SVD) 的张量奇异值分解及其在人脸识别上的应用 戴 超, 徐 宁, Kan Ramesh, Naji Wong, 喻文健 (287)

基于多种 LBP 特征集成学习的人脸识别 何 云, 吴怀宇, 钟 强 (292)

手指静脉的鲁棒性边缘检测算法 王明文, 唐东明, 于耀彪, 杨 进 (296)

基于西恩非鲁棒或用于运动检测算法 吴群峰, 侯 进 (300)

基于块稀疏度估计的压缩感知自适应重构算法 许华杰, 刘慧群, 胡小明 (305)

改进全变分保边平滑分割 孙建明, 何 军, 周德波 (309)

基于遗传优化的自适应凸松弛人体姿态估计 刘方军, 王 杰, 付兴武, 曲海成 (313)

基于 TD 框架的快速目标跟踪方法 董 磊, 曹树斌, 沈 捷 (317)

信息集萃

下期要目 (189)

期刊基本参数: (CN) 11-196/TP * 1984 = m * A4 * 320 * zh * P * ¥ 35.00 * 3840 * 68 * 2018-1 * n

本期责任编辑: 洪志全

并行计算和稀疏存储在模糊积分上的应用*

陈润健¹, 王金凤¹

(华南农业大学 数学与信息学院, 广州 510642)

摘要:近年来很多学者开展了模糊积分的相关研究,并将模糊积分应用于各种分类问题,而模糊测度的确定则是模糊积分计算的重点和难点。将并行计算和稀疏存储应用在模糊积分求解上,分别解决模糊积分计算中的时间复杂度和空间复杂度问题,并提出一种高效稀疏模糊积分算法——基于并行和稀疏框架的模糊积分(parallel and sparse frame based fuzzy integral, PSFI)。实验表明,随着计算资源的增加,PSFI算法的加速比和效率下降较低。在变量存储上,PSFI算法在较多特征的数据集上对存储空间减少数十倍。最后,提出的PSFI算法相比之前提出的多重模糊积分(multiple nonlinear integral, MNI)算法,有较高的分类准确率。

关键词:模糊测度;模糊积分;并行计算;稀疏存储;分类

中图分类号: TP301.5 **文献标志码:** A **文章编号:** 1001-3695(2018)01-0166-06

doi:10.3969/j.issn.1001-3695.2018.01.035

Application of parallel computing and sparse storage in fuzzy integral

Chen Runjian, Wang Jinfeng¹

(College of Mathematics & Information, South China Agricultural University, Guangzhou 510642, China)

Abstract: Recently, many scholars have studied the fuzzy integral which is widely used in classification. A large number of studies show that computing the fuzzy measure is the key and difficult problem in fuzzy integral. This paper used parallel computing and sparse storage to solve time and space complexity in fuzzy measure computation respectively, and then put forward an algorithm named parallel and sparse frame based fuzzy integral (PSFI). Experimental results show that with the increase of computing resources, the speedup and efficiency of PSFI algorithm decrease slowly. In variable storage, PSFI algorithm reduces storage space by several thousand times on datasets with many features. Compared with the multiple nonlinear integral (MNI) algorithm proposed previously, PSFI algorithm has higher classification accuracy.

Key words: fuzzy measure; fuzzy integral; parallel computation; sparse storage; classification

0 引言

Hadoop项目成立于2006年,是一个分布式存储和分布式计算框架^[1]。自Hadoop项目发布以来,大数据的技术和应用快速普及和发展。随着人们对计算速度要求的提高,起源于2009年的Spark项目逐渐取代Hadoop项目的MapReduce框架^[2]。Spark的核心是RDD(弹性分布式数据集),与Hadoop的MapReduce不同,其中间数据输出到内存,减少了I/O的存取时间,因此Spark的迭代计算速度更快。至今,Hadoop的分布式存储系统HDFS和Spark分布式处理框架在大数据领域被广泛应用。模糊测度和模糊积分起源于20世纪70年代,在信息融合中有着较广泛的应用。模糊积分属于非线性积分,其主要特点是集函数的非可加性,可以考虑特征之间的交互作用,从而弥补了经典模糊积分特征独立假设的缺点^[3],具有较好的分类效果。但是模糊积分在计算模糊测度上具有较大的时间复杂度和空间复杂度,随着特征数量的增加,其复杂度呈指数增长,这一点使得模糊测度的计算十分困难。目前,国内外的论文库中还没有发现模糊积分的并行算法。近几年CPU并行计算领域研究较多的是关于Spark和Hadoop Map-

Reduce并行计算算法的运行时间和加速比的比较。文献[4-8]显示,在SVM、关联规则、聚类等不同算法的并行计算上,Spark并行计算框架比MapReduce框架快2-3倍甚至几十倍,比Hadoop MapReduce有更大的效率提升,所以本文选取Spark作为模糊积分并行计算的框架。本文提出一种新的算法,将大数据中的并行计算和稀疏存储技术应用到模糊测度的计算上,减少程序的运行时间,提高了算法程序的性能。

1974年,日本学者 Sugeno^[9]在他的博士论文中提出了 Sugeno 模糊积分;1991年,Sugeno 和另一名日本学者 Murofushi 根据 Choquet 在 1954 年提出的容量原理定义了 Choquet 模糊积分^[6]。Choquet 和 Sugeno 模糊积分也是模糊积分最基本的两类积分。近年来,模糊积分的理论不断丰富和扩展,关于模糊积分的应用研究领域也越来越广泛。Choquet 和 Sugeno 模糊积分在木材质量评估^[10]、公众关于核电站使用的态度分析^[11]、彩色打印图片评估^[12]、扬声器的设计^[13]和人的可靠性分析^[14]等方面取得了较好的成果^[15]。2000年,Wang 等人^[16]提出了一种更直观可解释性强的模糊积分定义,简称 Zhou-Suan 模糊积分,并在五个特征的数据集上计算出一个最优结果。2003年,Xu 等人^[17]提出了 WCIIP 方法,将高维空间投影

收稿日期: 2016-08-29 **修回日期:** 2016-10-12 **基金项目:** 国家自然科学基金青年基金项目(61202295),广东省公益研究与能力建设资助项目(2015A020209150,2015A030401061)

作者简介: 陈润健(1992-),男,广东广州人,硕士研究生,主要研究方向为数据挖掘、并行计算;王金凤(1978-),女(通信作者),河北青县人,硕士,主要研究方向为机器学习、数据挖掘;wangjinfeng@163.com。

到低维空间上,相关的研究有文献[18,19]。同年,Wang 等人^[20]提出一种新的模糊积分计算方法,将非线性积分转换到线性函数上。2009年,Wang 等人^[21]使用套索回归方法学习模糊积分上的模糊测度值,发现套索回归能够使求得的模糊测度具有较少的非零值,并且提高了分类精度。模糊积分还常用于分类器的融合,将多个单分类器融合到一个混合的分类器模型,获得更高的分类效果,相关研究有文献[22,23]。这类研究结合其他分类模型优点和模糊积分考虑交互作用的优势,一般可以取得较好的分类效果,而且特征数量为分类器的个数,在模糊积分的计算量上不大。

在国内外关于模糊积分的研究中,大量文献[3,16,19,21]提到了获取模糊测度是一个关键问题,也是一个困难问题,随着特征数量增加,模糊积分的时间复杂度呈指数增长,这个问题在本文第二章中会进行讲述。目前关于模糊积分的研究,在较多特征的数据集上一股采取特征提取方法,然后再建立模糊积分模型。这些模型选取的特征数往往在六个或六个以下,总取的特征数量较少,因而特征信息损失较大,模型的预测准确率也相对降低。本文将结合大数据处理框架 Spark,对模糊积分中模糊测度的学习采用并行计算方法,在模糊测度的计算时间上有较大的提高;并在特征提取上选取一个较合适的特征数目值。

1 模糊积分概述

1.1 Zhenyuan 模糊积分的定义

下面给出本文算法所用到的 Zhenyuan 模糊积分的基本定义^[24]。

定义 1 给定一个集函数 $\mu: \mathcal{F} \rightarrow [0, +\infty)$, 满足 $\mu(\emptyset) = 0, A \in \mathcal{F}$, 函数 $f: X \rightarrow [0, +\infty)$ 在 A 上 f 关于 μ 的积分用符号 $(\mu) \int f d\mu$ 表示, 定义如下:

$$(\mu) \int f d\mu = \sup \left\{ \sum_{j=1}^k \lambda_j \mu(E_j) \mid f \geq \sum_{j=1}^k \lambda_j E_j, \right. \\ \left. \lambda_j \geq 0, E_j \in \mathcal{F} \cap A, \lambda_j \geq 0, j = 1, 2, \dots, k \right\}$$

其中: $\mathcal{F} \cap A = \{E \cap A \mid E \in \mathcal{F}\}$; f 表示特征函数; X 是 X 上的集类。

以上定义可得到下面的等价形式:

$$(\mu) \int f d\mu = \sup \left\{ \sum_{j=1}^k \lambda_j \mu(E_j \cap A) \mid f \geq \sum_{j=1}^k \lambda_j E_j \cap A, \right. \\ \left. \lambda_j \geq 0, E_j \in \mathcal{F}, \lambda_j \geq 0, j = 1, 2, \dots, k \right\}$$

f 不一定可测。当 μ 是经典测度且 f 可测时, Zhenyuan 模糊积分与 Lebesgue 积分是一致的。Zhenyuan 模糊积分是 Lebesgue 积分的推广。

设 X 为特征集, 另 $S = \{x_1, x_2, \dots, x_n\}$ 为有限集, 它的幂集 $P(X)$ 为 \mathcal{F} 。此时, 任何定义在 X 上的函数都是可测的, 并且由于所有的单点集都包含在 $P(X)$ 内, 所有上确界一定在等式 $f = \sum_{j=1}^k \lambda_j E_j$ 成立时达到, 所以积分的表示可简化为如下形式:

$$(\mu) \int f d\mu = \max \left\{ \sum_{j=1}^{n-1} \lambda_j \mu(E_j \cap A) \mid f \geq \sum_{j=1}^{n-1} \lambda_j E_j \cap A, \lambda_j \geq 0 \right\}$$

这里 $E_j = \{x_i \mid \frac{j-1}{2} < \frac{i}{2} \leq \frac{j}{2}, 1 \leq i \leq n\} \subseteq X, j = 1, 2, \dots, 2^n - 1$ 。

1.2 Zhenyuan 模糊积分的线性表示

Zhenyuan 模糊积分将分类器看做一个多输入单输出系

统, 从经典积分中求特征的加权平均过渡到求特征以及特征之间的交互作用的积分。文献[24]使用套索回归方法对模糊测度进行求解, 并给出 Zhenyuan 模糊积分的线性函数为

$$\int f d\mu = \sum_{i=1}^n c_i x_i$$

其中: c_i 是方程的系数; x_i 是待求解的模糊测度。一个包含三个特征的数据集用 $f(x_i)$ 表示其特征值, $i = 1, 2, 3$, c_i 的值由以下函数决定:

$$\begin{aligned} c_1 &= \min\{f(x_1)\} - \min\{f(x_2), f(x_3)\} \\ c_2 &= \min\{f(x_2)\} - \min\{f(x_1), f(x_3)\} \\ c_3 &= \min\{f(x_3)\} - \min\{f(x_1), f(x_2)\} \\ c_{12} &= \min\{f(x_1), f(x_2)\} - \min\{f(x_3)\} \\ c_{13} &= \min\{f(x_1), f(x_3)\} - \min\{f(x_2)\} \\ c_{23} &= \min\{f(x_2), f(x_3)\} - \min\{f(x_1)\} \\ c_{123} &= \min\{f(x_1), f(x_2), f(x_3)\} - \min\{0\} \end{aligned}$$

上述函数也可以看成由两个 $2^n - 1$ 维向量组成, 即向量 Z 和向量 R , 其中 $x_i \leq 0$ 时 $x_i = 0$ 。另 $\{f(x_1), f(x_2), f(x_3)\} = (60, 40, 10)$ 时, 计算得 $Z = (20, 0, 30, 0, 0, 0, 10)$ 。当特征数量增大时, 向量 Z 中的零值也越多。

模糊测度的计算需要求解所有特征组合的函数值, 该算法的时间复杂度为 $O(2^n)$, 所需存储空间随着特征个数增加呈指数增长。 $O(2^n)$ 的时间复杂度十分庞大, 当特征数目在 10 个或以上时, 一般先进行特征稀疏操作, 再建立模糊积分的分类模型。

2 关键技术与计算方法

2.1 并行计算

并行计算主要有两种分解模式: 任务分解和数据分解。任务分解是指把计算分解为一组独立的任务, 多个线程可以任意执行这些任务; 数据分解是指应用程序需要处理一个大数据集, 并且可以对数据集中的每个元素进行独立计算。一个 10 个特征的数据集中有 1 000 条记录, 用数据分解将数据分成四块, 则可以同时运行四个任务, 每个任务分别处理 250 条记录; 如果用任务分解, 则将一条记录的任務分解成多个任务并行处理。在模糊测度的计算上, 对数据集中的每条记录都要计算其模糊测度的系数。在整个计算过程中, 每条记录的计算都是独立的, 因此可以通过数据分解将数据集分成多个块, 使用并行计算减少程序的计算时间。

并行计算的框架主要包括 MPI、Hadoop 的 MapReduce 和 Spark。MPI 支持 C、C++ 语言, 程序可以编译生成类库, 但 C、C++ 语言相对于 Java 等面向对象语言, 数据处理能力较弱。Spark 和 MPI 相比, 容错性高, 一旦有任务出错, Spark 会自动重新计算该任务, 而无须重新计算整个数据集。而 Hadoop MapReduce 与 Spark 相比, Spark 的运行速度更快, 支持的编程语言更多。

图 1 为 Spark 中数据分块的并行编程架构图。图 1 左边是数据分块, 数据集被分成 n 个数据块, 每个数据块作为一个计算任务, 所有的任务进入一个待处理的队列中, 由调度器分配给计算节点。Spark 调度器的默认调度方法为 FIFO (先进先出)。调度器按先进先出的方式将任务分配到所有可用的计算单元, 当首个计算单元计算完分配的任务, 变为空闲单元, 调度器从队列取出新的任务并分配到空闲单元上, 直到所有任务分

配完成。一般地,数据分块个数大于计算单元数。一个工作节点有多个计算单元,计算单元的个数小于等于 CPU 的核心个数。

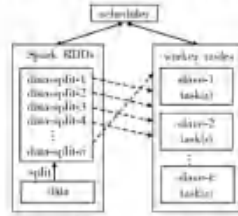


图1 并行计算架构图

2.2 稀疏存储

稀疏存储是指只存储非零值,而不存储的值默认为 0。稀疏存储主要用于存储较多零值的数据。由 1.2 节中模糊测度的计算可知,方程的系数向量存在较多的零值。当使用稠密存储时,随着特征数增加存储空间呈指数增长;而使用稀疏存储,只需要存储非零值,可节省大量存储空间。

