

应用统计 与数据科学

Applied Statistics and Data Science



ART AND TECHNOLOGY PRESS INC.

(517 666 0904)

263 S KENWOOD ST 560

CASPER, WY 82601

Copyright © 2026 by ART AND TECHNOLOGY PRESS INC. (United States)

Complimentary Copy



ART AND TECHNOLOGY PRESS INC.
(United States)

编委会名单

主 编



王国长 (Guochang Wang), 暨南大学 (Jinan University),
邮箱: wanggc023@amss.ac.cn
研究兴趣: 函数型数据分析、时间序列、充分性降维、统计与机器学习
functional data analysis, time series, sufficiency dimension reduction, statistics and machine learning

执行副主编



张兴发 (Xingfa Zhang), 广州大学 (Guangzhou University),
邮箱: xingfazhang@gzhu.edu.cn
研究兴趣: 金融统计、环境统计、机器学习
financial statistics, environmental statistics, machine learning

编 委



胡志勇 (Zhiyong Hu), 广州大学 (Guangzhou University)
邮箱: zhyhu65@163.com
研究兴趣: 大数据, 人工智能, 财务与会计
big data, artificial intelligence, finance and accounting



雷田礼 (Tianli Lei), 深圳职业技术大学 (Shen Zhen Polytechnic University)
邮箱: ltl@szpu.edu.cn
研究兴趣: 数量经济、高职数学教育
auantitative economics, higher vocational mathematics education



舒连杰 (Lianjie Shu), 澳门大学 (University of Macau),
邮箱: ljshu@um.edu.mo
研究兴趣: 量化金融, 统计学习, 质量控制及管理
quantitative finance, statistical learning, quality control and management



王纬 (Wei Wang), 山东财经大学 (Shandong University of Finance and Economics)
邮箱: wangwei_0115@sdufe.edu.cn
研究兴趣: 统计机器学习、应用统计、计量经济
statistical machine learning, applied statistics, econometrics



杨凯 (Kai Yang), 长春工业大学 (Changchun University of Technology)
邮箱: yangkai@ccut.edu.cn
研究兴趣: 时间序列分析、高维数据分析、贝叶斯分析
time series analysis, high-dimensional data analysis, bayesian analysis



周彦 (Yan Zhou), 深圳大学 (Shenzhen University)
邮箱: zhouy1016@163.com
研究兴趣: 生物统计, 数据科学
Biostatistics, data science



李永明 (Yongming Li), 上饶师范学院 (Shangrao Normal University)
邮箱: lym1019@163.com
研究兴趣: 非参数统计, 金融统计, 极限理论与统计大样本性质, 小波方法
nonparametric statistics, financial statistics, limit theory and statistical large sample theory, wavelet method



刘照德 (Zhao de Liu), 广东财经大学 (Guangdong University of Finance & Economics)
邮箱: lzhaode@163.com
研究兴趣: 经济统计、多元统计、计量分析
economic statistics, multivariate statistics, econometric analysis



王小刚 (Xiaogang Wang), 北方民族大学 (North Minzu University)
邮箱: wongone@163.com
研究兴趣: 经济与社会统计
economic and social statistics



严晓东 (Xiaodong Yan), 西安交通大学 (Xi'an Jiaotong University)
邮箱: yanxiaodong@xjtu.edu.cn
研究兴趣: 统计机器学习、智能体统计学
statistical machine learning, agent statistics



朱柯 (Ke Zhu), 香港大学 (The University of Hong Kong)
邮箱: mazhuke@hku.hk
研究兴趣: 统计建模、金融时间序列分析、计量经济、金融大数据
statistical modeling, financial time series analysis, econometrics, financial big data

应用统计与数据科学

Applied Statistics and Data Science

第2卷 第6期 2026年6月刊

主办单位: ART AND TECHNOLOGY PRESS INC.

(United States)

承办单位: 广州统计师事务所

指导单位: 广东省现场统计学会、广州市统计学会

主 编: 王国长

执行副主编: 张兴发

编 委: 胡志勇、李永明、雷田礼、刘照德、
舒连杰、王小刚、王 纬、严晓东、
杨 凯、朱 柯、周 彦

地 址: 263 S KENWOOD ST 560

CASPER,WY 82601

网 址: <https://arttechpress.com>

本刊说明:

凡向本刊所投稿件, 全体作者需签署论文著作权转让声明书和论文发表承诺书, 声明、承诺及相关事项如下:

- 作者将论文的复制权、发行权、网络传播权、翻译权、汇编权、信息网络传播权、改编权等著作权在世界范围内免费转让给本刊。
- 论文不侵犯他人著作权和其他权利, 否则作者将承担由此产生的全部责任, 并赔偿由此给出版单位造成的全部损失。
- 论文署名作者享有该作品的完全著作权, 署名作者的身份真实。
- 论文未曾以任何形式公开发表过。
- 作者所投本刊稿件, 本刊编辑部拥有修改权。



管理统计 | MANAGEMENT STATISTICS

- 001 探讨大数据技术对企业财务管理的影响 王慧娟
Discussion on the Influence of Big Data Technology on Enterprise Financial Management Wang Huijuan
- 004 第三方支付平台对跨境电商资金安全的影响探析 姚春光
Analysis on the Influence of Third-Party Payment Platform on the Safety of Cross-Border E-commerce Funds Yao Chunguang
- 007 集团化企业合并报表编制难点与内部交易抵消实务 郑婉文
Consolidated Financial Statements Preparation Challenges and Intercompany Transaction Elimination Practices in Group Enterprises Zheng Wanwen
- 012 农牧行业集团内控有效性评价体系构建与实施路径研究——以 A 集团下属子公司自查实践为例 丁兰
Research on the Construction and Implementation Path of the Internal Control Effectiveness Evaluation System in Agricultural and Animal Husbandry Industry Groups—A Case Study of Self-Inspection Practices of Subsidiaries of Group A Ding Lan

统计与教育教学 | STATISTICS AND EDUCATION TEACHING

- 016 新课标下构建数学师范生教学技能训练培养体系与实践 盛婷婷
Construction and Practice of Teaching Skill Training System for Mathematics Normal Students Under the New Curriculum Standard Sheng Tingting
- 020 教育数学思想下中学极限概念的教学衔接研究 胡志安, 徐洪焱, 李弘
Grounded in Educational Mathematical Thinking: Research on the Cognitive Progression and Teaching Integration of the Concept of Limits in Middle School Hu Zhi'an, Xu Hongyan, Li Hong
- 024 人工智能与线性代数课程深度融合的教学改革研究——基于实证数据的成效验证 吴嘉润, 赵娜娜, 苑孟理想, 唐子航, 刘新红
Research on Teaching Reform of Deep Integration of Artificial Intelligence and Linear Algebra Curriculum—Effectiveness Verification Based on Empirical Data Wu Jiarun, Zhao Nana, Yuan Menglixiang, Tang Zihang, Liu Xinhong

数据科学技术、方法与应用 | DATA SCIENCE TECHNOLOGY, METHODS AND APPLICATIONS

- 028 基于组合赋权云模型的自动驾驶安全综合评价研究 詹媛媛, 周金明
Combined Weighting and Cloud Model Based Comprehensive Safety Evaluation of Autonomous Driving Zhan Yuanyuan, Zhou Jinming
- 036 双模型校验驱动的 C 语言智能出题系统研究 江玉珍, 朱映辉
Research on Intelligent C Language Question Generation System Driven by Dual-LLM Verification Jiang Yuzhen, Zhu Yinghui
- 040 融合 BERT 语义表征与残差特征增强的 O2O 食品安全风险监测研究 金思彤, 李建波
Research on O2O Food Safety Risk Monitoring Based on BERT Semantic Representation and Residual Feature Enhancement Jin Sitong, Li Jianbo