如图 2 所示,以向量(0,0,5,0,0,2,0)为例来说明稀疏存储的优点。向量(0,0,5,0,0,2,0)含有 7 个特征值,其中包括 6 个零值和两个非零值。稀疏存储只记录这两个非零值,或者说具有实际意义的值,而不记录零值。因此,稀疏存储使用(2,6)和(5,2)这两个数组表示原向量,意思是从第 0 位开始计算,第 2 和第 6 位的值分别为 5 和 2,其余值为 0。由此可知,稀疏存储和稠密存储表示的内容是相同的,其意义并没有改变,而当特征向量包含很多零时,用稀疏存储可以节省大量的空间。

	0	1	2	3	4	5	6	7
稠密存储	0	0	5	0	0	0	2	0
稀疏存储			5			2		

图2 稀疏存储与稠密存储

2.3 模糊测度的计算方法

一个分类问题的数据集,用 $X = \{x_1, x_2, \dots, x_n\}$ 表示它的属性特征, Y 表示类别, l 表示数据集的记录数, n 表示特征个数。用 $(f_j(x_1), f_j(x_2), \dots, f_j(x_n), Y_j), j = 1, 2, \dots, l$ 表示第 j 条记录的特征值及其类别,并记录为 key_j 。

在模糊测度的计算中,数据集中每条记录的计算是独立的。根据数据分解的任务是并行计算,把数据集中的记录分块输入到算法 1 中,程序最终输出所有记录的 key 数组、value 数组及其类别 label 值。算法中的 binary 数据存储在二进制数值,实现一个二进制累加器。二进制累加器遍历所有组合数比多重循环和递归算法效率更高。数组 key 只存储非零系数的索引值,数组 value 存储非零系数的真实值,利用这两个数组实现方程系数的稀疏存储。

设 Zhenyuan 模糊积分的输出值为

$$F_\lambda = \int f d\mu = \sum_{i=1}^{n-1} \lambda_i \mu_i$$

设模型的误差为 ϵ ,采用最小化误差的方法求解最优化的模糊测度方案。

$$\epsilon = \sqrt{\frac{1}{l} \sum_{i=1}^l (F_i - Y_i)^2}$$

Spark MLlib 自带一个 LASSO 套索回归的实现库,并支持

稀疏向量和稀疏学习。LASSO 使用 L1-norm 正则化,它有特征提取的作用。结合 L1-norm 正则化求解方程,使模糊测度含有较多的零值。L1-norm 惩罚项的回归方程如下所示:

$$\frac{1}{2n} \left\| \sum_{i=1}^n \lambda_i \mu_i - Y \right\|_2^2 + \lambda \left\| \mu \right\|_1$$

其中: λ 的默认值为 0.01。

算法 1 求解方程系数

输入: line_j, 逐条记录输入。

初始化变量: $n = \text{length}(\text{line}_j), \lambda_2 = 2^{n-1}$;

初始化最大值和最小值: $\text{max} = 0, \text{min} = \infty$;

初始化存储空间: $\text{key} = \text{new array}(), \text{value} = \text{new array}()$;

初始化二进制数组: $\text{binary} = \text{new array}(n) = (0)$;

label = Y_j;

line = $(f_j(x_1), f_j(x_2), \dots, f_j(x_n))$;

for $j = 0$ to n do

binary = $+ 1$ //二进制累加器

max = 0, min = ∞ ;

for $j = 0$ to n do

if binary[j] == equal 0

if max < line[j]

max = line[j];

end

also if min > line[j]

min = line[j];

end

if min-max > 0

key.insert(j);

value.insert(min-max);

end

end

输出: key, value 和 label。

上述方法先计算出方程的系数,然后使用最优化方法求解出 2^{n-1} 个 μ 值。这些 μ 值即为该分类问题的特征以及特征交互作用的重要性程度表示,用于构建模糊积分分类器对数据进行分类。本文提出的算法 1 结合了并行计算和稀疏存储,对模糊积分的算法在运行时间和内存空间的使用均有较大的优化。为简便起见,本文后面称此基于并行计算和稀疏存储的模糊积分算法为 DSFI 算法。下一章实验部分将对 DSFI 算法在模糊积分计算效率上的提升进行详细分析。

3 实验部分

3.1 并行计算在模糊积分上的应用

本次实验采用了 UCI^[24] 上的四个数据集,分别为 Shuttle, Poker, Letter 和 Mushroom 数据集。Shuttle 是航天飞机的数据集,其中包含 9 个整型数值型特征和 1 个分类特征,分类特征为航天飞行的 7 种状态;Poker 是扑克牌游戏的数据集,其中包含 10 个整型数值型特征和 1 个分类特征,该 10 个特征描述了手上拥有的 5 张牌的花色和牌号,而分类特征为 10 个预先定义好的 5 张牌的等级大小;Letter 是关于识别 26 个英文字母的数据集,其中包含 16 个整型数值型特征和 1 个分类特征,该 16 个特征描述了英文字母的各种坐标特征,而分类特征为 26 个英文字母;Mushroom 是关于识别各种蘑菇是否有毒的数据集,其中包含 22 个整型数值型的类别特征和 1 个分类特征,该 22 个类别特征描述了蘑菇的各种特性,而分类特征为蘑菇是有毒的还是无毒的。表 1 对数据集的记录数、特征数和类别数进行了描述。

实验 1-3 的集群由一个主节点、两个从节点组成, 附加一台 HDFS 数据服务器。两个从节点分别为节点 1 和节点 2, 配置 Intel Xeon E3-1240 四核处理器和 2 GB 内存, 使用 CentOS 6 x64 系统并安装 Spark 并行计算环境。表 2 描述了集群使用的核心个数和节点使用的核心个数之间的关系。当集群使用三个核心时, 节点 1 贡献两个核心, 节点 2 贡献一个核心, 即核心总个数为奇数时, 节点 1 贡献的核心数比节点 2 贡献的核心数多一。表 3 是对集群中节点 1 和节点 2 进行整型运算和浮点型运算的性能测试数据。5 千万次 (1×1) 运算和 2 千万次 $\sqrt{25}$ 运算是在 Python 语言环境下的测试, 节点 2 相对于节点 1 分别在整型运算和浮点型运算上提升 33% 和 44%。5 千万次 $\sum_{i=1}^n x_i$ 运算和 5 千万次 $\sqrt{1, \dots, \sqrt{10}}$ 运算是在使用 Cython 对程序进行优化后的测试结果。Cython 采用者编写 Python 的 C 扩展程序, 其运行效率接近 C 语言。在 Cython 优化下, 节点 1 和节点 2 的运行效率均大幅提高, 而节点 2 在整型运算和浮点型运算中, 相对于节点 1 分别提高了 35% 和 37%。实验数据表明, 节点 2 比节点 1 在计算性能上高出 30%~40%。本文提出的 PSFI 算法主要使用到整型运算, 并使用 Cython 对代码进行优化, 节点 2 相对于节点 1 在性能上提升了 35.4%。

表 1 数据集描述

数据集	记录数	特征数	分类数
Shuttle	43 500	9	7
Poker	1 500 000	10	10
Letter	20 000	16	26
Mushroom	8 124	22	2

表 2 集群中节点的性能测试

测试项目	节点 1	节点 2	提升
5 千万次 (1×1) 运算	3.197	2.128	0.334
2 千万次 $\sqrt{25}$ 运算	4.228	2.772	0.438
5 千万次 $\sum_{i=1}^n x_i$ 运算	6.020	3.917	0.354
5 千万次 $\sqrt{1, \dots, \sqrt{10}}$ 运算	3.255	2.209	0.322

a) 实验集群中, 节点之间的性能不能保证完全相同。节点 1 和节点 2 在性能上存在差异, 节点 2 的运行速度更快。Poker 数据集并行计算模糊积分方程系数的实验数据如表 4 所示。

表 4 Poker 数据集运行时间对比

对比项	核心数			
	1	2	3	4
耗时	120.045	52.499	38.518	28.927
总耗时	120.045	104.999	115.551	115.748

表 4 的结果表明, 随着集群核心数的增加, 数据转换的运行时间是下降的。但是, 两个核心的运行时间比一个核心快一倍多。理论上, 这并不合理, 因为两个核心相对于单个核心, 循环执行还会多一个进程间通信的开销, 所以两个核心的运行时间应该比单核心运行的时间接近两倍, 但小于两倍。由此说明, 实验集群采用的两个节点之间是存在差异的, 当 CPU 核心数为 2 个时, 节点 2 的 CPU 核心加入到集群, 性能有所提高。对比 CPU 核心个数 3 和 4, 也可以发现节点 2 的加入使集群的性能有所提高, 所以节点 2 比节点 1 的性能稍好, 节点 2 执行单个任务的运行时间会比节点 1 的运行时间短。集群之间的节点差异会对程序的运行时间造成一定影响, 如果集群存在性能较差的节点, 这个较差的节点可能会成为集群计算的瓶颈, 当最后一个任务恰好由较差的节点执行时, 整个计算任务的执

行结束时间会以这个较差节点的执行结束时间为任务的完成时间。

b) 集群的核心数越多越好并不适用于所有数据集。当任务的计算量较小, 核心数较多时, 整个计算时间的快慢取决于最慢执行的任务。

图 3 显示了 Shuttle 数据集并行计算模糊积分方程系数的实验数据。图 3 的结果也表明偶数核心集群比奇数核心集群运行时间的下降速度更快, 原因在于奇数核心采用的节点 1 的核心个数比节点 2 的核心个数多 1, 这也是节点 1 和节点 2 之间的差异造成的。当核心的个数大于等于 5 时, 奇数核心集群的运行时间随着核心数增加, 运行时间反而增加了, 这表明集群的核心数并不是越多越好。当数据集较小时, 增加核心数不一定能够降低运行时间。核心数越大, 进程的通信开销也增大, 而且进程的开销也与集群节点的性能有关; 当进程开销大于并行计算带来的效率时, 运行时间反而降低。当核心数等于 8 时, 运行时间最短, 这时最后一个任务恰好由节点 2 执行。因此, 在中小型数据集上, 应合理选取核心数, 特别是要考虑最慢节点完成任务的时间, 以提高集群整体的运行效率和节省集群资源。

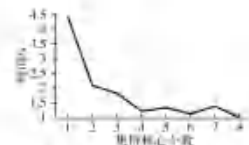


图 3 Shuttle 数据集不同核心数的运行时间对比

c) 一般情况, 并行程序的核心数增加, 并行程序的加速比增加, 程序的整体运行时间减少, 但并行程序的效率降低。

用 T_s 表示串行计算的运行时间, T_p 表示并行计算的运行时间, 用 N 表示 CPU 个数, S 表示加速比, E 表示并行程序的效率。则并行程序的加速比计算方法为 $S = T_s / T_p$; 并行程序的效率计算方法为 $E = \frac{S}{N}$ 。

使用 PSFI 算法对 Poker, Letter 和 Mushroom 三个数据集分别在 1、2、4、6 和 8 个核心下进行实验, 其中一个核心的实验在性能较好的节点 2 中进行, 实验的运行时间如表 5 所示。并行程序的加速比和效率计算结果如表 6 所示。

表 5 串行计算和并行计算运行时间

核心个数	Poker	Letter	Mushroom
1	81.620	215.256	6.264.599
2	52.499	114.41	3.573.232
4	28.917	60.959	1.953.009
6	21.079	44.501	1.367.393
8	18.957	34.86	1.041.325

表 6 并行计算的加速比和效率

核心个数	Poker		Letter		Mushroom	
	加速比	效率	加速比	效率	加速比	效率
2	1.55	0.78	1.89	0.94	1.78	0.89
4	2.82	0.71	3.53	0.86	3.26	0.81
6	3.54	0.59	4.84	0.81	4.65	0.78
8	3.51	0.54	6.17	0.77	6.11	0.76

表 5 的结果说明随着计算资源的增多, 模糊积分的计算时间接近呈线性减少。在 CPU 核心个数达到 6 个或 8 个时, 运行时间减少较慢, 原因在于进程之间的通信开销增加, 资源使用达到集群的峰值, 对程序运行造成一定的影响。CPU 核心数

越大,进程的开销越大;数据的计算量越大,进程的开销也越大。

表 6 的结果表明,核心个数增加,并行计算的加速比增加,效率降低。并行计算的效率降低是因为系统资源的开销增大。但对于大型数据集,并行任务对运行时间的减少量远高于进程开销所花的时间。在 100 万条记录 10 个特征的 Poker 数据集上,8 个核心并行计算花费 18.957 s,在 2 万条记录 16 个特征的 Letter 数据集上,8 个核心并行计算花费 34.88 s。这个时间是比较快的,当数据集增大时,本文还可以扩展集群的节点数,提高任务并行度。基于并行计算,本文在模糊积分的计算上所花的时间降低了几倍。在进行特征选取选择适合的参数时,具有较高的效率。

当集群节点性能接近,并行计算损失的效率降低。实验采用三个性能相近的节点,一个主节点安装 HDFS 和 Spark 主节点,两个从节点分别安装 Spark 从节点。三个节点均使用 CentOS 6 x64 系统,配置 4 核心 1.5630 GPU,频率 2.13 GHz。在内存方面,主节点配置 4 GB 内存,从节点配置 2 GB 内存。三个节点均为同一台西数 1.5630 GPU(每个 GPU 4 个物理核心)的 XeonServer 服务器下的虚拟机,其性能相同或非常接近。表 7 显示了其中一个从节点的性能测试数据。

表 7 集群中节点的性能测试

测试项目	集群节点
5 千万次(1+1)运算	1.334
2 千万次 $\sqrt{23}$ 运算	1.957
5 千万次 $\sum_{i=1}^n i$ 运算	0.006
2 千万次 $\sqrt{2^i}$ 运算	0.217

本实验的集群使用上购价较低的 CPU,运算能力相对前面实验使用的集群较弱,但节点间性能相近,对并行计算的加速比和效率计算的影响较小。本实验主要考虑对模糊积分计算中的运行效率,在实验前采取预先对数据进行字符转数值的处理。Poker 数据集本身是数值型数据,无须处理;Letter 数据集原先对 26 个字母进行字符转数值的处理,所以运行时间相对前一部分的实验较快。表 8 和 9 分别为 Poker 和 Letter 数据集的实验结果。

表 8 Poker 数据集实验结果

核心个数	计算时间	加速比	效率
1	180.744	1.000	1.000
2	96.685	1.869	0.935
3	66.702	2.707	0.902
4	51.278	3.498	0.875

表 9 Letter 数据集实验结果

核心个数	计算时间	加速比	效率
1	166.280	1.000	1.000
2	85.711	1.921	0.961
3	60.235	2.760	0.922
4	46.172	3.605	0.902

实验结果表明,在性能接近的集群上测试,并行计算的加速比和效率下降较为稳定。使用 4 个核心实验的加速比理想状态是 4,Poker 和 Letter 数据集分别下降了 0.51 和 0.59。在效率上,Poker 数据集和 Letter 数据集分别在 3 个核心和 4 个核心保持 0.9 的效率,损失的 0.1 效率消耗在系统和进程的开销上。当核心个数为 2 时,此时使用的节点个数为 2,系统间的开销较大,并行系统损失的效率也较大,Poker 和 Letter 数据集在 2 个核心的实验中,分别损失 0.065 和 0.039 的效率,此时应尽可能多地使用系统空闲资源来计算。由于集群资源上的不足,不能进行更高核心个数的实验,但集群资源越多,计算的并行度就越高,范围越大,对算法运行时间的效率提升越明显。在优化并行计算的效率上,可以对系统的网络,使调度进行优化,对集群调度的参数进行优化,减少系统和进程的开销,让更多的 CPU 资源都用于计算,减少并行计算在效率上

的损失。

文献[6]的实验表明 512 k 数据集比 428 k 数据集的加速比更高。对于聚类算法,数据集越大,每次迭代的计算次数越多,计算也越密集,因此加速比也较高。本文的 PSF 算法和文献[6]提到的基于算法在 Spark 上运行,加速比和效率较高,是因为这些算法都是计算密集型算法,而计算越密集的算法,并行程序的运行效率相对也较高。Letter 数据集有 16 个特征,每条数据计算 2¹⁶ 次,而 Poker 数据集有 10 个特征,每条数据计算 2¹⁰ 次,相对来说,计算次数一定时,Letter 数据集数据交换的次数更少。所以,并行计算在模糊积分或其他密集型算法的计算上有重要的应用。当数据集或者计算量达到单机不能处理的情况,并行计算对效率的提升是明显的,此时系统开销只占总资源使用量中很少一部分。

3.2 稀疏存储对模糊积分影响测试

模糊积分的计算存在高时间复杂度和高空间存储量问题。本节使用并行计算减少了计算的运行时间,本节将通过对比实验说明稀疏存储在模糊积分计算上的重要应用。

在模糊积分的算法中,随着特征数量的增加,特征值的个数呈指数级增长,而且特征值含有很多零值。表 10 列出了模糊积分算法中使用稀疏存储和稠密存储的存储空间对比。

表 10 模糊积分算法应用于不同数据集的存储空间对比

数据集	特征数	记录数	稀疏存储/MB	稠密存储/KB	效率
Shuttle	9	14 500	36.42	808.25	33.00
Poker	10	32 000	97.79	1.64	59.67
Letter	16	50	20	7.3	2.805.49
Mushroom	22	8	50	656	102.300.1

表 10 的结果表明,稀疏存储相对于稠密存储减少了大量的存储空间,在特征数较少的数据集上,稀疏存储能降低数十倍存储空间。在本例 Mushroom 数据集上有 22 个特征,仅 4 条记录,稠密存储需要 64 MB 空间,而稀疏存储仅占用了 656 B 空间,存储空间减少了 102.300 倍;在数据集较少的 Shuttle 数据集上,存储空间降低 33 倍。这对于基于内存计算的 Spark 框架来说,可以节省大量的内存空间。

稀疏存储在模糊积分算法上对存储空间的减少非常显著。随着特征数增加,存储空间也呈指数级增长,如果使用稠密存储,不管是用内存还是磁盘,效率都是十分低效的,而且还存在存储空间不足导致程序中断等问题。因此,稀疏存储是模糊积分计算的重要应用。

稀疏存储在模糊积分的计算上,不仅能够降低存储空间,还加快了数据训练的速度。表 11 显示了 Shuttle 数据集关于稀疏存储和稠密存储训练模型所花时间的对比,稀疏存储训练模型的时间为 0.631 s,比稠密存储快 3.6 倍。

表 11 Shuttle 数据集在不同存储模型下的训练时间对比

存储存储	训练时间
稀疏存储	0.631
稠密存储	2.295

3.3 并行模式下模糊积分的应用

本节使用本文提出的 PSF 算法对 UCI 上的 7 个数据集进行分类测试,并与 2015 年提出的 MNI^[27]算法进行实验对比,数据集的描述如表 12 所示。实验采用十折交叉验证,即将数据集分为十份,每次取其中一份作为测试集,其余九份作为训练集,重复十次,保证所有数据都能用于测试集,并取这十次实验结果的平均值作为最终的准确率,结果如表 13 所示。

表 12 经典数据集描述

数据集	已标数	未标特征数	分类数
Monk1	556	6	2
Monk2	603	6	2
Monk3	554	6	2
Heart	270	7	2
Vote	768	7	2
Wilt	500	18	2
Vote	629	4	2

表 13 PSFI 与 MFI 算法的实验结果对比

数据集	特征数	MFI	PSFI
Monk1	训练准确率	0.978	0.961
	测试准确率	0.948	0.937
Monk2	训练准确率	0.994	0.988
	测试准确率	0.688	0.980
Monk3	训练准确率	0.996	0.972
	测试准确率	0.992	0.976
Heart	训练准确率	0.867	0.864
	测试准确率	0.844	0.862
Vote	训练准确率	0.782	0.781
	测试准确率	0.752	0.764
Wilt	训练准确率	0.930	0.940
	测试准确率	0.958	0.947
Vote	训练准确率	0.961	0.957
	测试准确率	0.956	0.958

表 13 实验结果表明, 本文提出的算法在七个数据集中, Monk1、Monk2、Heart 和 Vote 数据集的测试准确率有较大提升, 特别是 Monk2 和 Heart 数据集, 而其他数据集的预测效果则相对接近。总的来说, 本文提出的算法在预测准确率上占据优势。

4 结束语

本文介绍了模糊积分在分类问题中的应用, 并结合 Spark 大数据框架, 提出一种优化的基于并行计算和稀疏存储的 PSFI 算法。本文提出的 PSFI 算法首次将并行计算和稀疏存储引入模糊积分, 提高了模糊积分的计算效率, 使模糊积分的研究可以扩展到更高维的大数据。

本文提出的算法有较高的效率和准确率, 但模糊积分在大数据应用中还有值得改进的方面: a) 将本文的算法应用到多分类问题, 应用到更多节点的集群和更高维的大数据集; b) 对并行计算的调度器进行动态调优, 减少最慢任务对整体计算完成时间的影响; c) 稀疏存储在模糊积分上有着重要的应用, 关于稀疏存储的最优化问题的相关算法也值得研究。

参考文献:

- [1] Apache Hadoop 2.6.4 [EB/OL]. (2016-02-11) (2016-08-11). <http://hadoop.apache.org/docs/r2.6.4/>.
- [2] Apache Spark overview 1.6.2 [EB/OL]. (2016-08-25) (2016-08-11). <http://spark.apache.org/docs/1.6.2/index.html>.
- [3] Gohlich M. The application of fuzzy integrals in multicriteria decision making [J]. *European Journal of Operational Research*, 1996, 89(3): 445-456.
- [4] Qiu Hongbin, Gu Rong, Yuan Guiming, et al. YAKM: a parallel frequent itemset mining algorithm with Spark [C]//Proc of IEEE International Parallel & Distributed Processing Symposium, 2014: 1664-1671.
- [5] Deng Jie, Qiu Zhiguo, Zhu Yanguo, et al. Transmit efficient and scalable data mining using Spark [C]//Proc of International Conference on Information and Communications Technology, 2014: 1-6.
- [6] Liu Chen, Liu Ruiqian, Chen Zhengzhang, et al. A scalable hierarchical clustering algorithm using Spark [C]//Proc of the 1st International Conference on Big Data Computing Service and Applications, Washington DC: IEEE Computer Society, 2015: 418-426.
- [7] Chen Dehua, Shen Changjun, Feng Jinyang, et al. An efficient parallel

big- k similarity join for massive multidimensional data using Spark [J]. *International Journal of Database Theory and Application*, 2015, 8(3): 57-68.

- [8] Kalogeropoulos A K, Yiparinos P, Tekiner E, et al. A parallel distributed Weka framework for big data mining using Spark [C]//Proc of IEEE International Congress on Big Data, 2015: 9-16.
- [9] Sugeno M. *Theory of fuzzy integrals and its applications* [D]. Tokyo: Tokyo Institute of Technology, 1975.
- [10] Mizumoto T, Sugeno M. A theory of fuzzy measures, representation, the Choquet integral, and null sets [J]. *Journal of Mathematical Analysis and Applications*, 1991, 159(2): 532-549.
- [11] Ishii K, Sugeno M. A model of human evaluation process using fuzzy measure [J]. *International Journal of Man-Machine Studies*, 1985, 22(1): 19-38.
- [12] Ohmura T, Sugeno M, Nishiyaki Y, et al. Fuzzy measure analysis of public attitude towards the use of nuclear energy [J]. *Fuzzy Sets and Systems*, 1989, 20(3): 259-289.
- [13] Tanaka K, Sugeno M. A study on subjective evaluation of color processing images [J]. *International Journal of Approximate Reasoning*, 1991, 5(5): 213-222.
- [14] Iwano K, Anzai T. A study on the industrial design evaluation based upon non-additive measures [C]//Proc of the 7th Fuzzy System Symposium, 1991: 521-524.
- [15] Washio T, Takahashi H, Kitamura M. A method for supporting decision making on plant operation based on human reliability analysis by fuzzy integral [C]//Proc of the 2nd International Conference on Fuzzy Logic and Neural Networks, 1992: 841-845.
- [16] Wang Zhenyuan, Liang K S, Wang M L, et al. A new type of nonlinear integrals and the computational algorithm [J]. *Fuzzy Sets and Systems*, 2000, 112(2): 223-231.
- [17] Xu Kabin, Wang Zhenyuan, Hong P A, et al. Classification by nonlinear integral projections [J]. *IEEE Trans on Fuzzy Systems*, 2003, 11(2): 182-200.
- [18] Wang Wei, Wang Zhenyuan, Kuo G J. Genetic algorithms for determining fuzzy measures from data [J]. *Journal of Intelligent & Fuzzy Systems*, 1998, 6(2): 171-183.
- [19] Wang Zhenyuan, Liang K S, Wang Jia. A genetic algorithm for determining nonadditive set functions in information basis [J]. *Fuzzy Sets and Systems*, 1999, 102(3): 463-469.
- [20] Wang Zhenyuan, Guo Hailong. A new genetic algorithm for nonlinear multi-regression based on generalized Choquet integrals [C]//Proc of the 12th IEEE International Conference on Fuzzy Systems, 2003: 819-821.
- [21] Wang Jiefeng, Lee K H, Liang K S. LI-norm regularization based nonlinear integrals [C]//Proc of International Symposium on Neural Networks, Berlin: Springer, 2009: 204-208.
- [22] Gader P D, Mohamed M A, Keller J M. Fusion of hammerstein word classifiers [J]. *Pattern Recognition Letters*, 1996, 17(6): 577-584.
- [23] Zhai Junhui, Wang Jiefeng, He Wenxiang. Combination of ONELM classifiers with fuzzy integral for large-scale classification [J]. *Journal of Intelligent & Fuzzy Systems*, 2015, 28(5): 2257-2268.
- [24] American A, Newman D J. UCI machine learning repository [EB/OL]. (2007-04-20). <http://mlacm.ics.ucr.edu/MLRepository.html>.
- [25] Wang Jiefeng, Liang K S, Lee K H, et al. Multiple nonlinear integral for classification [J]. *Journal of Intelligent & Fuzzy Systems*, 2015, 28(4): 1635-1645.

3.8. 基于蚁群算法求解 Choquet 模糊积分模型



山东大学学报(工学版)

Shandong Daxue Xuebao (Gongxue Ban)

第48卷第3期

2018年6月

目次

机器学习与数据挖掘

- 基于正反向异质性的遥感图像变化检测..... 李士进,王声特,黄乐平(1)
- 基于人工蜂群和 SVM 的基因表达数据分类..... 叶明全,高凌云,万春圆(10)
- 电商商品嵌入表示分类方法..... 龙柏,曾宪宇,李微,刘淇(17)
- 基于 SVD 和 DCNN 的彩色图像多功能掌水印算法..... 赵彦霞,王熙照(25)
- 基于卷积神经网络的中文财经新闻分类方法..... 谢志峰,吴佳萍,马利庄(34)
- 基于词向量的领域情感词共构建..... 林江豪,周咏梅,阳爱民,陈锦(40)
- 基于核 K-means 的增量多视图聚类算法..... 张佩瑞,杨燕,郝继来,喻琰(48)
- 基于 HBase 和 SimHash 的大数据 K-近邻算法..... 王婷婷,崔俊海,张明阳,郝璞(54)
- 深度卷积神经网络嵌套 fine-tune 的图像美感品质评价..... 李雨鑫,普园镜,徐丹,钱文华,刘和娟(60)
- 一种长文本辅助短文本的文本理解方法..... 阎盈盈,黄瑞章,王瑞,马灿,刘博伟,黄庭(67)
- 有序稀疏惩罚的一维下网问题..... 梁泽华,崔耀东,张雨(75)
- 基于双群算法求解 Choquet 模糊积分模型..... 陈嘉杰,王金凤(81)
- 一种集成卷积神经网络和深信网的步态识别与模拟方法..... 何正文,曾宪华,郭姜(88)
- 一种融合社交网络的叠加联合聚类推荐模型..... 谈习习,刘华锋,景丽萍(96)
- 一种基于加权图模型的手指静脉识别方法..... 叶子云,杨金锋(103)
- 压缩感知重构算法的并行化及 GPU 加速..... 何文杰,何伟超,孙叔森(110)
- 基于 BFOA 和 K-means 的复合入侵检测算法..... 肖苗苗,魏本征,尹义龙(115)
- 面向短文本情感分析的词扩充 LDA 模型..... 沈冀,马志强,李国雅,张力(120)
- 基于单调约束的径向基函数神经网络模型..... 曹雅,邓赵红,王士同(127)
- 一种基于聚类的过抽样算法..... 王换,周忠盾(134)
- 非均匀数据的变异性系数聚类算法..... 杨天鹏,徐鲲鹏,陈黎飞(140)

责任编辑:陈燕 封面设计:山东大学王全军

[期刊基本参数] CN37-1391/T * 1956 * b * A4 * 146 * zh * P * ¥ 10.00 * 1000 * 21 * 2018-06

基于蚁群算法求解 Choquet 模糊积分模型

陈嘉杰, 王金凤*

(华南农业大学数学与信息学院, 广州 510642)

摘要: 为了提高 Choquet 模糊积分模糊测度的搜索效率, 提出改进的蚁群算法求解模型, 将迭代次数大大降低。根据特征数量构建 Choquet 模糊积分模型, 搜索过程中对每只蚂蚁按状态转移概率进行全局搜索或局部搜索, 迭代搜索最优解, 最后由 Fisher 判别进行分类。试验使用 3 组癌症基因数据集, 利用 R 语言的 Bioconductor 工具箱进行数据预处理, 并分析对比新模型和主流算法的分类效果。结果表明, 在 DLBCL 数据集和 Colon 数据集中, 基于蚁群算法的 Choquet 模糊积分得到最好的分类效果; 在 Prostate 数据集中, 虽然和基于遗传算法的 Choquet 模糊积分分类效果接近, 但是蚁群算法仍然更快收敛, 验证了改进的蚁群算法可以作为一种求解模糊测度的快速方法。
关键词: Choquet 模糊积分; 模糊测度; 蚁群算法; 遗传算法; 癌症分类

中图分类号: TP399 **文献标志码:** A

Method for solving Choquet integral model based on ant colony algorithm

CHEN Jiajie, WANG Jinfeng*

(College of Mathematics and Information, South China Agricultural University, Guangzhou 510642, China)

Abstract: An improved ant colony algorithm for Choquet integral was investigated to enhance the search efficiency of fuzzy measure. Choquet integral model was built according to the characteristic quantity and solved by the process of searching globally or locally according to the state transition probability. Finally, it was classified by Fisher discriminates. The experiment used three sets of cancer gene datasets preprocessed by R language Bioconductor toolkit, and analyzed classification results between new model and the mainstream algorithm. The results showed that in DLBCL dataset and Colon dataset, ant colony algorithm had the best effect; in Prostate dataset, although the classification results were about the same, ant colony algorithm still had faster convergence than Genetic Algorithm. Therefore, the improved ant colony algorithm presented a feasible and effective way to solve fuzzy measures in Choquet integral model.