049	基于代价敏感 XGBoost 模型的企业 ESG 评级预测 Enterprise ESG Rating Prediction Based on Cost-Sensitive XGBoost Model	鲍金敏 Bao Jinmin
055	基于神经网络和多目标粒子群优化的智能路径规划方法及其应用 Intelligent Path Planning Method Based on Graph Neural Network and Multi-Objective Particle Swarm Optimization and its Application	吴荣康, 吴伟佳, 黄惠婷 Wu Rongkang, Wu Weijia, Huang Huiting

统计方法及应用 | STATISTICAL METHODS AND APPLICATIONS

061	广东省绿色低碳创新效率的统计测度与时空演化分析 Statistical Measurement and Spatiotemporal Evolution Mechanisms of Green and Low-Carbon Innovation Efficiency in Guangdong Province	张倩生, 朱曹了了 Zhang Qiansheng, Zhu Caoliaoliao
069	公司创业投资对创业企业 ESG 表现的影响——基于深圳创业板的实证分析 The Impact of Corporate Venture Capital on the ESG Performance of Startups — An Empirical Analysis Based on Shenzhen Growth Enterprise Market	乔明哲, 余俊, 肖文慧 Qiao Mingzhe, Yu Jun, Xiao Wenhui
076	科技金融助推中国西部地区全要素生产率提升的作用路径及对策研究 The Action Path and Countermeasure Research of Sci-Tech Finance Boosting Total Factor Productivity Growth in Western China	石光周, 钱成 Shi Guangzhou, Qian Cheng
083	广州市智慧场馆平台的使用现状及消费意愿调查 A Survey on the Usage Status and Consumption Intention of Guangzhou's Smart Venue Platform	王婷, 李迪, 陈颖熙, 刘诺柠 Wang Ting, Li Di, Chen Yingxi, Liu Nuoning

探讨大数据技术对企业财务管理的影响

王慧娟

广州金不换咨询有限公司，广东 广州 510080

DOI:10.61369/ASDS.2026060001

摘 要： 在数字经济深度发展背景下，企业面临的经营环境复杂。传统财务管理模式因数据处理能力欠缺、决策依据单一等问题，较难适配企业发展新需求。大数据技术凭借规模庞大、类型多元、处理高效等特性，成为企业财务管理转型升级重要支撑。此文从大数据技术核心内涵与特征出发，探讨大数据技术对企业财务决策、运营成本、管理职能等方面的具体影响。提出开展信息化平台部署、构建多层次财务指标监测体系、加强数据全生命周期管理机制等策略。大数据技术不但可推动企业财务工作从人工操作到智能化操作、从事后核算到事前规划事中控制转变，更能重新构建财务管理范式，推动财务部门从成本中心向价值创造中心转型，为企业财务管理数字化变革提供实践思路。

关 键 词： 大数据技术；财务管理；影响

Discussion on the Influence of Big Data Technology on Enterprise Financial Management

Wang Huijuan

Guangzhou Jinbuhuan Consulting Co., Ltd., Guangzhou, Guangdong 510080

Abstract： Under the background of the deep development of digital economy, the business environment faced by enterprises is complex. The traditional financial management model is difficult to adapt to the new needs of enterprise development because of the lack of data processing ability and single decision-making basis. Big data technology has become an important support for the transformation and upgrading of corporate financial management due to its large scale, diverse types and high processing efficiency. Based on the core connotation and characteristics of big data technology, this paper discusses the specific impact of big data technology on financial decision-making, operating costs and management functions of enterprises. Some strategies are put forward, such as deploying information platform, constructing multi-level financial indicators monitoring system and strengthening data life cycle management mechanism. Big data technology can not only promote the transformation of enterprise financial work from manual operation to intelligent operation, from post-accounting to in-process control, but also reconstruct the financial management paradigm, promote the transformation of financial department from cost center to value creation center, and provide practical ideas for the digital transformation of enterprise financial management.

Keywords： big data technology; financial management; affect

引言

现代市场经济的迅速演变和技术革新的不断前进，使企业遭遇了以往未曾有过的发展难题。财务管理作为企业运营管理的关键部分，其效能和水平直接影响着企业的市场竞争能力和发展空间。一直以来，企业财务管理主要依靠人工操作和经验判断，特别是中小型企业，数据获取范围仅限于内部财务系统，存在着处理效率不高、分析不够及时和深入、预测精准度欠佳等问题^[1]。而大数据技术的兴起和运用，冲破了传统数据处理的技术阻碍。其借助分布式计算、云存储、数据挖掘等先进方式，对大量不同类型的数据进行高效收集、存储和智能分析，从繁杂的数据中提炼出有价值的商业信息^[2]。在这样的形势下，研究大数据技术对企业财务管理的作用，探寻其在财务管理中的应用途径，成为企业突破管理瓶颈、实现数字化转型的重要议题。

一、大数据技术概述

（一）大数据技术

1. 大数据技术的定义

大数据技术是采用分布式计算、云存储、数据挖掘等先进的手段，对海量的异构的数据进行开展采集、实施存储、执行管理、开展分析以及进行应用等操作的一种综合技术^[3]。

2. 大数据技术的主要特征

第一方面，具有规模庞大性。数据的总体数量呈现出按指数级的方式进行增长的态势，单个的数据集合能够达到数百 TB 的量级，远远超过了传统数据库所具备的承载能力^[4]。

第二方面，存在类型多元性。数据的形态包含具有结构化特征的数据库记录、呈现半结构化状态的 XML 文档、属于非结构化性质的社交媒体内容、传感器发出的信号、影像方面的资料等^[5]。

第三方面，具有处理高效性。和人工操作所具有的低效以及误差情况相比较，大数据技术达成了毫秒级的响应以及自动化的处理。系统能够在实时状态下捕捉数据的变化情况，快速地完成清洗、归类、统计等操作，大幅度地提升工作的效率。

第四方面，存在信息完整性。技术平台通过冗余备份以及校验的机制来保障数据的完整状态，避免在传统方式当中因为人为的疏漏而导致信息出现缺失的情况。

第五方面，存在价值密集性。尽管单条数据所具有的价值是有限的，但是通过关联分析以及模式识别的手段，能够从海量的数据当中挖掘出隐藏着的商业规律。企业凭借此来洞察市场的趋势、客户的偏好、运营的瓶颈等情况，把数据资产转化成竞争的优势，对战略的调整以及创新的发展起到支撑的作用^[6]。

二、大数据技术对企业财务管理的影响

（一）赋能财务决策，实现从经验驱动到数据驱动的转型

在传统的模式下，财务部门主要是依靠内部的财务系统来获取相关的数据，信息的来源是单一、滞后的，很难去反映出市场的动态变化情况。大数据的技术打破了这样的局限状况。它借助云计算的平台以及数据的接口，能够在实时的情况下抓取供应链的交易记录、客户的消费行为、行业的政策动向、竞争对手的财报等多个维度的信息，经过智能的算法进行清洗以及整合之后形成立体化的信息资源库。财务管理者不再局限于对历史的财务报表进行分析，开始结合外部的市场数据来进行前瞻性的研判工作。比如通过社交媒体的舆情去预测产品的销量波动情况、依据原材料的期货价格来调整采购的策略、利用区域的经济指标去评估投资项目的可行性。这样全景式的信息视野让财务的决策从经验驱动转变为数据驱动，明显地提升了战略规划的科学性以及时效性^[7]。财务部门在大数据技术的加持下，从“账房先生”转型为“战略参谋”，真正成为企业价值创造的核心驱动力。与此同时，数据获取的自动化机制大幅度地缩短了信息传递的链条，管理层可以在较短时间之内获得完整的经营分析报告，而不是等待月度的财务结算，为快速地响应市场的变化赢得了宝贵的时间。