Keywords: Choquet integral; fuzzy measures; ant colony algorithm; Genetic algorithm; cancer classification

0 引言

在实际生活中, 很多问题并不能用简单的线性相加解决, 比如工人的工作效率问题, 工人 A 和工人 B 一同工作, 两人的工作效率不能简单地

用工人 A 的效率和工人 B 的效率相加, 因为他们之间会相互影响。模糊测度^[1], 也称为非可加测度, 能够很好地表示各个特征以及特征组合对决策的贡献度, 每一个模糊测度对应一个特征集合

收稿日期: 2017-05-09 网络出版时间:

网络出版地址: <http://www.cnki.net>

基金项目: 国家自然科学基金资助项目(61202295); 广东省公益研究与能力建设基金资助项目(2017A040406023); 广东省公益研究与能力建设基金资助项目(2015A030401081)

作者简介: 陈嘉杰(1993-), 男, 广东东莞人, 硕士, 主要研究方向为模糊积分 E-mail: chen_jia_jie@scau.edu.cn

*通讯作者: 王金凤(1978-), 女, 河北黄骅人, 副教授, 博士, 主要研究方向为数据挖掘、机器学习 E-mail: wangjinfeng@163.com

的子集, 对于有 n 个特征的数据集, 模糊测度的个数为 $2^n - 1$ 。

模糊积分^[3]是近 20 年来出现的新的有效的分类器, 尤其适用于特征之间存在相互影响的情况, 它是基于模糊测度的非线性积分。1901 年, 文献[3]提出 Choquet 模糊积分。2000 年, 文献[4]提出 Zhenyuan 模糊积分。模糊积分理论不断扩展延伸。Choquet 积分是 Lebesgue 积分的严格推广, 而且计算相对简单, 因此广泛应用于实际问题中。在疾病分类^[5]、害虫预测^[6]、物流供应链选择决策^[7]以及企业的绩效评估^[8]等实际问题中取得很好的效果。文献[9]等提出基于 Choquet 积分的非线性回归模型, 并在往后的研究中给出一个基于遗传算法的求解方案^[10]。使用遗传算法搜索 Choquet 积分模型的解是一个确定模糊测度的过程, 虽然能取得一定的效果, 但是收敛速度慢, 搜索十分耗时。因此本研究提出使用蚁群算法来求解 Choquet 积分模型参数, 给出搜索过程的算法, 最后使用几个经典的疾病分类的数据集, 通过试验来比较蚁群算法和遗传算法的结果差异, 并且也和其他常用的分类器进行比较。

1 模糊积分概述

1.1 模糊测度

经典测度也称可加测度, 具有可加性, 比如容器 A 和容器 B 的体积, 他们的总体积等于两者之和。与可加测度相对, 模糊测度具有不可加性, 这里给出模糊测度的严格定义^[1]。

定义 1 设 X 是任意集合, \mathcal{F} 是 X 的子集构成的 σ 代数, 集函数 $\mu: \mathcal{F} \rightarrow [0, 1]$ 当且仅当满足以下条件时, 称为定义在 \mathcal{F} 的模糊测度:

- (1) 平凡性 若 $\emptyset \in \mathcal{F}$, 那么 $\mu(\emptyset) = 0$;
- (2) 单调性 若 $E \in \mathcal{F}, F \in \mathcal{F}, E \subset F$, 则有 $\mu(E) \leq \mu(F)$;
- (3) 下连续性 如果 $E_n \in \mathcal{F}(n = 1, \dots, \infty), E_1 \subset E_2 \subset \dots, \cup_{n=1}^{\infty} E_n \in \mathcal{F}$, 那么 $\lim_n \mu(E_n) = \mu(\cup_{n=1}^{\infty} E_n)$;
- (4) 上连续性 如果 $E_n \in \mathcal{F}(n = 1, \dots, \infty), E_1 \supset E_2 \supset \dots, \mu(E_1) < \infty, \cap_{n=1}^{\infty} E_n \in \mathcal{F}$, 那么有

$$\lim_n \mu(E_n) = \mu(\cap_{n=1}^{\infty} E_n)。$$

定义 2^[1] 给定集函数 $\mu: \mathcal{F} \rightarrow (-\infty, +\infty)$, 如果 $\mu(\emptyset) = 0$, 则称作符号型模糊测度。

定义 1 中模糊测度的非负性会对一些实际问题产生限制; 定义 2 定义更一般的模糊测度, 允许模糊测度为负数, 因此, 模糊测度可以看成是符号型模糊测度的一个特例。

1.2 Choquet 模糊积分

常见的模糊积分有 Sugeno 模糊积分^[11], Choquet 模糊积分^[1]和 Zhenyuan 模糊积分^[12]。这里只对 Choquet 模糊积分做介绍, 并且后面提到模糊积分不加说明都是指 Choquet 模糊积分。

定义 3^[11] 设 (X, \mathcal{F}) 是可测空间, μ 是 \mathcal{F} 上符号型模糊测度, 实值函数 $f: X \rightarrow (-\infty, +\infty)$ 关于 μ 的 Choquet 模糊积分定义如下:

$$(c) \int f d\mu = \int_{-\infty}^0 [\mu(E_x) - \mu(X)] dx + \int_0^{\infty} \mu(E_x) dx, (1)$$

其中: $\mu(E_x)$ 为符号型模糊测度, $E_x = \{x | f(x) \geq x, x \in X\}$, 对任意的 $\alpha \in (-\infty, +\infty)$ 称作 α 截集。

当 X 是一个有限集合时候, 即 $X = \{x_1, x_2, \dots, x_n\}$, 为了方便计算, 将 f 值 $f(x_1), f(x_2), \dots, f(x_n)$, 按照非递减的顺序排列, 如 $f(x_1^*) \leq f(x_2^*) \leq \dots \leq f(x_n^*)$, 其中 $(x_1^*, x_2^*, \dots, x_n^*)$ 是 (x_1, x_2, \dots, x_n) 的一个重排列, Choquet 模糊积分可以表示成:

$$(c) \int f d\mu = \sum_{j=1}^n [f(x_j^*) - f(x_{j-1}^*)] \mu(\{x_j^*, x_{j+1}^*, \dots, x_n^*\}), (2)$$

其中: $f(x_0^*) = 0$ 。为了方便计算, 文献[10]提出等价公式计算 Choquet 模糊积分:

$$(c) \int f d\mu = \sum_{j=1}^{2^n-1} z_j \mu_j, (3)$$

其中:

$$z_j = \begin{cases} \min(\frac{j}{2^n}, \text{frc}(\frac{j}{2^n})) f(x_j) - \max(\frac{j}{2^n}, \text{frc}(\frac{j}{2^n})) f(x_j) & \text{if } \mu > 0 \\ 0 & \text{otherwise} \end{cases}$$

上式中, $j = 1, 2, \dots, 2^n - 1$; $\text{frc}(\frac{j}{2^n})$ 是分数 $\frac{j}{2^n}$ 的小数部分的数值。

1.3 Choquet 模糊积分分类

对于给定的一组样本各个特征值, 和一组模糊测度值, 利用公式(3), 即可求出每一个样本对

应的 Choquet 模糊积分值。对于这组模糊积分值,可以使用 Fisher 线性判别进行分类^[11],其基本思想是,寻找一个最佳的投影方向,使得样本在投影后,类内方差尽可能小,类间的方差尽可能大。Fisher 判别函数:

$$g(x) = w^T x + b, \quad (4)$$

当 $g(x) > 0$ 时,模糊积分值归类为第一类;当 $g(x) < 0$ 时,模糊积分值归类为第二类。公式(4)中, w^T 是分类面的法向量, b 为分类面的偏移量。

2 蚁群算法

文献[12]通过模拟蚁群的觅食行为提出蚁群算法。除了在离散空间问题,在连续空间优化问题也适用。本研究参考已有的求解连续空间约束优化问题的蚁群算法^[13],对蚁群算法做进一步改进,并应用于基于 Choquet 模糊积分模型的模糊测度的求解,再代入到 Choquet 模糊积分公式中,进行分类。

2.1 算法流程

(1)初始化蚁群算法相关变量,设置蚂蚁数目 m , m 只蚂蚁随机分布,作为各自蚂蚁搜索的起点。第 i 只蚂蚁所代表的解为 $M_i = (\mu_1^i, \mu_2^i, \dots, \mu_p^i, \dots, \mu_n^i)$, 其中 $\mu_j^i \in [a, b]$, $j = 1, 2, \dots, n$, 对每一只蚂蚁的具体值进行初始化:

$$\mu_j^i = a + (b - a) \times \text{rand}(1), \quad (5)$$

其中, a 表示模糊测度的最小取值, b 表示模糊测度的最大取值, $\text{rand}(1)$ 表示区间 $[0, 1]$ 之间的随机数。设置蚂蚁 i 所在位置的信息素强度初始值为:

$$T(i, 1) = \text{fitness}(M_i). \quad (6)$$

$\text{fitness}()$ 是一个适应度函数,信息素强度初始值直接取决于蚂蚁对应的适应度。这个适应度函数是按照以下步骤进行计算的:

①根据当前蚂蚁 $M(i)$ 所表示的模糊测度,计算出每个样本对应的 Choquet 模糊积分值;

②使用 Fisher 线性判别对 Choquet 模糊积分值进行分类;

③根据分类结果计算出分类的准确度 accuracy 和灵敏度 sensitivity , $\text{fitness}()$ 函数值由这两个指标按照一定的权重比例决定:

$$\text{fitness}(M_i) = w_1 \times \text{accuracy}(M_i) \times \text{sensitivity}(M_i), \quad (7)$$

其中, $w_1 \in [0, 1]$, $w_2 \in [0, 1]$, $w_1 + w_2 = 1$ 。

(2)计算第 t 次迭代蚂蚁 i 的状态转移概率 $P(i, t)$, 状态转移概率 $P(i, t)$ 的计算式如下:

$$P(i, t) = \frac{\exp(T_{\text{best}}^t - T(i, t))}{\exp(T_{\text{best}}^t)}, \quad (8)$$

其中, T_{best}^t 表示当前第 t 次迭代中,信息素强度的最大值。从式子中观察可知,蚂蚁所在点的信息素强度 $T(i, t)$ 越强,状态转移概率 $P(i, t)$ 越小。结合本题,状态转移概率它实际上是控制着蚂蚁搜索下一个解的策略。状态转移概率大的,则会以较大的概率使用全局搜索的策略;状态转移概率小的,则会以较大的概率使用局部搜索的策略。

(3)根据状态转移概率 P 搜索各个蚂蚁的下一个位置,将搜索结果保存到临时数组 temp 。首先需要预先设置一个状态转移阈值 P_0 , 搜索步长 step , 和步长参数 λ , 步长参数 λ 是由当前迭代次数决定的:

$$\lambda = \frac{1}{t}. \quad (9)$$

随着迭代次数的增加,步长参数越来越小,蚂蚁搜索的跨度也就变小,利于收敛。对于蚂蚁 i , 若 $P(i, t) \leq P_0$, 则进行局部搜索,局部搜索的步长为 step , 局部搜索算法如下:

$$\text{temp}(i, j) = \begin{cases} M(i, j) + \text{step} \times \lambda, & \text{rand}(1) < 0.5 \\ M(i, j) - \text{step} \times \lambda, & \text{else} \end{cases} \quad (10)$$

若 $P(i, t) \geq P_0$ 则进行全局搜索,全局搜索算法如下:

$$\text{temp}(i, j) = \begin{cases} M(i, j) + (b - a) \times 0.5, & \text{rand}(1) < 0.5 \\ M(i, j) - (b - a) \times 0.5, & \text{else} \end{cases} \quad (11)$$

值得注意的是,无论在全局搜索还是在局部搜索,若 $\text{temp}(i, j)$ 的搜索结果超出 $[a, b]$ 范围,则需要做越界处理,比 a 小则重置成 a , 比 b 大的则重置成 b 。从式子(10)和(11)可以看出,蚂蚁的状态转移概率比较小的时候,当前点的信息素强度较大,进行局部搜索,有利于能够在该点附近找到更优解;蚂蚁的状态转移概率比较大的时候,当前点的信息素强度较低,进行全局搜索,加快搜索的步长和随机性,加快寻优。

(4)蚂蚁的位置更新。搜索出下一个位置后,还需要判断蚂蚁 i 需不需要移动到该位置,判断的式子如下:

$$M_i = \begin{cases} M_i, & \text{fitness}(M_i) < \text{fitness}(\text{temp}_i) \\ \text{temp}_i, & \text{fitness}(\text{temp}_i) \geq \text{fitness}(M_i) \end{cases} \quad (12)$$

式子中,如果搜索的下一个位置蚂蚁的适应度比当前的适应度还要低,则放弃移动,否则就会移动到下一个位置。

(5) 信息素强度更新:信息素挥发系数 ρ , $\rho \in (0,1)$ 。参数 ρ 的取值直接影响到整个蚁群算法的收敛速度和全局搜索性能。 ρ 较小时,搜索的随机性就会增强; ρ 较大时,能够加快收敛速度。当蚂蚁 i 移动后,需要进行信息素强度更新的操作。信息素强度更新式子:

$$T(i,t+1) = (1-\rho) \times T(i,t) + \text{fitness}(M_i), \quad (13)$$

信息素挥发系数更新式子:

$$\rho = 0.1 \times \exp\left(\frac{t}{T}\right) + \ln 9 \quad (14)$$

(6) 经过 T 次迭代后,选出适应度最大的蚂蚁 M_{best} 。 M_{best} 表示当前符合 Choquet 模糊积分模型的一组最优模糊测度。

2.2 算法总结

根据 2.1 描述的具体算法流程,这里对算法进行步骤总结

Step1 初始化参数,包括蚂蚁数量 m ,需要估计的模糊测度个数 n ,模糊测度区间 $[a,b]$,迭代次数 times ,状态转移阈值 P_0 ,搜索步长 step ,信息素挥发系数 ρ ;

Step2 使用式子(5),(6),(7)初始化蚂蚁集合 $M(i,j)$ 和信息素强度 $T(i,1)$,其中 $i = 1,2,\dots,n$, $j = 1,2,\dots,m$;