（二）优化资源配置，显著降低企业财务运营成本

在资金管理层面，传统的财务管理依靠定期进行盘点以及静态的预算，很难捕捉到资金使用的动态变化情况，容易出现部门之间资金闲置与紧张这种失衡的状况。大数据系统能够实时地监控各个业务单元资金的占用情况，通过算法预测未来资金需求的峰谷，自动地生成最优的调配方案。比如，系统分析历史的采购数据，发现原材料价格存在季节性波动的规律，建议企业在低价的时期集中进行采购并且合理地安排库存的周期，仅仅这一项就可以节省数百万元的成本。在费用管控层面，大数据技术能够穿透组织的架构，识别出隐藏的浪费环节。通过对差旅费、办公耗材、能源消耗等支出进行多维度的交叉分析，系统可以发现异常的支出模式，如某部门人均的差旅费用远远超过行业的均值，或者特定时段电力消耗急剧增加却没有产能的提升，为管理层提供成本压缩的具体方向。智能化的供应商评估体系整合价格、质量、交付周期等指标，帮助采购部门筛选出性价比最优的合作伙伴，避免人情采购和暗箱操作所带来的隐性损失，从根本上提升企业的盈利能力^[8]。

（三）推动职能升级，实现财务部门从成本中心到价值创造中心的转变

在传统的财务工作的开展过程之中，大量的时间消耗在对凭证进行整理、对账目进行核对、对报表进行编制等重复性的劳动工作之上。财务人员一直疲于应付基础的业务工作，很难深入地参与到经营分析工作当中。财务共享服务中心的建立彻底地改变了这种情况。标准化的业务流程固化到系统中，数据采集、分类汇总、报表生成等工作进行全自动运行。财务人员从繁琐的计算工作中解放出来，工作的重心开始转向数据解读、趋势研判、方案设计等高质量的活动内容。这种转变情况倒逼财务团队提升综合能力水平。财务人员开始深入地介入到产品定价、成本控制、绩效考核等经营环节当中，以专业的视角为业务决策提供量化的支持内容。实时的数据监控能力使得财务管理从后核算工作转向事中的控制工作，通过设置预警的阈值和自动触发的机制，在问题的萌芽阶段就启动干预的措施^[9]。这种管理模式的革新工作不仅提升了财务工作的效率水平，更重要的是将财务职能嵌入到企业价值创造的全过程当中，真正地实现了从成本中心向利润中心的角色转换。

三、大数据技术促进企业进行财务管理的策略

（一）部署适配企业需求的信息化平台

企业在做系统选型时，不应该随便地去追求功能全面的高端产品，而是要根据自身的规模、业务存在的复杂程度以及管理方面的需求来做理性的评估工作。中小型企业可以优先考虑在云端做部署操作的 SaaS 模式财务软件，这种软件在开始投入的成本比较少，系统的维护工作由服务商负责，能快速实现上线并投入使用。大型集团企业需要定制化的 ERP 系统，该系统要对财务、供应链、人力资源等模块做整合操作，从而实现数据的集中管理以及跨部门的协同作业^[10]。系统部署完成之后，企业应该对原来的

财务流程做优化和重组工作，把报销审批、费用分摊、资产折旧等常规业务嵌入到系统的自动化流程中，以此减少人工干预。在系统的安全防护方面，除了要配置基础防火墙和杀毒软件之外，还应该引入多因素身份认证技术，通过密码加上动态验证码或者生物识别的办法来提高账户的安全程度。针对不同岗位要设置不一样的操作权限，出纳只能查看并处理资金收付业务，成本会计只能访问成本核算模块。总账会计有报表编制的权限，财务总监有全局查看和审批的权限。通过权限的隔离来降低内部舞弊的风险，确保系统能高效、安全且可控。

（二）构建多层次财务指标监测体系

企业需要充分利用大数据技术的多维分析能力，建立一个综合评价的框架，这个框架包括财务基础指标以及拓展指标。在基础指标上，需要确保对资产负债率、流动比率、净资产收益率等传统核心指标进行准确地计算。之后，企业按照自身的行业属性以及经营特点，引入更具有针对性的补充指标。制造型企业将存货周转天数、设备利用率、单位产品能耗等运营效率方面的指标纳入监测的范围之中，通过大数据平台实时地去追踪生产环节当中的资源消耗情况。零售企业应该关注客单价变化的趋势、会员复购率、区域市场渗透度等市场表现方面的指标，借助消费数据的分析去把握市场的动向。企业需要建立指标之间的关联分析机制，通过大数据算法去识别不同指标之间的因果关系以及传导路径。例如，当应收账款周转率出现下降的情况时，系统应该自动地关联分析客户信用等级的变化、销售政策的调整、行业回款周期等相关的因素，帮助管理层快速地定位问题的根源^[11]。同时，指标体系不应该是一成不变的，企业要根据战略的调整以及市场环境的变化，定期地评估指标的有效性。淘汰失去参考价值的指标，补充新的观测维度，确保财务分析始终能够贴合企业发展的实际需求。

（三）强化数据全生命周期管理机制

数据的质量直接决定了财务分析工作的可靠程度，企业必须建立覆盖数据采集、数据存储、数据使用、数据销毁全流程的管理规范内容。在数据采集时期，应该制定统一的数据标准以及录入规则，避免因格式不一致、口径不统一等情况而出现的数据

混乱现象。财务系统需要与业务系统达成无缝式对接状态，通过接口自动地抓取销售订单、采购合同、库存变动等业务方面的数据内容，减少人工录入的环节，降低差错概率^[12]。在数据存储方面，企业构建分类数据库，将原始的交易数据、汇总统计的数据、分析结果的数据进行分层式的管理操作。对于涉及商业机密的财务信息内容，应该采用加密存储的措施。在数据使用的环节，建立数据调用的审批流程以及使用日志的记录机制，追踪每一次数据提取的目的、范围以及使用者，防止数据的滥用以及泄露情况。企业制定数据备份以及灾难恢复的预案，定期将关键的财务数据备份到异地的服务器中。一旦出现系统故障或者网络攻击现象，能够迅速地恢复数据内容，保障财务工作的连续性。

四、结束语

大数据技术是一种超越了财务工具简单更新换代范畴的事物，它把财务部门从大量繁杂又琐碎的事务性工作的束缚当中解脱出来，让它能够转变成为企业价值创造方面的关键动力来源。而这一转变不但体现了管理模式层面的深层次变革，更突出了它在企业整体发展当中的重要战略意义。大数据技术通过对来自内部各个业务环节以及外部市场环境等多个维度的数据进行全面汇聚和整合，能够给企业的财务决策提供一个包含多方面信息的全面视角，从而推动决策模式从传统的主要依靠经验来驱动的模式，逐渐向高度依赖数据支撑的新型模式转变。这种转变在很大程度上增强了决策本身的合理性和及时性，为企业在复杂多变的 market 环境中做出科学决策奠定了坚实的基础。在资金管理领域，大数据技术通过实施实时监控手段以及深入开展成本分析工作，能够实现企业各类资源的精准化配置。这种精准配置不但有力地降低了企业的财务运营成本，还明显提高了企业的整体盈利水平，显示出它在企业资源管理中的强大效能。企业只有紧密结合自身发展的实际情况和战略需求，在信息化平台匹配、财务指标体系搭建、数据全生命周期管理等多个关键环节做好相应的工作，才能充分释放大数据技术所蕴含的巨大价值，让财务管理真正成为企业战略发展进程中的核心动力。

参考文献

- [1] 张志敏. 利用大数据技术提高企业财务管理水平 [J]. 中国商界, 2025, (07): 204-206.
- [2] 王青艳. 大数据背景下企业财务管理水平提升路径研究 [J]. 中国乡镇企业会计, 2025, (05): 176-178.
- [3] 周飞. 大数据背景下企业财务管理数字化转型研究 [J]. 老字号品牌营销, 2025, (04): 182-184.
- [4] 孙惠梅. 大数据技术在企业财务管理中的应用与效益分析 [J]. 中国集体经济, 2025, (05): 153-156.
- [5] 廉西瑞. 数字经济视域下企业财务管理中大数据技术的运用 [J]. 纳税, 2025, 19(02): 88-90.
- [6] 何冬冬. 大数据技术在财务管理中的应用 [J]. 电子技术, 2024, 53(12): 302-303.
- [7] 王晨, 温馨. 大数据时代企业财务管理创新路径探析 [J]. 产业创新研究, 2024, (22): 163-165.
- [8] 邓艳莉. 大数据技术对企业财务管理的创新应用 [J]. 经济, 2024, (11): 94-96.
- [9] 江芳芳. 大数据背景下国有企业财务管理水平提升路径研究 [J]. 中国市场, 2024, (29): 151-154.
- [10] 陈松. 大数据背景下的企业财务管理创新研究 [J]. 张江科技评论, 2024, (09): 156-158.
- [11] 王心茹, 王文霞, 陆星好, 等. 大数据时代的企业财务管理研究 [J]. 江苏科技信息, 2024, 41(8): 104-107.
- [12] 陈俊花. 大数据背景下企业财务风险管理与控制 [J]. 活力, 2024, 42(24): 106-108.

第三方支付平台对跨境电商资金安全的影响探析

姚春光

深圳市联众长信实业有限公司，广东，深圳 518000

DOI:10.61369/ASDS.2026060002

摘 要： 随着数字经济的快速发展，第三方支付平台已成为跨境电商资金流转的核心载体，在优化资金结算流程、搭建专业风控体系、助力合规结 / 售汇等方面发挥了积极作用。部分平台存在资质不全、违规资金的套利、资金清算效率低下等问题，由此引发资金运营风险、安全风险与账户管理风险，对跨境电商的资金安全构成现实威胁。本文从规范资质审核、健全跨境监管协同、优化清算体系三个方面提出对策，以期为支付平台提高风控能力提供参考。

关 键 词： 第三方支付平台；跨境电商；资金安全；影响

Analysis on the Influence of Third-Party Payment Platform on the Safety of Cross-Border E-commerce Funds

Yao Chunguang

Shenzhen Lianzhong Changxin Industrial Co., Ltd., Shenzhen, Guangdong 518000

Abstract： With the rapid development of the digital economy, third-party payment platforms have become the core carriers of cross-border e-commerce fund flows, playing a positive role in optimizing fund settlement processes, establishing professional risk control systems, and facilitating compliant foreign exchange settlement. However, problems such as incomplete qualifications, being exploited for illegal fund arbitrage, and inefficient fund clearing exist on some platforms, giving rise to fund operational risks, security risks, and account management risks that threaten the fund security of cross-border e-commerce. This paper proposes corresponding countermeasures from three aspects — standardizing qualification review, improving cross-border regulatory coordination, and optimizing the clearing system — so as to provide references for payment platforms to enhance their risk control capabilities.

Keywords： third-party payment platform; cross-border e-commerce; fund security; impact

引言

在数字经济迅速发展的大环境下，第三方支付平台成了跨境电商资金流转的主要渠道，其发展规模不断增大，给跨境电商资金安全带来的影响也越来越深远。据统计，2024年我国第三方互联网支付市场交易额在约33.9万亿元人民币左右^[1]。2024年支付宝、微信支付等第三方跨境支付系统跨境电商支付和B2B外贸支付总金额是1.534万亿元^[2]。第三方支付平台在改善资金结算效率、推进合规结，售汇的同时，产生出一系列的资金安全风险，需要引起业界和学术界的关注。

一、第三方支付平台对跨境电商资金安全的积极影响

（一）优化资金结算流程，降低流转风险

第三方支付平台凭借自身电子支付技术的优势，有效地简化了传统跨境支付的复杂流程，降低了资金流转过程中的风险。传统的跨境支付到账时间是3到7个工作日，且受银行营业时间，以及国家外管局外汇申报的影响。第三方支付可以做到秒级或者T+1到账，支持7×24小时不间断运行^[3]。支付平台可以对资金流向进行实时追踪，包括第三方支付平台可以直接将外汇结算至国内私人

人账户，简单高效，整个交易能清楚地呈现资金流转的全过程，有效地降低由于截留、挪用而造成资金流失的风险。此外区块链技术以及智能合约的应用后，进一步增强了资金流动的安全性及效率，提升了资金结算的透明度。平台同跨境电商合作推出支付、物流、供应链一体化服务，从多个维度为跨境电商资金安全提供全方位的支持，使流转风险得到控制。

（二）搭建专业风控体系，防范交易风险

第三方支付平台依靠身份认证、交易监测、反欺诈系统这些技术手段，给跨境电业资金流动构筑起第一道防线，有助于减少

财务损失。区块链技术结合加密支付使资金交易更透明，安全，可以防止企业资金被挪用以及篡改。平台引入人工智能以及大数据技术之后，基于 AI 引发的反欺诈系统可以分析交易规律，识别异常行为，实现实时监控与预警，从而降低财务坏账率^[4]。在跨境交易中，本地化支付效果显著，采用 GrabPay、Paytm 等当地化支付方式更符合当地人的交易习惯，降低了支付失败、拒付的风险，进而提升了企业的财务安全度。

（三）助力合规结售汇，规避政策违规风险

第三方支付平台大多与银行等金融机构深度合作，严格遵循国家外汇管理政策，确保跨境电商资金结售汇流程合法合规，从源头规避政策违规带来的资金风险。同时，平台能够简化企业结售汇操作流程，避免因操作失误、流程不规范导致的政策违规风险，减轻企业合规操作压力。部分支付平台通过完善的合规审核机制、强化交易监管等创新方式，保证数字货币交易资金流动的合法合规性^[5]。平台还借助数字化手段同监管机构建立实时对接，及时把交易信息传送到监管机构，使企业的结售汇过程更加透明，进一步降低政策违规的风险。

二、第三方支付平台对跨境电商资金安全的消极影响

（一）持牌支付平台资质不全面，存在资金运营风险

部分平台因为缺少资质审核的公开透明，甚至用技术手段来规避监管，从而放大了资金运营风险。如果平台常用的跨境资金监管存在漏洞，极易导致资金被挪用，直接危及跨境电商的资金安全。除支付平台自身存在的安全漏洞外，其所对接的电商网站若存在安全隐患，也可能被黑客攻击成为突破口，进而造成电商企业的资金损失。近些年来监管愈加严格，已经有多家支付机构因为合规问题遭到处罚。其中部分支付机构因未满足跨境支付资质要求而被限制业务。此外，一些聚合平台虽然不具备完整的跨境支付资质，却将各种支付渠道整合起来开展业务，给电商的资金造成安全隐患。

（二）平台协助违规资金套利，存在资金安全风险

第三方支付平台容易被不法分子利用来从事违规资金套利活动，不法分子用虚构跨境电商交易、跨境套现等手段来套取资金，给跨境电商资金安全造成了严重的危害。他们还在支付平台上借助外币、虚拟货币、人民币之间的转换模式实现跨境洗钱，达到违规资金套利的目的，以 TW711 平台、火速平台等网站为例，其以虚拟货币泰达币（USDT）为媒介，为客户提供外币与人民币的汇兑服务。这些平台通过非法渠道兑换人民币 2.2 亿余元^[6]。这种行为不仅扰乱金融管理秩序，还会进一步加剧资金安全风险。

（三）平台资金清算效率低下，存在账户管理风险

部分第三方支付平台存在资金清算效率低下的问题，造成资金长时间滞留平台，直接影响跨境电商企业的资金周转效率。同时部分支付平台账户管理不规范，出现多头开户、资金混用等现象，增加了企业的资金管理的难度。此外，支付平台与跨境电商企业之间存在信息不对称，平台对清算进度、账户明细的披露不

够及时，导致企业无法准确掌握资金动态，埋下资金安全隐患。PayPal 资金清零事件是第三方支付平台资金安全风险的典型案例。深圳大卖通拓科技被划扣或冻结资金共计 7849.1 万元，其中被直接划走 5424.5 万元^[7]。充分暴露出第三方支付平台在财务管理方面存在的风险。