Step3 迭代 times 次,更新状态转移概率,蚂蚁集合,信息素强度;

Step3.1 使用式子(8)计算状态转移概率 $P(i,t)$,其中 $t = 1,2,\dots,\text{times}$;

Step3.2 计算 $\text{temp}(i,j)$,使用式子(10)和式子(11)进行局部搜索或全局搜索;

Step3.3 使用式子(12)更新集合 $M(i,j)$;

Step3.4 使用式子(13)更新信息素强度 $T(i,t+1)$,式子(14)更新信息素挥发系数;

Step4 迭代结束,使用式子(7)选出适应度最大的蚂蚁 M_{best} 。

2.3 算法改进思路

对于本研究改进的蚁群算法中,最大的改进是每一只蚂蚁都能进行全局搜索和局部搜索,这取决于状态转移概率。而改进前的蚁群算法^[10],只允许当前最优的蚂蚁进行局部搜索。在实践的

过程中发现,除了当前最优解,当前次最优解如果也进行局部搜索,会有很大可能搜索到一个更优解。因此,本研究中的算法允许每一只蚂蚁根据状态转移概率随机选用局部搜索,是为了尽可能地找出更优解。

3 仿真试验

Choquet 模糊积分模型模糊测度的求解也是一个在连续空间上的优化问题,为了分析遗传算法和改进后的蚁群算法在连续空间上的优化问题,使用如下简单的二元函数做仿真测试:

$$f(x) = 100 \times (x_2 - x_1^2)^2 + (1 - x_1)^2,$$

其中 $x_1 \in [-0.5,0.5]$, $x_2 \in [-0.5,0.5]$,已知最小值为 0.25。分别使用遗传算法和本研究使用的蚁群算法寻找这个二元函数的最小值。

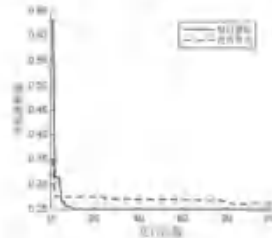


图1 使用蚁群算法和遗传算法求解目标函数最小值

Fig.1 Solving the minimum of the objective function by ant colony algorithm and Genetic Algorithm

试验中,遗传算法中个体编码使用实数编码,迭代次数取 100,种群规模取 100,交叉概率取 0.7,变异概率取 0.2;蚁群算法中迭代次数取 100,蚂蚁数目取 100,搜索步长取 0.1。试验结果如图 1,遗传算法在 100 次迭代内仅能找到一个近似的最优值,而蚁群算法在第 10 次迭代就找到最优值,收敛性要好于遗传算法。虽然遗传算法可以通过增加迭代次数和群体规模获取更优解,但是这样运算量就会大大增加。因此在同等条件下,由于蚁群算法的收敛性要好于遗传算法,因此蚁群算法在连续空间上寻找最优值的搜索效率要比遗传算法高,用时更少。

4 试验设计与分析

4.1 试验数据集

本试验将 Choquet 模糊积分应用在癌症分类问题上,并使用改进后的蚁群算法求解模型中的模糊测度。选用 DLBCL^[14]、Prostate^[15]和 Colon^[16]这3组经典癌症基因数据集作为试验分析对象,数据集情况请见表1、2、3。

表1 DLBCL数据集

基因数	样本数	DLBCL	FL
7129	77	58	19

表2 Prostate数据集

基因数	样本数	正常	病变
12690	102	50	52

表3 Colon数据集

基因数	样本数	正常	病变
2000	62	22	40

本研究使用R语言的Bioconductor工具箱对芯片癌症数据集进行一系列预处理^[17],包括背景校正,标准化和汇总,预处理后会得到一个基因表达矩阵,然后进行基因表达差异的显著性分析,使用经验贝叶斯的分析方法^[18],最后识别出显著变化的基因,并根据变化程度进行降序排列,对于这些显著变化的基因,这里分别取前3个,前5个,前8个特征进行重复试验。

数据集分组使用10折交叉,把数据随机分成10份,每次使用9份进行模型训练,剩下1份进行模型的试验验证,获取准确度,循环进行10次试验,最后取平均值。

4.2 算法参数

蚁群算法中不同参数的取值组合对试验结果影响很大,受数据集类型的影响。本试验中蚁群算法的参数取值如表4。

表4 蚁群算法的参数设置

蚂蚁数量	模糊测度区间	迭代次数	状态转移阈值	搜索步长
20	[0,1]	5	0.5	0.1

4.3 试验结果与分析

试验中,使用决策树(DT)^[19]、SVM^[20]、KNN^[21]等分类算法和模糊积分(分别基于遗传算法实现的FI-GA和基于蚁群算法实现的FI-ACO),模糊积分中模糊测度的求解分别使用改进的遗传算法^[22]和蚁群算法两种方案。在使用遗传算法解决模糊积分问题中,文献[9]提出针对此类问题的染色体编码方法,适应度评价方法和遗传算子。对本研究中使用的癌症基因数据集分类,只会出现4种结果^[1]:真正类,诊断结果为患病,实际情况也是患病;假正类,诊断结果为患病,实际情况是正常;真负类,诊断结果为正常,实际情况也是正常;假负类,诊断结果为正常,实际情况是患病。令TP、FP、TN和FN分别表示真正类、假正类,真负类和假负类的数目,本研究将采用准确度和敏感度作为分类指标,公式如下:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{sensitivity} = \frac{TP}{TP + FN}$$

和准确度相比,灵敏度在癌症分类中具有更加积极的意义,因为高灵敏度代表着对患病样本误判的概率很低。针对不同的数据库,测试结果分别如表5、6、7所示。

表5 DLBCL数据集试验

特征数	指标	DT	SVM	KNN	FI-GA	FI-ACO
3特征	CPU时间	0.55s	0.52s	0.47s	331s	15s
	训练准确度	1.000	0.980	0.940	0.987	0.950
	测试准确度	0.905	0.914	0.924	0.948	0.975
	测试灵敏度	0.940	0.901	0.935	0.900	0.950
5特征	CPU时间	0.59s	0.54s	0.45s	1077s	40s
	训练准确度	1.000	0.973	0.934	0.990	0.956
	测试准确度	0.910	0.920	0.904	0.948	0.980
	测试灵敏度	0.941	0.930	0.912	0.950	0.930
8特征	CPU时间	0.66s	0.59s	0.49s	9587s	178s
	训练准确度	1.000	0.978	0.932	0.993	0.940
	测试准确度	0.910	0.951	0.920	0.948	0.970
	测试灵敏度	0.942	0.932	0.921	0.950	0.950

表6 Prostate数据集试验

Table 6 Experiments of Prostate dataset

特征数	指标	DT	SVM	KNN	FI-GA	FI-ACO
3特征	CPU时间	0.6s	0.56s	0.43s	405s	20s
	训练准确度	1.000	0.953	0.923	0.942	0.932
	训练灵敏度	1.000	0.936	0.880	0.980	0.969
	测试准确度	0.901	0.912	0.921	0.922	0.920
	测试灵敏度	0.872	0.923	0.870	0.960	0.950
4特征	CPU时间	0.65s	0.57s	0.44s	1324s	30s
	训练准确度	1.000	0.943	0.923	0.949	0.920
	训练灵敏度	1.000	0.923	0.848	0.980	0.950
	测试准确度	0.902	0.903	0.895	0.895	0.920
	测试灵敏度	0.907	0.872	0.830	0.980	0.970
8特征	CPU时间	0.68s	0.58s	0.44s	1371s	20s
	训练准确度	0.992	0.960	0.932	0.967	0.963
	训练灵敏度	0.995	0.930	0.870	0.978	0.980
	测试准确度	0.920	0.930	0.927	0.913	0.910
	测试灵敏度	0.921	0.910	0.860	0.920	0.940

表7 Colon数据集试验

Table 7 Experiments of Colon dataset

特征数	指标	DT	SVM	KNN	FI-GA	FI-ACO
3特征	CPU时间	0.45s	0.35s	0.5s	267s	12s
	训练准确度	0.890	0.920	0.921	0.878	0.900
	训练灵敏度	0.783	0.810	0.818	0.947	0.950
	测试准确度	0.792	0.840	0.845	0.824	0.860
	测试灵敏度	0.650	0.800	0.760	0.925	0.975
5特征	CPU时间	0.57s	0.39s	0.51s	864s	27s
	训练准确度	0.910	0.920	0.850	0.901	0.880
	训练灵敏度	0.790	0.800	0.660	0.992	1.000
	测试准确度	0.796	0.862	0.830	0.793	0.861
	测试灵敏度	0.620	0.770	0.600	0.900	0.950
8特征	CPU时间	0.59s	0.39s	0.51s	7501s	199s
	训练准确度	0.990	0.930	0.840	0.909	0.900
	训练灵敏度	0.980	0.810	0.667	0.983	0.960
	测试准确度	0.800	0.860	0.845	0.774	0.870
	测试灵敏度	0.590	0.783	0.688	0.925	0.970

表5为DLBCL数据集试验结果,从表格数据中,可以分析出以下几点:

(1)在DLBCL数据集中,模糊积分CPU时间是要比另外三种算法要长,但是,模糊积分的测试准确度和测试灵敏度明显高于另外三种分类算法;

(2)使用蚁群算法求解的模糊积分准确性要

比使用遗传算法的要略高,同时在时间上也要比遗传算法的要快。

表6为Prostate数据集试验结果,这个表格中的数据 displays,模糊积分总体上比另外三种分类算法略高,这方面优势不明显,但是在测试灵敏度上,模糊积分仍然能够保持着高灵敏度。这意味着更少地将患者分类为正常,因此,模糊积分表现出的高灵敏度在这个领域中有积极作用。在模糊积分的实现上,蚁群算法和遗传算法的分类结果差异不大,但是在CPU运算时间上蚁群算法明显低于遗传算法。

表7为Colon数据集试验结果,这个试验中 can be found,从总体来说,使用蚁群算法实现的模糊积分其分类效果更加好,而使用遗传算法实现的模糊积分仅仅表现出灵敏度上的优势,而没有保证在准确度上的优势。

从这几个试验中,可以发现,模糊积分运算时间较长,但是表现出较好的准确度和较高灵敏度,在特征数量较少模糊积分是一个很好的选择。而在实现方式上比较,分别使用蚁群算法和遗传算法求解模糊测度,在分类结果中,没有必然性,也就是说不能给出使用蚁群算法实现比使用遗传算法实现的分类效果会更好这样的结论,因为有可能分类结果会很接近。但是在收敛性方面,蚁群算法要比遗传算法表现得更加好,因此时间消耗上是明显比遗传算法要少。

5 结论

本研究在已有的蚁群算法上做改进,并应用到 Choquet 模糊积分中模糊测度的求解。实际运算中发现非最优解的基础上做全局搜索也可能得到更优解,因此在算法中做出改进,允许每一只蚂蚁有一定的概率做全局搜索。试验中分别基于遗传算法和改进后的蚁群算法构建 Choquet 模糊积分的分类器,应用于癌症基因数据集,结果表明,在求解模糊测度这个问题上,蚁群算法比遗传算法具有更快的搜索解的速度,甚至会得到更好的分类结果。在仿真试验中也证明,蚁群算法的收敛性要好于遗传算法,大大降低搜索的迭代次数。在后续的工作中,我们将研究更多的不同领域的数据集,尝试用不同的参数值组合进行测试,探究参数对分类效果的具体影响。

参考文献:

- [1] WANG Z, KLIR G J. Fuzzy Measure Theory[J]. Springer Berlin, 1992, 35(1-2): 3-10.
- [2] SUGENO M. Fuzzy measures and fuzzy integrals—a survey[J]. Readings in Fuzzy Sets for Intelligent Systems, 1993, 6: 251-257.
- [3] MIYOFUSHI T, SUGENO M. A theory of fuzzy measures: Representations, the Choquet integral, and null sets[J]. Journal of Mathematical Analysis & Applications, 1991, 159(2): 532-549.
- [4] WANG Z, LEUNG K S, WONG M L, et al. A new type of nonlinear integrals and the computational algorithm[J]. Fuzzy Sets & Systems, 2000, 112(2): 223-231.
- [5] LEUNG K S, LEE K H, WANG J F, et al. Data Mining on DNA Sequences of Hepatitis B Virus[J]. IEEE/ACM Transactions on Computational Biology & Bioinformatics, 2011, 8(2): 428-440.
- [6] 冯慧敏, 闫雄, 李雪非. 基于 Choquet 积分的非线性虫洞问题[J]. 湖北农业科学, 2013, 52(22): 5485-5487.
- FENG Huimin, YAN Wei, LI Xuefei. Non-linear prediction of insects based on Choquet integral[J]. Hubei Agricultural Sciences, 2013, 52(22): 5485-5487.
- [7] 秦娟, 李延来, 陈振恩. 基于极大熵配置模型与 Choquet 积分的物流配送网络选择决策方法[J]. 计算机集成制造系统, 2015, 21(10): 2746-2759.
- QIN Juan, LI Yanlai, CHEN Zhensong. Group decision making method for supplier selection based on maximum entropy-optimization model and Choquet integral[J]. Computer Integrated Manufacturing Systems, 2015, 21(10): 2746-2759.
- [8] 王文刚, 陈维蒙, 林明夫. 基于 Choquet 积分的绩效评估模型研究—以建筑企业为例[J]. 中国海洋大学学报(社会科学版), 2015(5): 79-85.
- WANG Wenzhou, SHI Limeng, LIN Zhen. A study on the model of performance-evaluation based on Choquet integral: a case study of construction enterprise. Periodical of Ocean University of China, 2015(5): 79-85.
- [9] WANG Z, LEUNG K S, WONG M L, et al. Nonlinear nonnegative multiregressions based on Choquet integrals[J]. International Journal of Approximate Reasoning, 2000, 25(2): 71-87.
- [10] WANG Z, GUO H F. A new genetic algorithm for nonlinear multiregressions based on generalized Choquet integrals[C]//The IEEE International Conference on Fuzzy Systems. IEEE, 2003: 819-821. vol 2.
- [11] YANG R, WANG Z, HENG P A, et al. Fuzzy numbers and fuzzification of the Choquet integral[J]. Fuzzy Sets & Systems, 2005, 153(1): 95-113.
- [12] DORIGO M, GAMBARDILLA L M. Ant colony system: a cooperative learning approach to the traveling salesman problem[J]. IEEE Transactions on Evolutionary Computation, 1997, 1(1): 53-66.
- [13] 熊留成, 邵创创, 程志平. 一种求解连续空间约束优化问题的蚁群算法[J]. 郑州大学学报(工学版), 2015, 36(1): 20-23.
- XIAO Liucheng, SHAO Chuangchun, CHENG Zhiping. Ant colony algorithm for solving continuous space constrained optimization problems[J]. Journal of Zhengzhou University(Engineering Science), 2015, 36(1): 20-23.
- [14] SHIPP M A, ROSS K N, TAMAYO P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning [J]. Nature Medicine, 2002, 8(1): 68-74.
- [15] SINGH D, FEBHO P G, ROSS K, et al. Gene expression correlates of clinical prostate cancer behavior[J]. Cancer Cell, 2002, 1(2): 203-209.
- [16] ALON U, BARKAI N, NOTTERMAN D A, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays[J]. Proceedings of the National Academy of Sciences of the United States of America, 1999, 96(12): 6745.
- [17] GAO S, OU J, KAI X. R language and Bioconductor in bioinformatics applications(Chinese Edition)[M]. Tianjin: Tianjin Science and Technology Translation and Publishing Co, 2014: 106-150.
- [18] SMYTH G K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments [J]. Stat Appl Genet Mol Biol, 2004, 3(3): e3.
- [19] QUINLAN J R. Induction on decision tree[J]. Machine Learning, 1986, 1(1): 81-106.
- [20] CHANG C C, LIN C J. LIBSVM: A library for support vector machines[J]. Acm Transactions on Intelligent Systems & Technology, 2011, 2(3): 27.

-
- [21] COVER T, HART P. Nearest neighbor pattern classification[J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.

四、科研成果

1.知识产权

1.1. 专利授权证书：一种基于深度学习的司机状态实时监测方法



证书号第6292385号

专利权人应当依照专利法及其实施细则规定缴纳年费。本专利的年费应当在每年07月10日前缴纳。未按照规定缴纳年费的，专利权自应当缴纳年费期满之日起终止。

申请日时本专利记载的申请人、发明人信息如下：

申请人：

华南农业大学

发明人：

王金凤;王文中;刘君扬;何振宇

1.2. 专利授权证书：基于多任务深度学习的情感识别方法、系统及存储介质



证书号第6233564号

专利权人应当依照专利法及其实施细则规定缴纳年费。本专利的年费应当在每年02月27日前缴纳。未按照规定缴纳年费的，专利权自应当缴纳年费期满之日起终止。

申请日时本专利记载的申请人、发明人信息如下：

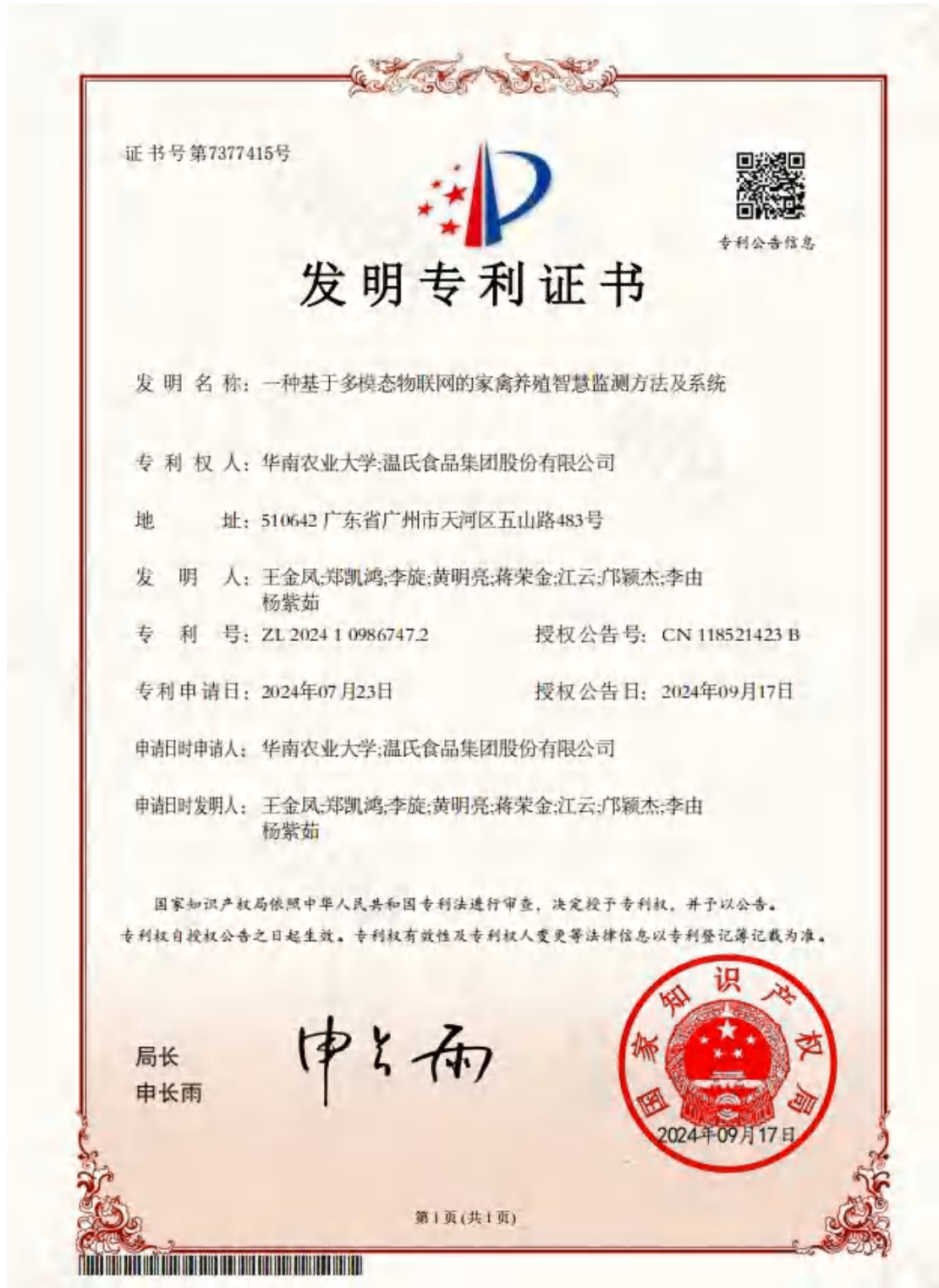
申请人：

华南农业大学

发明人：

王金凤;郑志桑;苏志坚;黄可;李杏圆;许健恒;尤茵茵;刘呈宇

1.3. 专利授权证书：一种基于多模态物联网的家禽养殖智慧监测方法及系统



1.4. 专利授权证书：一种基于大数据的企业外贸风险预警方法及系统



证书号第6102343号

专利权人应当依照专利法及其实施细则规定缴纳年费。本专利的年费应当在每年06月24日前缴纳。未按照规定缴纳年费的，专利权自应当缴纳年费期满之日起终止。

申请日时本专利记载的申请人、发明人信息如下：

申请人：

华南农业大学

发明人：

王文中,王金凤,陈嘉杰,陈润健,刘君扬,周展博,何振宇,郭达暨

1.5. 软件著作权：基于微信小程序的会议室预约系统 V1.0



1.6. 软件著作权：基于微信小程序的华农智能导航系统 V1.0



1.7. 软件著作权：基于深度学习的鸡群跟踪计数系统[简称：
鸡群计数系统]V1.0



1.8. 软件著作权：基于全局规划的禽类配送站点智能排班系统[简称：禽类配送站点智能排班系统]V1.0



1.9. 软件著作权：操作系统辅助教学系统 V1.0



1.10. 软件著作权：货物优化运输与排班系统 V1.0



1.11. 软件著作权：基于深度学习的人脸疲劳识别系统 V1.0



1.12.软件著作权：美丽华农导航系统[简称：华农导游系统]V1.0



1.13. 软件著作权：基于深度学习的鱼群密度检测系统 V1.0



1.14. 软件著作权：基于深度学习的司机安全带检测系统 V1.0



1.15. 软件著作权：基于微信小程序的教室管理系统[简称：
教室管理系统]V1.0



1.16. 软件著作权：基于区块链技术的校园公益平台系统[简称：校园公益平台]V1.0



1.17. 软件著作权：基于图卷积神经网络的学生课程个性化推荐系统[简称：触手课及]V1.0



1.18. 软件著作权：基于质点弹簧的血管手术模拟系统



1.19. 软件著作权：农村耕地评价及智能决策系统



1.20. 软件著作权：自然环境特效渲染插件系统



1.21. 软件著作权: SuperSocket——基于手机 APP 管理的智能插座配套 APP



1.22. 软件著作权：基于深度学习的蝴蝶种类识别系统[简称：
蝴蝶种类识别系统]V1.0



1.23. 软件著作权：基于神经网络的企业外贸汇率风险智能化决策系统[简称:企业外贸汇率风险智能化决策系统]V1.0



1.24. 软件著作权：汽车行业开放式创新 APP 系统 V1.0



1.25. 软件著作权：基于大数据的渔业外贸数据分析平台[简称：渔业宝]V1.0



1.26. 软件著作权：面向大数据的汽车行业创新资源交易平台[简称:汽车行业创新资源交易平台]V1.0



1.27. 软件著作权：外贸事务专家在线咨询平台[简称:外贸在线
咨询平台]V1.0



1.28. 软件著作权：企业外贸风险预警云平台[简称:外贸风险预警云平台]V1.0



1.29. 软件著作权：汽车行业创新资源交易大数据管理系统
[简称:汽车资源交易管理系统]V1.0



1.30. 软件著作权：基于路径优化的旅游推荐系统[简称：旅游推荐系统] V1.0



1.31. 软件著作权：基于物联网数据分析的智慧养殖监测系统[简称：智慧养殖监测系统]V1.0



五、其他业绩

1.指导学生学科竞赛

1.1. 2024年(第7届)“泰迪杯”数据分析技能赛本科及以上组 一等奖



1.2. 2024年(第7届)“泰迪杯”数据分析技能赛本科及以上组 二等奖



1.3. 第十四届蓝桥杯全国软件和信息技术专业人才大赛全国总决赛 C/C++程序设计大学 B 组二等奖



1.4. 第十四届蓝桥杯全国软件和信息技术专业人才大赛全国总决赛 C/C++程序设计大学 B 组二等奖



1.5. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区 C/C++程序设计大学 A 组二等奖



1.6. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区 C/C++程序设计大学 A 组二等奖



1.7. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区 C/C++程序设计大学 B 组一等奖



1.8. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区 C/C++程序设计大学 B 组一等奖



1.9. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区 C/C++程序设计大学 B 组三等奖



1.10. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区 C/C++程序设计大学 B 组三等奖



1.11. 第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区 C/C++程序设计大学 B 组二等奖

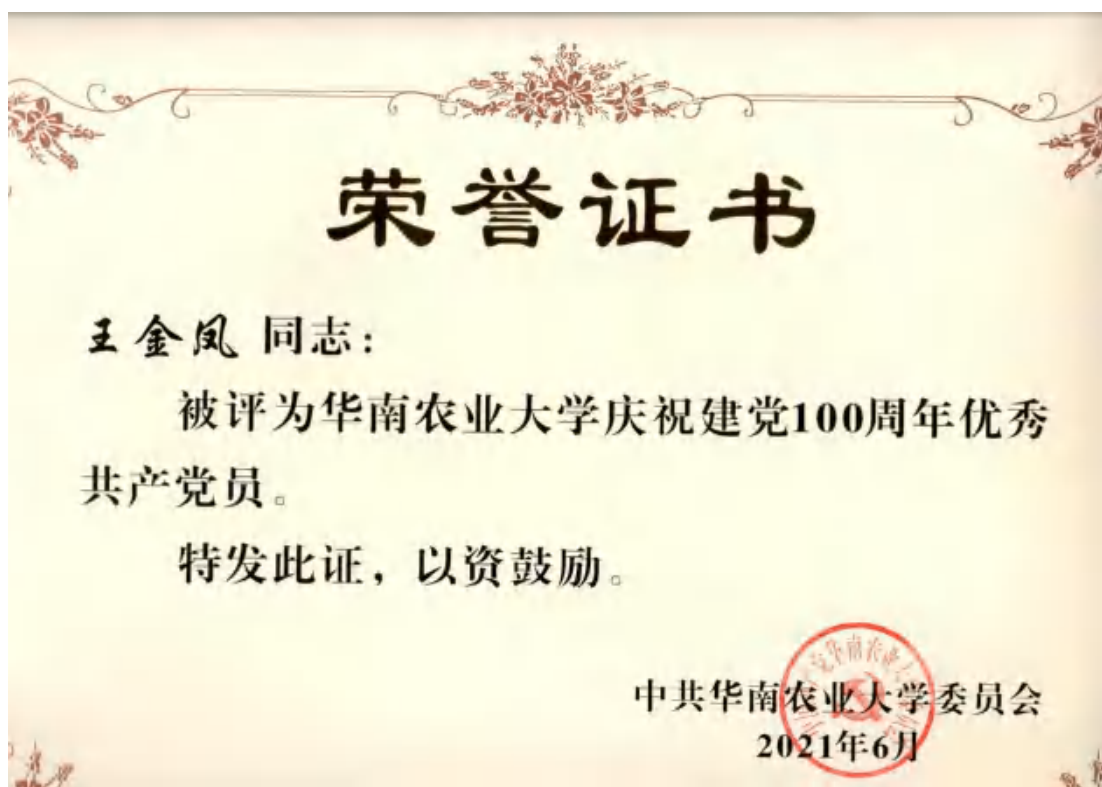


2.个人荣誉

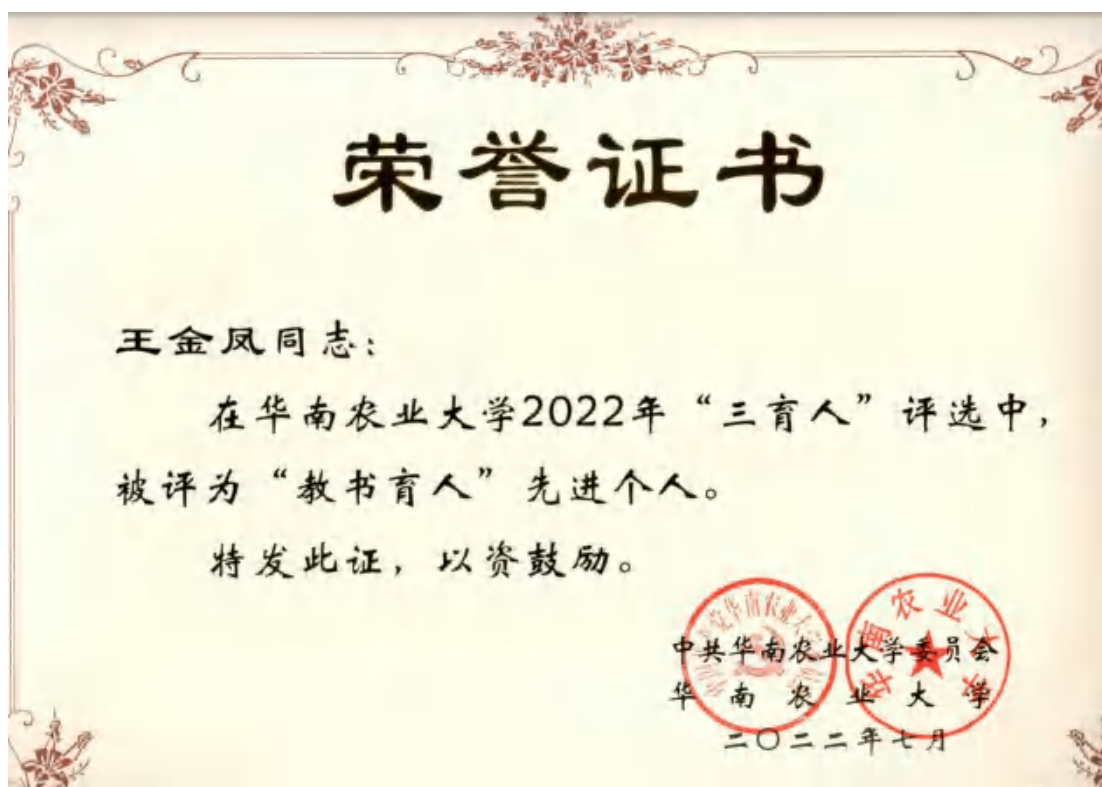
2.1. 校级“教学名师”证书



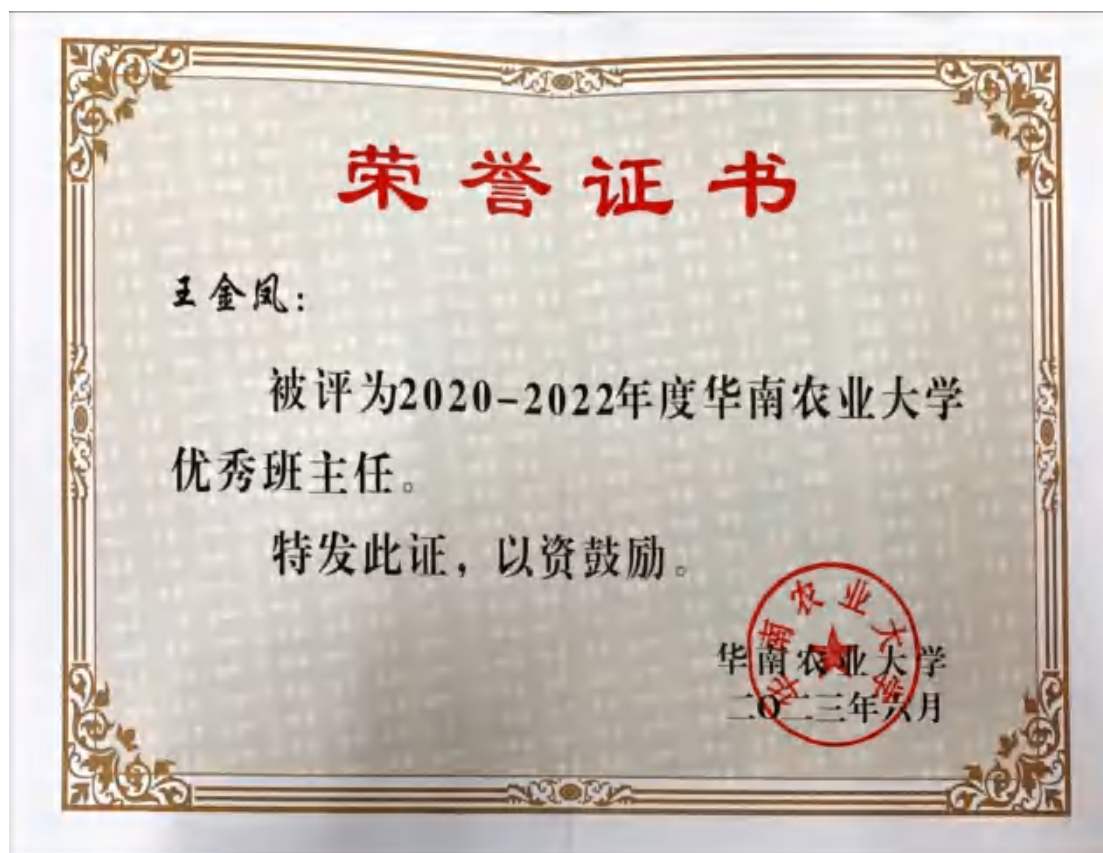
2.2. 校级“庆祝建党 100 周年优秀共产党员”证书



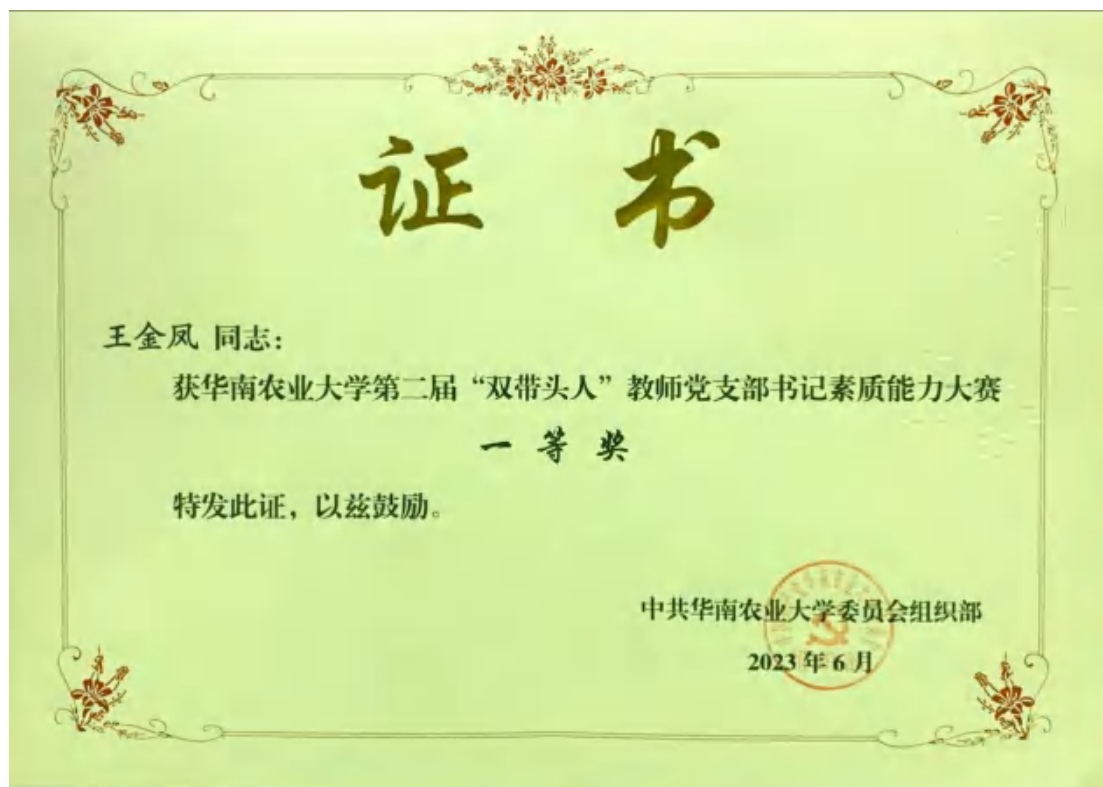
2.3. 校级“教书育人先进个人”证书



2.4. 校级“优秀班主任”证书



2.5. 第二届“双带头人”教师党支部书记素质能力大赛一等奖证书



2.6. 院级“十佳工作者”证书



3.其他业绩

3.1 全国党建工作样板支部

教育部办公厅

教思政厅函〔2024〕7号

教育部办公厅关于公布第四批全国党建工作 示范高校、标杆院系、样板支部 培育创建单位名单的通知

各省、自治区、直辖市党委教育工作部门，新疆生产建设兵团教育局，部属各高等学校党委、部省合建各高等学校党委：