（四）平台服务手续费不透明，加剧企业账户财务风险

第三方支付平台收取的服务手续费差异大，存在竞争压力下形成的这种“低费率获客—增值服务创收”的模式，拿主流平台公开的费率来看，以跨境电商卖家最常使用的 PayPal、派安盈（Payoneer）和万里汇（WorldFirst）为例，派安盈 payoneer 一般新用户的话提现手续费是 1.2%，随着累计入账金额越多费用会逐步降低，在实际工作中，甚至有 0 手续费的情况；PayPal 采用分层定价体系：中国大陆账户默认费率为交易金额的 4.4% 加币种固定费用（美元 0.30/ 欧元 0.35），提现成本：渠道差异与隐性风险，直接提现至国内银行单笔固定收取 35 美元；万里汇则通过“0 费率 + 0.3% 封顶”策略将成本压缩至 1% 以内 0.6%）、实时锁汇及突破额度限制等优势，将到账时效缩短至 1-3 个工作日。

对于百万 / 千万 / 亿级卖家来说，不同的第三方支付平台，仅手续费而言一年的差距就高达几万甚至几十万美元的差异，本质上是跨境电商企业账户中隐性流失的成本。

三、应对第三方支付平台对跨境电商资金安全消极影响的策略

（一）严格规范支付平台资质审核，强化平台资金运营风控建设

监管部门应健全审核机制，对第三方支付平台的跨境支付许可及运营资质进行全面核查。支付平台自身也需加强资金运营风控的建设，完善资金池管理规定，明晰资金所有权和使用权，在交易全过程对资金进行全方位的监控，及早发现系统的漏洞并及时加以修复。平台要加快技术更新和革新，运用区块链、大数据等先进的技术手段增强资金运营的安全性及透明度。同时应建立支付机构跨境业务专项备案制度，针对跨境电商场景出台《非银行支付机构监督管理条例》配套细则^[8]，明确资金冻结的触发条件、申诉时限和争议解决机制。

（二）健全跨境监管协同机制，严打违规资金套利行为

加快创建国内外监管机构常态化协同工作机制，统一监管标准和执法程序，联合打击利用支付平台实施的资金套利和洗钱行为，消除跨境监管盲区。同时，需加快制定和完善相关法律法规，对第三方支付平台跨境资金流动的监管边界和平台责任作出明确规定，细化交易审核要求，给监管执法提供明确的依据。可参照 G20 金融科技工作组给出的跨境支付监管建议，依靠国际清算银行发起的全球协同机制，促使各国加强多边监管信息交换以及执法合作^[9]。

（三）优化平台资金清算体系，完善账户规范化管理流程

第三方支付平台应当主动优化资金清算流程，缩短资金滞留时间。同时要建立严格的账户管理制度，对账户的开立、使用、

注销等作出规定，严禁多头开户、资金混用等违规行为。另外支付平台要主动把交易明细、清算进度这些核心信息与跨境电商企业共享，从而改善资金流动的透明度。平台依靠区块链简化清算过程，使资金可以快速到账，从而提高资金清算速度。加强对区块链，智能合约的研发投资。按照央行2024年工作会议关于加强跨境支付系统建设的有关要求，设立专项扶持基金，支持支付机构参与多边央行数字货币桥项目，进一步提高跨境结算效率和透明度^[10]。

（四）将第三方支付平台纳入供应商体系，加强企业财务风险管理

跨境电商企业需要加强资金管理外，还要建立多平台成本比对机制，将第三方支付平台纳入供应商定期评估体系，避免平台费率优势的营销宣传而忽视综合财务风险。第三方支付平台的手续费的变动会直接影响企业的不确定成本，影响企业资金流动性的平稳管控。监管部门应推动建立第三方跨境支付手续费披露的行业标准，保障商户的知情权与选择权。

四、结论

综上所述，第三方支付平台在优化跨境电商的资结算、助力合规结售汇、提升资金周转效率等方面的发挥了积极的作用。区块链、人工智能、大数据等技术的广泛应用可以精准预警和智能风控能力。平台推出新的供应链金融、退税等增值服务，以及和亚马逊、eBay 等大型电商平台的合作模式不断升级，进一步加速了跨境资金的流动性。加密货币、稳定币的广泛应用，数字人民币跨境试点等新的安全问题也随之出现。部分平台的市场垄断行为也因为影响资金安全公平性而受到监管的关注。SWIFT GPI 升级、ISO 20022 标准等规则的改变，对平台合规经营提出了更高的要求。第三方支付平台持续改进风险防控水平，手续费披露标准化建设，多方合力筑牢跨境电商资金安全防线，促进跨境电商的高质量发展。

参考文献

[1] 徐苗. 第三方支付对商业银行中间业务的影响分析 [J]. 中国管理信息化, 2024, 27(18): 164–166.
[2] 朝瑾. 2024 第三方支付行业纵览 [J]. 互联网周刊, 2025, (05): 30–31.
[3] 汤莹. 第三方支付机构跨境电商收款业务及监管对策 [J]. 河北金融, 2022(2): 60–64.
[4] 陈尾云. 第三方支付电子支付为导向的跨境电商供应链金融模式及风险防范 [J]. 三明学院学报, 2024, 41(05): 28–33+104.
[5] 韩旭. 互联网金融中跨境第三方支付行业的风险和合规监管 [J]. 商业观察, 2024, 10(4): 117–120.
[6] 程雪军. 加密数字货币洗钱的技术路径与系统监管 [J]. 国际金融, 2025, (01): 44–52.
[7] 周泽中, 谭楚雄. 第三方支付领域的行政合规管理 [J]. 阴山学刊, 2025, 38(01): 79–86+96.
[8] 尹成远, 王深艳, 郭娜. 非银行支付机构监管的演进逻辑与完善建议 [J]. 河北金融, 2024, (04): 3–9.
[9] 程涣. 《G20 转型金融框架》的主要内容及启示 [J]. 黑龙江金融, 2023, (01): 72–76.
[10] 万存知. 以强大的货币为基石助力金融强国建设 [J]. 金融博览, 2025, (05): 18–20.

集团化企业合并报表编制难点与内部交易抵消实务

郑婉文

广州市番禺粮食储备有限公司，广东 广州 511400

DOI:10.61369/ASDS.2026060003

摘 要： 随着国有粮食企业集团化发展，合并报表成为反映整体财务状况的核心工具。然而，组织架构复杂、内部交易频繁、信息化水平参差不齐等因素导致合并报表编制面临诸多难点，尤其内部交易抵消处理极易出错。本文以 A 集团公司为例，系统梳理合并报表编制中的组织协调、数据收集、抵消处理等实务难点，提出建立标准化数据收集机制、构建内部交易台账、推进信息化建设、加强业财融合等优化对策。文中通过内部交易抵消逻辑图和内部交易类型及抵消方法示例表直观展示抵消处理流程，并借助合并报表系统功能需求对比表为信息化建设提供参考。本文旨在为同类企业提供可操作的合并报表编制参考，提升财务管理精细化水平。

关 键 词： 集团化企业；合并报表；内部交易抵消；财务信息化；国有粮食企业

Consolidated Financial Statements Preparation Challenges and Intercompany Transaction Elimination Practices in Group Enterprises

Zheng Wanwen

Guangzhou Panyu Grain Reserve Co., Ltd., Guangzhou, Guangdong 511400

Abstract： With the group-based development of state-owned grain enterprises, consolidated financial statements have become a core tool for reflecting the overall financial position. However, factors such as complex organizational structures, frequent intercompany transactions, and uneven levels of informatization pose numerous difficulties in the preparation of consolidated financial statements, with intercompany transaction elimination being particularly prone to errors. Taking Group A as an example, this paper systematically reviews practical challenges in the preparation of consolidated financial statements, including organizational coordination, data collection, and elimination processing. It proposes optimization measures such as establishing standardized data collection mechanisms, building intercompany transaction ledgers, advancing informatization, and strengthening business finance integration. The paper visually illustrates the elimination process using a logic diagram of intercompany elimination and a table of intercompany transaction types and corresponding elimination methods. Additionally, a comparison of functional requirements for consolidated reporting systems is provided as a reference for informatization development. This study aims to offer practical guidance for similar enterprises in preparing consolidated financial statements and enhancing the precision of financial management.