根据《教育部办公厅关于开展第四批新时代高校党建示范创建和质量创优工作的通知》（教思政厅函〔2023〕17号）安排和遴选工作方案，经省级党委教育工作部门推荐、资格审查、专家推荐、教育部党的建设和全面从严治党工作领导小组成员单位集中审议、结果公示，遴选产生了10个全国党建工作示范高校、100个全国党建工作标杆院系、1001个全国党建工作样板支部培育创建单位，现将名单予以公布（见附件1、2、3）。各培育创建单位建设周期为2年，自通知发布之日起至2026年3月。有关工作安排和要求如下。

一、认真培育建设

各培育创建单位要按照教思政厅函〔2023〕17号文件所附

的《第四批新时代高校党建“双创”工作重点任务指南》要求，坚持软件建设和硬件建设相结合、统筹规划和分步实施相结合、整体提升和品牌塑造相结合，按计划、分步骤开展培育创建工作。教育部在全国高校思想政治工作网（<http://www.sizhengwang.cn>）上（以下简称全国高校思政网），为每个培育创建单位搭建网上工作平台（平台登录账号另行通知），各单位要及时在平台上发布工作进展、经验成效。

二、加强动态管理

坚持目标管理和过程管理相结合，委托高校思想政治工作创新发展中心（华中师范大学）对培育创建单位开展跟踪评价。各培育创建单位应于2025年3月前提交中期建设进展情况报告，建设期满提交工作总结报告、成果汇编和工作案例。教育部组织开展验收，公布验收评定结果，验收通过的创建单位建设成果将面向全国高校推广应用。培育创建单位所在属地党委教育部门和高校党委，要加强工作指导，提供条件保障，加强示范引领，确保建设实效，引领带动高校党建工作质量整体提升。

三、联系方式

1. 教育部思想政治工作司：010-66096675。
2. 高校思想政治工作创新发展中心（华中师范大学）：027-67862663。
3. 全国高校思政网：010-56803673；010-56803621。

- 附件：1. 第四批“全国党建工作示范高校”培育创建单位
名单
2. 第四批“全国党建工作标杆院系”培育创建单位
名单
3. 第四批“全国党建工作样板支部”培育创建单位
名单

教育部办公厅

2024年4月16日

(此件主动公开)

部内发送：有关部领导，办公厅

教育部办公厅

2024年4月18日印发

附件 3

第四批“全国党建工作样板支部”培育 创建单位名单

(排名不分先后)

序号	单 位
1	北京大学心理与认知科学学院本科生党支部
2	北京大学前沿交叉学科研究院教工第一党支部
3	北京大学口腔医院修复科党支部
4	清华大学车辆学院车辆博 21 党支部
5	清华大学自动化系系统工程所党支部
6	清华大学环境学院大气所第一党支部
7	中国人民大学理学院化学系硕士生联合党支部
8	中国人民大学经济学院政治经济学、经济史和经济思想史 教研室联合党支部
9	北京师范大学生命科学院乐育党支部
10	北京师范大学教育学部国际与比较教育研究院党支部
11	中国农业大学信息与电气工程学院数据科学研究生党支部
12	中国农业大学工学院农业机械化工程教职工党支部
13	北京航空航天大学能源与动力工程学院航空推进系教师党支部
14	北京理工大学光电学院博士光工 3 班党支部
15	北京理工大学集成电路与电子学院微电子技术研究所党支部
16	北京理工大学求是书院第一党支部
17	北京语言大学语言科学院教师党支部

序号	单位
713	暨南大学化学与材料学院直属超分子配位化学研究所党支部
714	华南农业大学数学与信息学院、软件学院计算机应用教工党支部
715	广东海洋大学电子与信息工程学院电子工程系教工党支部
716	广州医科大学药学院教工第三党支部
717	广州医科大学附属第一医院呼研院第一党支部
718	广州中医药大学针灸康复临床医学院学生第四党支部
719	广州中医药大学第一临床医学院急诊中心党总支第一党支部
720	广东药科大学公共卫生学院王烁青年志愿者服务队党支部
721	广东药科大学基础医学院教工第六党支部
722	华南师范大学教育科学学院长江学者创新团队党支部
723	惠州学院电子信息与电气工程学院学生第一党支部
724	广州美术学院雕塑与公共艺术学院教工党支部
725	广东技术师范大学管理学院学生第三党支部
726	广东技术师范大学美术学院教工第二党支部
727	广东技术师范大学美术学院艺心党支部
728	深圳大学经济学院本科金融学党支部
729	深圳大学化学与环境工程学院第三教职工党支部
730	顺德职业技术学院酒店与旅游管理学院旅游会展党支部
731	广东轻工职业技术学院轻化工技术学院教师第三党支部
732	广东交通职业技术学院轨道交通学院教师一支部
733	广州大学经济与统计学院本科生第一党支部
734	广州大学化学化工学院分析科学技术研究中心党支部
735	广州大学土木工程学院工程抗震研究中心党支部

3.2 教育部内地与港澳大中小学师生交流计划大学生项目

教育部司局函件

教港澳台办〔2022〕7号

关于2022年内地与港澳大中小学师生交流计划大学生项目的批复

有关高校:

你校报送的2022年内地与港澳大中小学师生交流计划大学生项目(以下简称“万人计划”)申请书收悉。

根据中央财政预算有关规定及教育部对推动内地与港澳教育交流合作提质增效的要求,现将批复意见通知你校(详见附件),具体工作要求如下。

一、严格履行项目程序。请你校精心组织实施,注重质量和效果,确保项目顺利完成,于项目完成后15个工作日内提交项目执行情况、项目总结、师生信息表及费用结算表等材料,并于12月31日前提交学校年度总结与决算报告。项目如需调整,请于项目申报实施日期前15个工作日内提交项目调整报告。

二、严格使用项目经费。根据专家评审结果,2021年你校项目执行情况并扣除2021年项目结余经费后,向你校下拨资助经费。教育部直属高校经费已纳入2022年中央财政预算,由财政部统一下拨;非教育部直属高校经费计划于2022年4月下拨。

请你校严格按照有关规定使用经费，在经费下拨前执行的项目，请先行垫付有关费用。如有多个项目获批，可在万人计划项目经费内由你校统筹使用。

三、严格开展项目监管。请你校严格把关，提高项目执行质量及经费使用率，并配合我办做好委托有关机构开展的项目实施效果调研工作。

此复。

联系人：毛亚芳 景平康 010-66096937/6281

附件：2022年内地与港澳大中小学师生交流计划大学生项目批复意见

教育部港澳台事务办公室

2022年2月25日



附件

2022年内地与港澳大中小学师生交流计划大学生项目批复意见

高校名称	序号	项目名称	活动类别	时间类型	审批人数	批复金额 (万元)	项目批复 总金额 (万元)	使用2021年 结余金额 (万元)	2022年实际 拨付总额 (万元)
华南农业大学	1	慧治城乡，情满湾区	实践交流类	短期	10	1.38	12.73	0.00	12.73
	2	“AI+”助力智慧农业—学术研讨班	教学科研类	短期	11	2.12			
	3	广东经典民谣（民歌）传唱研修体验营	实践交流类	短期	20	0.00			
	4	昆虫世界奇妙物语-粤港澳生南岭国家自然保护区调查	实践交流类	短期	15	2.88			
	5	文化传承视域下的影视动画创作研究(线上)	实践交流类	短期	20	0.00			
	6	心手相连、共筑粤港澳学子手球情	实践交流类	短期	32	0.00			
	7	农业机械技术运用对乡村振兴促进研究与实践	实践交流类	短期	11	6.35			
	8	粤港澳茶文化交流及制茶体验活动	实践交流类	短期	15	0.00			

附件 4

2022 年内地与港澳大中小学师生交流计划
大学生项目确认书

为落实教育部内地与港澳大中小学师生交流计划，经友好协商，华南农业大学与澳门科技大学将合作申请下列项目，列入“2022 年内地与港澳大中小学师生交流计划大学生项目”。并由华南农业大学负责向教育部申报。