Keywords： group enterprises; consolidated financial statements; intercompany transaction elimination; financial informatization; state-owned grain enterprises

引言

合并报表是集团企业对外披露财务状况、经营成果和现金流量的法定要求，更是内部管理决策的重要依据。对于承担国家粮食储备任务的国有粮食企业集团而言，业务链条长、内部交易频繁、资金流量大，合并报表编制的准确性和及时性直接影响企业资源配置效率和风险防控能力。然而，实务中由于组织层级多、数据分散、业财脱节等问题，合并报表编制常陷入“数据难收集、抵消难准确、效率难提升”的困境^[1]。本文结合 A 集团公司（以下简称“集团公司”）的实际情况，深入剖析合并报表编制中的难点，并围绕内部交易抵消这一核心环节，提出切实可行的实务对策，也为同类企业提供有益借鉴。

一、集团化企业合并报表编制的组织架构与业务特点

A 集团公司与 B 公司实行一套人马两块牌子管理模式，总部设 7 个职能部门。合并范围共涉及 23 家企业，其中子公司 21 户和 B 公司由本集团公司负责运营和管理。另外，B 公司及其下属 4 个分公司需先编制“小合并”，再纳入集团“大合并”。这种多层次组织架构，要求合并报表编制必须清晰界定控制关系，确保并表单位完整、合并层级清晰。

集团业务涵盖粮油储备和轮换、乳酸菌饮料制造和物业出租等多个领域，由此衍生出丰富的内部交易类型，包括粮油购销、资金拆借、使用权资产租赁、支付费用、代垫款项、计提分配股利及借款利息等，每年发生的内部交易笔数和涉及金额较大。这些交易涉及资产负债表、利润表、现金流量表的多个项目，若不能准确抵消，将导致合并报表资产、负债、收入、费用等项目的虚增或虚减。

在此背景下，合并报表能够消除内部交易影响，真实反映集团整体财务状况和经营成果，为管理层战略制定、资源配置提供可靠依据，同时也是上级考核、审计监督的重要基础^[1]。因此，合并报表编制必须遵循实质重于形式原则，确保数据准确、抵消完整、披露充分，满足及时性、准确性、完整性和可理解性的要求。

二、合并报表编制中的主要难点分析

（一）组织协调与业财脱节导致数据收集困难

集团内部部分下属公司对合并工作重视不够，认为这是总部任务，数据报送拖沓，尤其在月末、年末结账期，往往延迟报送，导致总部合并时间紧迫。更突出的问题是业财融合不足：业务部门签订内部交易合同时，未及时将合同信息传递财务，财务往往在收到发票或付款申请时才得知交易发生，无法提前规划抵消处理；交易双方记账口径不一致，例如内部代垫水电费收入，代垫方计入“其他业务收入”，支付方计入“管理费用”，导致抵消时科目不匹配；各单位未定期进行内部交易对账，问题积累到月末才发现。又如某月 a 下属公司上报了与 b 下属公司的内部交易，但 b 下属公司未上报，月末合并抵消时才发现差异并排查是哪家属公司出现错误，经多次沟通才达成一致，导致严重影响合并进度。

（二）内部交易抵消处理复杂且依赖手工操作

内部交易抵消是合并报表核心环节，也是最易出错环节。实务中常见问题：记账时间差异导致单边挂账，如销售方已确认收入并开具发票，采购方因发票未到未入账；金额不一致，因折扣、返利、运费等因素导致双方记账金额存在差异；科目对应不清，如同笔资金拆借，一方计入“其他应收款”，另一方计入“短期借款”；历史挂账长期未清理，部分内部往来款因年代久远、经手人变动后未及时梳理导致长期挂账，双方余额不符难

以核实；特殊交易处理复杂，如拨付贷款采购粮油、支付委派人员劳务费、使用权资产租赁和计提分配股利等，抵消涉及多个科目^[2]。目前集团尚未建立统一合并报表系统，抵消处理以 Excel 为主，各单位上报单体报表和内部交易明细表，财务人员手工编制抵消底稿逐笔核对，再手工编制抵消分录录入模板，反复检查调整直至平衡。这种方式效率低下，每月合并耗费大量人力；容易出错，手工操作难免疏漏；无法动态监控，只能事后处理；缺乏审计痕迹，不利于追溯。例如一次月度合并中，因手工漏抵一笔 200 万元内部销售，导致合并收入虚增，直到年度审计时才被发现更正。

（三）预算管理与合并报表衔接不畅

预算编制与执行情况未能有效反馈至合并报表编制中，导致预算数据与合并结果脱节。部分下属公司预算执行偏差大，但合并报表未能及时反映偏差原因，影响集团对整体经营情况的判断。某年度预算执行情况内部审计报告指出，部分下属公司未完成年度财务预算报告和决算报告，缺乏对预算执行情况的分析说明，这使得合并报表难以发挥预算控制作用，削弱了财务管理的闭环效应。此外，预算数据未融入合并流程，无法实现预算与实际的自动对比分析，管理层难以及时掌握预算执行动态，影响经营决策的针对性和有效性。

三、内部交易抵消实务操作要点

内部交易抵消的准确与否，直接决定合并报表的真实性。实务中需从交易识别、数据收集、分录编制、历史清理到特殊事项处理等环节逐一规范，形成闭环管理。

（一）内部交易识别与数据收集

准确识别内部交易是抵消的基础。集团应建立内部交易申报制度，要求各单位在交易发生时及时填报《内部交易明细表》，内容包括交易日期、交易类型（如购销、资金拆借、租赁、费用支付等）、交易对手单位名称、交易金额、合同编号、发票号码、记账凭证号，并在期末填写对方记账情况。财务部门每月汇总所有明细表，交叉核对，确保无遗漏^[3]。同时，应建立定期对账机制，各单位每月 5 个工作日内完成上月内部交易对账，双方确认一致后签字留存；发现差异须在当月处理完毕，避免问题积累。通过这种日常化的数据收集和对账，可将抵消风险化解在萌芽状态。

（二）抵消分录的编制原则与方法

抵消分录编制应遵循“谁销售、谁确认收入，谁采购、谁确认成本”的基本原则，按交易类型分别处理。各类常见内部交易的抵消方法如表 1 所示。实务中需注意：对于购销类交易，不仅要抵消收入和成本，还需同步抵消对应的往来款项；对于资金拆借，需区分本金和利息分别处理；对于费用支付和代垫款项，要确保双方记账科目对应，否则应先调整再抵消。此外，同一笔交易可能涉及多个抵消分录，需逐笔核对，防止漏抵或重抵。

表 1：内部交易类型及抵消方法示例表

交易类型	交易方 A (销售 / 服务提供方)	交易方 B (采购 / 服务接受方)	合并抵消分录
粮油购销	借：应收账款——B 贷：主营业务收入 合并税金差异	借：管理费用 / 主营业务成本 贷：其他应付款 / 应付账款——A	借：主营业务收入 合并税金差异 贷：管理费用 / 主营业务成本 借：其他应付款 / 应付账款——A 贷：应收账款——B
资金拆借 (本金)	借：其他应收款——B 贷：银行存款	借：银行存款 贷：短期借款 / 长期应付款 / 其他应付款——A	借：短期借款 / 长期应付款 / 其他应付款——A 贷：其他应收款——B
资金拆借 (利息)	借：其他应收款——B (利息) 贷：其他业务收入 (利息收入) 合并税金差异	借：财务费用 (利息支出) 贷：应付利息——A (利息)	借：其他业务收入 (利息收入) 合并税金差异 贷：财务费用 (利息支出) 借：应付利息——A (利息) 贷：其他应收款——B (利息)
使用权资产租赁	每月： 借：其他应收款 / 应收账款——B 贷：其他业务收入 合并税金差异	合同签订后： 借：使用权资产 租赁负债 (未确认融资费用) 贷：租赁负债 (租赁付款额) 每月： 借：管理 / 销售费用 (折旧费) 贷：使用权资产累计折旧 借：财务费用 (未确认融资费用) 贷：租赁负债 (租赁付款额) 借：其他应付款 / 应付账款——A	借：其他业务收入 合并税金差异 租赁负债 (租赁付款额) 使用权资产累计折旧 贷：管理 / 销售费用 (折旧费) 租赁负债 (未确认融资费用) 财务费用 (利息支出) 使用权资产 合并价差 借：其他应付款 / 应付账款——A 贷：其他应收款 / 应收账款——B
管理费支付	借：其他应收款——B 贷：其他业务收入 (管理费) 合并税金差异	借：管理费用 贷：其他应付款——A	借：其他业务收入 (管理费) 合并税金差异 贷：管理费用 借：其他应付款——A 贷：其他应收款——B
代垫款项	借：其他应收款——B 贷：其他业务收入 (代垫款项) 合并税金差异	借：管理费用 贷：其他应付款——A	借：其他业务收入 (代垫款项) 合并税金差异 贷：管理费用 借：其他应付款——A 贷：其他应收款——B
计提分配股利	借：应收股利——B 贷：投资收益	借：利润分配——应付股利 贷：应付股利——A	借：投资收益 贷：利润分配——应付股利 借：应付股利——A 贷：应收股利——B

(三) 历史挂账与往来款清理

历史挂账是内部交易抵消的“顽疾”。由于经手人变动、原始单据缺失等原因，部分往来款长期挂账且双方余额不符，直接影响抵消质量。集团应尽快聘请会计师事务所进行专项审计，由财务部牵头，业务部门配合。首先，全面梳理所有内部往来科目余额表，逐笔分析挂账原因、时间和经办人等；其次，与交易对手方逐笔核对，查找差异根源；然后，分类处理：属于记账错误的原因应及时调整账务，因未及时结算的应督促尽快结算，存在争议的地方应协调解决，确实无法收回或无需支付的款项按规定程序报批并根据专项审计报告和相关批复进行核销；最后，建立清理台账，跟踪后续情况，防止新挂账再次积累。清理结果应在合并报表附注中充分披露，以增强信息透明度^[4]。

(四) 特殊事项处理要点

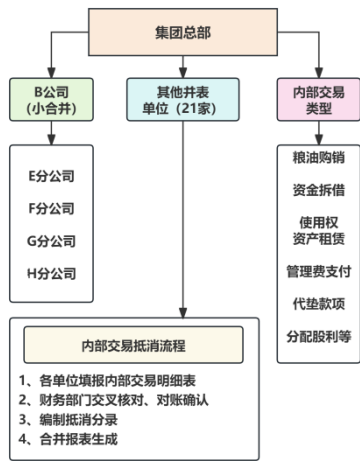
除常规交易外，集团还存在一些特殊内部事项，处理不当极

易出错。例如，B公司向银行贷款后拨付下属分公司用于采购粮油，下属分公司确认为负债（其他应付款），B公司确认为债权（其他应收款），合并时应抵消相关款项。又如，A公司支付给B公司委派至其他公司的人员劳务费或B公司支付下属分公司相关管理费，合并时需将收入、费用等同步抵消。再如，C公司向D公司支付使用权资产租金，D公司确认收入，C公司确认未确认融资费用、租赁付款额、财务费用和折旧费，合并时应抵消以上会计科目，并同步抵消剩余使用权资产和租赁负债，合并价差计入资本公积。这些特殊事项往往涉及多个科目，需在编制抵消分录时格外谨慎，必要时可编制操作指引，统一全集团处理口径。

(五) 内部交易抵消的逻辑流程

为直观展示内部交易抵消的全过程，下图汇总了从内部交易发生到合并报表生成的逻辑关系。该图以集团组织架构为基础，展示了不同层级单位间的交易类型、数据收集与核对、抵消分录

编制及最终合并报表生成的关键节点，可作为实务操作的参考流程图。



图：集团内部交易抵消逻辑图

通过上述操作要点，集团可将内部交易抵消工作从被动纠错转向主动规范，为高质量合并报表奠定基础。

四、合并报表编制与内部交易抵消的优化对策

针对前述难点，集团应从数据标准化、信息化建设、业财融合和闭环管理四个维度系统优化，提升合并报表质量与效率。

表2：合并报表系统功能需求对比表

功能模块	需求描述	重要性
数据采集	支持多种数据源接入，自动取数、校验	★★★★★
合并范围管理	灵活定义合并范围、合并层级，支持多级合并	★★★★★
内部交易对账	支持在线对账、差异处理、自动生成抵消分录	★★★★★
抵消分录处理	自动生成标准抵消分录，支持手工调整	★★★★☆
合并报表生成	自动生成合并资产负债表、利润表、现金流量表及附注	★★★★★
多维度分析	支持按板块、区域、业务等维度分析	★★★☆☆
审计追踪	保留所有调整痕迹，便于审计	★★★★☆

（三）加强业财融合，确保交易数据一致

推行合同管理财务前置，所有内部交易合同签订前须经财务部门审核，明确交易条款对账务处理影响。合同签订后业务部门须在3个工作日内（不可跨月）将合同副本提交财务备案。统一记账口径，财务部编制操作指引，明确双方记账科目和处理规则。每月组织由各相关业务部门参加的对账会，通报内部交易执行情况，协调解决差异^[6]。对历史挂账及时组织专项清理，逐笔分析原因、分类处理，清理结果在合并报表附注中披露。

（四）构建预算—执行—合并闭环管理体系

预算编制阶段，要求各单位按合并口径编制年度预算并分解到月度，预算项目与会计科目对应。预算数据导入合并报表系统作为执行监控基准。每月合并报表生成后，系统自动对比实际数

（一）建立标准化数据收集机制，夯实合并基础

数据源头质量决定合并报表质量。集团应统一全集团会计科目体系、辅助核算项目和报表格式，确保数据口径一致。制定《合并报表数据报送手册》，明确报送时间、内容、格式要求，将数据质量纳入下属公司绩效考核。在数据收集环节内置校验规则，如勾稽关系检查、合理性检查，发现问题及时退回修改。建立内部交易申报制度，要求各单位在交易发生时填报《内部交易明细表》，财务部门每月汇总交叉核对。建立定期对账机制，各单位每月5个工作日前完成上月内部交易对账，双方确认一致后签字留存，发现差异须当月处理完毕。同时制定《集团合并报表管理办法》^[5]，明确各部门职责、工作流程、时间节点和考核标准，使合并工作有章可循。定期组织合并报表培训，提高全员专业素养，建立问题反馈机制持续优化流程。

（二）推进合并报表信息化建设，提升处理效率

引入专业合并报表系统是解决手工操作弊端的根本途径。集团应根据业务特点选择合适系统，核心功能需求如表2所示。实施需分步推进：需求调研与方案设计、系统选型采购、数据标准化整理、系统开发测试、历史数据迁移、用户培训上线、持续优化。关键在系统集成，应与ERP、资金系统、预算系统对接，实现数据自动传输，避免人工重复录入；与银行系统对接实时获取资金流水，便于资金类交易对账。系统应具备审计追踪功能，保留所有调整痕迹，便于审计核查。

与预算数，对偏差超过10%的项目自动预警，财务部门组织专题分析形成报告。年度决算时，各单位编制预算执行情况分析报告，集团汇总形成年度分析报告作为下年度预算提供依据^[7]。将预算执行情况纳入下属公司绩效考核，强化预算约束力，使合并报表真正发挥管理工具的作用。

五、结语

集团化企业合并报表编制是一项系统性、复杂性的财务工作，涉及组织协调、数据处理、信息化建设等多个方面。内部交易抵消作为合并报表的核心环节，直接影响报表的真实性和决策支持功能。国有粮食企业应结合自身业务特点，优化合并流程，