序号	项目名称	项目实施日期 (港澳师生人数及天数)	港澳负责单位 及联系人	内地负责单位 及联系人
1	“AI+”助力智慧农业—学术研讨班	2022 年 6 月 10 日至 2022 年 6 月 16 日， 7 天 师生人数：11 人	澳门科技大学 姓名：梁勇 职务：无 电话：18063869506 邮箱： yliang@ust.edu.mo	华南农业大学 姓名：王金凤 职务：无 电话：15815860686 邮箱： wangjinfeng@scau.edu.cn

双方已就上述合作事项达成共识，并将合力推动相关项目的落实。该确认书由双方授权代表签署，自签署之日起生效。



姓名：陈延夫
职务：港澳台办公室主任
日期：2021.11.11

澳门科技大学
(授权代表) 签署



姓名：伍乃骥
职务：澳门科技大学系统工程研究所主任
日期：11/16/2021

3.3 全省党建工作样板支部

中共广东省委教育工作委员会

粤教工委组函〔2023〕3号

中共广东省委教育工委关于公布第四批全省 党建工作示范高校、标杆院系、样板支部 培育创建单位名单的通知

各高校党委：

根据《中共广东省委教育工委关于开展第四批新时代高校党建示范创建和质量创优工作的通知》安排和遴选工作方案，经各高校申报、资格审查、专家评审、省委教育工委会议审议、结果公示，遴选产生了第四批全省党建双创工作培育创建单位“党建工作示范高校”5所、“党建工作标杆院系”51个、“党建工作样板支部”292个。现将名单予以公布（见附件1、2、3）。各培育创建单位建设周期为2年，自通知发布之日起至2025年3月。有关工作安排和要求如下。

一、加强培育，引领带动质量提升。各高校党委要高度重视，加强组织领导，按计划、分步骤组织实施立项的党建工作示范高校、标杆院系、样板支部开展创建工作，真正做到党建示范校“六

个过硬”、标杆院系“五个到位”、样板支部“七个有力”，真正成为我省高校党建示范标兵。要及时总结发掘、宣传推广入选党组织的探索经验、培育成果、创建成效，充分发挥引领示范、辐射带动作用，有计划有步骤地把点上的经验做法推广到面上去，进一步引领带动高校党建工作质量整体提升，为加快一流大学和学科建设、实现高等教育内涵式发展、办好人民满意的教育提供坚强保证。

二、加强管理，推进取得建设成效。坚持目标管理和过程管理相结合，加强对创建单位的常态化跟踪和日常工作指导，及时解决创建中遇到的困难和问题，及时总结、推广建设经验和成果，推进取得建设成效。省委教育工委将开展创建培育期中评估和期满考核验收，未通过期中评估的培育创建单位，将视情况停止拨付下一年度建设经费并督促整改落实；期满验收不达标的单位，将予以通报，并严格追责问责。

三、加强支持，确保培育创建有保障。省委教育工委将按照各类项目建设标准，在建设周期内共给予全省党建工作示范高校10万/校、标杆院系2万/院（系）、样板支部1万/支部的经费支持，分两年建设周期下拨。各高校要结合实际，为入选的培育创建单位提供创建配套经费、资源条件等支持；原则上高校支持经费不少于省委教育工委的经费标准。要加强对资金的管理使用，专款专用，按相关规定负责履行监管经费使用管理，确保发挥成效。

四、联系方式

联系人：徐鹏飞，联系电话：020-37627553，电子邮箱：
zzc@gdedu.gov.cn

- 附件：1.第四批“全省党建工作示范高校”培育创建单位
名单
2.第四批“全省党建工作标杆院系”培育创建单位
名单
3.第四批“全省党建工作样板支部”培育创建单位
名单



附件 3

第四批“全省党建工作样板支部”

培育创建单位名单

(排名不分先后)

序号	名单
1	中山大学中山眼科中心临床师生第一党支部
2	中山大学智能工程学院智能交通专任教师党支部
3	中山大学孙逸仙纪念医院药学部教职工第一党支部
4	中山大学生命科学学院生态系统演化与修复党支部
5	中山大学马克思主义学院马克思主义基本原理教研室 专任教师党支部
6	中山大学海洋科学学院物理海洋专任教师党支部
7	中山大学公共卫生学院医学统计学系专任教师党支部
8	中山大学附属第一医院重症医学科教职工第一党支部
9	中山大学附属第五医院肿瘤中心教职工党支部
10	中山大学附属第三医院手术麻醉中心 教职工第一党支部
11	中山大学哲学系专任教师第二党支部
12	中山大学心理学系教工党支部

— 8 —

49	华南农业大学数学与信息学院、软件学院 计算机应用教工党支部
50	华南农业大学工程学院农业工程系教工党支部
51	南方医科大学珠江医院（第二临床医学院） 康复医学科党支部
52	南方医科大学南方医院（第一临床医学院） 检验医学科党支部
53	南方医科大学口腔医院海珠广场院区第一党支部
54	南方医科大学基础医学院人体解剖学教研室党支部
55	南方医科大学第三附属医院（第三临床医学院） 骨科第二党支部
56	广州中医药大学第一临床医学院（第一附属医院） 直属内科第二党支部
57	广州中医药大学第一临床医学院（第一附属医院） 直属急诊中心第一党支部
58	广州中医药大学第一临床医学院（第一附属医院） 大学城学生党支部
59	广州中医药大学第二临床医学院（第二附属医院） 大德路总院药剂第一党支部
60	广州中医药大学第二临床医学院（第二附属医院） 大德路总院机关第一党支部
61	广州中医药大学医学信息工程学院学生第一党支部
62	广州中医药大学护理学院研究生党支部
63	广州中医药大学公共卫生与管理学院 公共事业管理专业学生党支部
64	广州中医药大学第三临床医学院（第三附属医院） 关节中心第一党支部

3.4 华南农业大学第二批“党建工作样板支部”

中共华南农业大学委员会文件

华农党发〔2022〕61号



中共华南农业大学委员会 关于公布第二批“党建工作样板支部” 培育创建单位的通知

各二级党组织，各学院、部处，各单位：

根据学校《关于开展第二批“党建工作样板支部”培育创建单位申报工作的通知》要求，学校党委组织开展了第二批校级“党建工作样板支部”培育创建单位评选工作。经各二级党组织遴选推荐和学校党委审核，确定农学院党委种子科学技术系教工党支部等32个党支部为我校第二批“党建工作样板支部”培育创建单位（具体名单见附件），培育创建时间为2022年10月至2024年

— 1 —

10月。

入选培育创建单位要按照《中共教育部党组关于高校党组织“对标争先”建设计划的实施意见》和学校《关于开展第二批“党建工作样板支部”培育创建单位申报工作的通知》要求，坚持整体提升和品牌塑造相结合、党建工作与事业发展相融合，明确建设周期的总体思路、建设目标、年度计划、工作举措等，严格“三会一课”、创新工作方法、发挥典型示范，真正做到“七个有力”（教育党员有力、管理党员有力、监督党员有力、组织师生有力、宣传师生有力、凝聚师生有力、服务师生有力），真正成为学校党支部建设的示范标兵。

各二级党组织要对入选培育创建单位加强常态化跟踪和日常工作指导，及时解决建设过程中遇到的困难和问题，提升创建成效；要及时总结凝练、宣传推广入选党支部的建设经验和特色做法，充分发挥典型示范和辐射带动作用，有计划有步骤地把点上的经验做法推广到面上去，引领带动党建工作质量整体提升；要结合实际提供配套经费、资源场地、人力物力等支持；要建立激励保障机制，对建设成效显著的党支部、党员骨干给予表彰奖励。

学校党委将开展培育创建中期评估和期满考核验收，验收达标的，予以结项，深化宣传推广；验收未通过的单位予以撤项；造成严重不良影响的，给予追责问责。培育创建期内，每年对每个支部提供2000元建设经费支持。

附件：华南农业大学第二批“党建工作样板支部”培育创建单位名单

中共华南农业大学委员会

2022年10月25日

— 3 —

附件

华南农业大学第二批“党建工作样板支部” 培育创建单位名单

农学院党委种子科学技术系教工党支部
植物保护学院党委农药学系硕士生党支部
林学与风景园林学院党委林木遗传育种教研室教工党支部
园艺学院党委本科生园艺丁颖班党支部
兽医学院党委临床兽医学博士研究生党支部
动物科学学院党委智慧养殖系教工党支部
资源环境学院党委耕地保育创新团队党支部
海洋学院党委海洋资源与环境系教工党支部
生命科学学院党委植物学教工党支部
工程学院党委农业工程系教工党支部
食品学院党委食品质量与安全团队教工党支部
水利与土木工程学院党委研究生党支部
材料与能源学院党委研究生木材科学与工程党支部
数学与信息学院、软件学院党委计算机应用教工党支部
电子工程学院（人工智能学院）党委研究生第一党支部
经济管理学院党委农林经济系教工党支部
公共管理学院党委社会工作系教工党支部

— 4 —

3.5 全省高校“双带头人”教师党支部书记“广东行”专项行动 建设项目

中共广东省委教育工作委员会

粤教工委组函〔2024〕12号

中共广东省委教育工委关于公布全国高校“ 双带头人”教师党支部书记“强国行”专项 行动推荐名单和全省高校“双带头人” 教师党支部书记“广东行”专项行动 建设名单的通知

各高校党委：

根据《教育部办公厅关于开展高校“双带头人”教师党支部书记“强国行”专项行动的通知》有关精神，以及《中共广东省委教育工委转发教育部办公厅关于开展高校“双带头人”教师党支部书记“强国行”专项行动的通知》要求，省委教育工委组织开展了全国高校“双带头人”教师党支部书记“强国行”专项行动（以下简称“强国行”专项行动）和全省高校“双带头人”教师党支部书记“广东行”专项行动（以下简称“广东行”专项行动）建设单位的遴选工作。经高校自主申报、专家评审和省委教育工委审定，确定中山大学附属第五医院感染病防治中心教职工党支部等 98 个单

位为“强国行”专项行动推荐单位（其中，20个第一、二批全国高校“双带头人”工作室以及第三批全国高校“双带头人”工作室建设对象，不占推荐名额），确定华南理工大学造纸与污染控制国家工程中心党支部等154个单位为“广东行”专项行动建设单位。现将名单予以公布（见附件），并提出如下工作要求。

一、加强管理，强化工作保障。各高校党委要高度重视“强国行”“广东行”专项行动，将其作为提升教师党建和思想政治工作质量的重要抓手，作为提升育人实效的有效载体，不断强化各级“双带头人”工作室的培育创建，积极动员“双带头人”参加到专项行动中，加强资源统筹，加大经费、政策等配套支持力度，增强在述职考核、评优评先等环节的导向力度，确保专项行动取得实效。

二、压实责任，深化行动成效。专项行动周期为3年，行动周期内将对各“双带头人”开展专项行动情况进行总结，形成一批高校“双带头人”教师党支部书记专项行动典型案例。将基层党组织动员“双带头人”参与专项行动情况，作为党建“双创”工作成效的重要参考；将专项工作成效作为全国、全省高校“双带头人”工作室建设验收的重要内容，作为推荐全国模范教师、全国优秀教师、全国教书育人楷模、最美教师、全国高校黄大年式教师团队等的参考因素。各“双带头人”每年定期要在高校思政网提交专项行动开展情况。校级“双带头人”工作室参加“专项工作”成绩突出的，在申报全省高校“双带头人”工作室同等条件下优先

考虑。

三、加强经验总结，注重宣传推广。要围绕中心，服务大局，在党建联建、教育服务、科技赋能和实践育人等方面，强化学校、院（系）各级党组织的作用发挥，及时总结专项行动开展情况和实际成效，梳理好经验好做法，推动“强国行”“广东行”专项行动有序有力开展，不断提升工作实效。加大先进典型发掘力度，积极宣传报道“专项行动”中的先进经验和典型事迹，着力形成品牌效应，不断扩大社会影响，营造良好氛围。

四、联系方式

联系人：徐鹏飞、龚炜华

联系电话：020-37626696

- 附件：1.第一、二批全国高校“双带头人”工作室以及第三批全国高校“双带头人”工作室建设对象申报“强国行”专项行动名单（20个）
- 2.全国高校“双带头人”教师党支部书记“强国行”专项行动推荐名单（78个）
- 3.全省高校“双带头人”教师党支部书记“广东行”专项行动建设名单（154个）



附件 3

全省高校“双带头人”教师党支部书记“广东行” 专项行动建设名单

(154 个)

序号	党支部名称
1	中山大学大气科学学院海洋科学考察中心党支部
2	中山大学测绘科学与技术学院教职工第二党支部
3	中山大学岭南学院金融学专任教师第二党支部
4	中山大学附属第一医院心血管教职工第二党支部
5	中山大学孙逸仙纪念医院神经科教职工党支部
6	中山大学公共卫生学院(深圳)教职工第一党支部
7	中山大学海洋科学学院物理海洋专任教师党支部
8	中山大学孙逸仙纪念医院骨外科教职工党支部
9	中山大学航空航天学院力学教职工党支部
10	华南理工大学造纸与污染控制国家工程中心党支部
11	华南理工大学电子与信息学院移动超声探测党支部
12	华南理工大学吴贤铭智能工程学院教工党支部
13	暨南大学公共管理学院/应急管理學院行政管理系教工党支部
14	暨南大学基础医学与公共卫生学院粤港澳中枢神经再生研究院第一党支部
15	暨南大学环境与气候学院教工第四党支部
16	暨南大学药学院生物活性分子与成药性优化全国重点实验室党支部

17	华南农业大学动物科学学院蚕丝科学系教工党支部
18	华南农业大学数学与信息学院、软件学院计算机应用教工党支部
19	南方医科大学珠江医院（第二临床医学院）妇产医学中心党支部
20	南方医科大学第三附属医院（第三临床医学院）骨科第二党支部
21	广州中医药大学第一附属医院（第一临床医学院）肿瘤中心第一党支部
22	广州中医药大学第二附属医院（第二临床医学院、广东省中医院）大德路总院骨伤科第一党支部
23	华南师范大学心理学院“黄大年式教师团队”创新党支部
24	华南师范大学数学科学学院青年教师先锋党支部
25	广东工业大学艺术与设计的学院工业设计系教工党支部
26	广东工业大学机电工程学院机械电子工程系第一党支部（国家重点实验室团队教工党支部）
27	广东工业大学轻工化工学院应用化学系党支部
28	广东外语外贸大学西方语言文化学院教工第一党支部
29	汕头大学理学院生物海洋教工党支部
30	广东财经大学人文与传播学院新闻学系教工党支部
31	广东财经大学旅游管理与规划设计研究院、岭南旅游研究院（合署）党支部
32	广东医科大学公共卫生学院流行病与卫生统计学系党支部
33	广东海洋大学水产学院海洋渔业系教工党支部
34	广东海洋大学滨海农业学院生物技术系教工党支部
35	仲恺农业工程学院动物科技学院水产教工党支部