推进信息化建设，加强业财融合，提升合并报表的编制效率和准确性。唯有如此，才能在保障国家粮食安全的同时，实现企业财务管理的精细化、科学化和现代化^[8]。通过建立标准化数据收集机制、构建内部交易台账、引入合并报表系统、强化预算与合并的衔接等措施，A 集团公司可以逐步解决当前合并报表编制中的难点，为管理层提供更高质量的财务信息，助力企业高质量发展。

参考文献

[1] 孙丹. 集团企业合并财务报表编制的难点及应对策略探究 [J]. 中国乡镇企业会计, 2025, (04): 179-181.
[2] 曾阳春. 集团企业合并报表编制的难点与优化建议分析 [J]. 市场周刊, 2024, 37(18): 108-111.
[3] 徐凤仙. 企业合并报表的编制与难点解析 [J]. 中国中小企业, 2023, (04): 99-101.
[4] 邱修文. 国有企业合并财务报表编制的难点与优化路径 [J]. 上海企业, 2025, (11): 253-255.
[5] 高凯. 合并报表财务信息是否可靠内部交易抵消是关键 [J]. 中国商人, 2025, (20): 34-36.
[6] 姚莎莎. 内部交易对账问题分析 [J]. 合作经济与科技, 2025, (18): 157-159.DOI: 10.13665/j.cnki.hzjjykj.2025.18.030.
[7] 高才清. 权益法下顺流交易和逆流交易中会计处理的底层逻辑 [J]. 大众投资指南, 2025, (03): 98-100.
[8] 於晨婧. 企业内部存货交易的合并报表抵销处理研究 [J]. 商业 2.0, 2024, (26): 73-75.

农牧行业集团内控有效性评价体系构建与实施路径研究 ——以 A 集团下属子公司自查实践为例

丁兰

温氏食品集团股份有限公司, 广东 广州 510630

DOI:10.61369/ASDS.2026060016

摘 要 : 农牧行业因生产链条覆盖范围广、生物资产占比高、基层业务分散等固有属性, 内控体系在子公司层面落地难度显著高于传统制造业。2026 年国资委 2 号文和 15 号文出台后, 内控评价标准正从“制度是否健全”转向“能否支撑智能化穿透式监管”。本文立足内部审计第三道防线定位, 结合 A 集团内控评价的顶层设计与子公司自查经验, 将穿透式监管要求融入评价体系, 构建了涵盖“评价标准统一化、实施模式分层化、整改闭环全流程化、穿透能力系统化”的四维框架, 并梳理出调研前置、底稿标准化、缺陷分级管控、穿透能力评估、规则模型建设等落地路径。实践表明, 该体系可精准识别饲料私下流转、生物资产死淘确认不规范等农牧行业特有风险, 对农牧企业集团强化风控能力具有参考价值。

关 键 词 : 农牧企业集团; 内控有效性评价; 穿透式监管; 内部审计; 自查实践

Research on the Construction and Implementation Path of the Internal Control Effectiveness Evaluation System in Agricultural and Animal Husbandry Industry Groups — A Case Study of Self-Inspection Practices of Subsidiaries of Group A

Ding Lan

Wens Foodstuff Group Co., Ltd., Guangzhou, Guangdong 510630

Abstract : Due to the inherent attributes of the agricultural and animal husbandry industry—including a wide production chain coverage, a high proportion of biological assets, and dispersed grassroots business—the internal control system is significantly more difficult to implement at the subsidiary level than in traditional manufacturing. After the issuance of Documents No. 2 and No. 15 by the State-owned Assets Supervision and Administration Commission in 2026, the internal control evaluation standards are shifting from "whether the system is sound" to "whether it can support intelligent, penetrative supervision." Based on the positioning of the third line of defense in internal audit, combined with Group A's top-level design of internal control evaluation and subsidiary self-inspection experience, this paper integrates penetrative regulatory requirements into the evaluation system, constructing a four-dimensional framework covering "unified evaluation standards, layered implementation models, full-process closed-loop rectification, and systematic penetration capabilities," and outlines implementation paths including pre-research preparation, standardized working papers, graded defect control, penetration capability assessment, and rule model construction. Practice shows that this system can accurately identify unique risks in the agricultural and animal husbandry sector, such as private feed circulation and irregular confirmation of biological asset death, providing valuable reference for agricultural and animal husbandry enterprise groups to strengthen risk control capabilities.

Keywords : agricultural and animal husbandry enterprise group; evaluation of internal control effectiveness; penetrative supervision; internal audit; self-inspection practice

引言

(一) 研究背景与意义

近年来, 国内农牧行业规模化进程持续加快, 分子公司数量快速增长, 业务覆盖种养殖、食品加工、终端销售等多个环节, 内控体

系的有效性直接决定集团经营安全底线。农业农村部2025年印发《农业产业化龙头企业合规经营指引》，要求农牧企业建立覆盖全业务链条的内控评价机制^[1]。2026年，国资委连续出台《关于加强中央企业穿透式监管的指导意见（试行）》（2号文）和《关于做好2026年中央企业内部控制体系建设与监督工作有关事项的通知》（15号文），明确将“实现智能化穿透式监管”作为衡量内控体系有效性的核心依据^[2、3]。从行业实际看，多家农牧集团因子公司内控执行不到位引发生物资产损失、养殖数据造假等事件，建立适应穿透式监管要求的内控评价体系已迫在眉睫。与过去相比，内控建设的逻辑正发生根本转变：过去强调“制度有没有、流程全不全”，现在更关注“数据是否贯通、系统是否联通、风险是否可预警、整改是否能闭环”。

（二）国内外研究现状

国际层面，COSO于2025年发布《内控框架农业行业适配指南》，在原有五要素基础上新增生物资产管控、产业链分散风险识别两项专项原则，为涉农企业内控体系建设提供了针对性标准^[4]。Smith等（2024）针对全球农牧集团的实证研究表明，适配行业特性的内控评价体系可使风险识别效率提升42%^[5]。国内层面，李军等（2025）提出集团化农牧企业内控评价需兼顾“合规底线”与“行业特性”双重目标，建立分层分级的差异化指标体系^[6]。王佳等（2026）基于12家头部农牧集团审计数据，构建了包含生物资产死淘确认、饲料流转管控等特色指标的评价体系，验证了行业适配指标对提升缺陷识别精准度的显著作用^[7]。2025年《农牧行业内控白皮书》显示，当前仅38%的规模化农牧集团建立了适配行业特性的评价机制，实践需求迫切^[8]。2号文和15号文进一步将评价焦点从“制度是否存在”升级为“数据是否贯通、系统是否联通、风险是否可预警、整改是否能闭环”，标志着内控建设从“制度导向”向“穿透能力导向”的根本转变^[2、3]。

（三）研究内容与方法

本文采用理论推演与案例实证相结合的方法，首先梳理农牧行业内控核心特征与现存痛点，其次将穿透式监管要求融入评价体系框架，再结合A集团子公司内控自查实操案例，提出具体落地路径与保障措施。研究数据主要来源于集团正式发布的内控制度、自查方案与工作底稿，保障研究结论的可操作性与可复制性。

一、农牧行业内控核心特性与现存问题分析

（一）农牧行业内控的核心特性

与其他行业相比，农牧行业的内控管理存在三个显著的差异化特征。第一，业务场景分散化特征突出，管控穿透难度较大。规模化农牧集团的养殖基地、乡镇服务站通常布局在偏远县域，一线业务人员规模大、流动性高，集团内控要求传导至基层容易出现执行偏差，饲料私转、生产数据填报不实等问题隐蔽性较强，常规检查手段难以有效识别。第二，生物资产占比高，管控节点具有特殊性。活禽、生猪等生物资产的计量核算、死淘确认、投入品领用等环节涉及大量人工判断空间，内控关键控制点与制造业存在明显差异，容易出现管理漏洞与合规风险。第三，产业链条跨度大，风险传导效应明显。从种源繁育、畜禽养殖、食品加工到终端销售，任意环节出现内控缺陷都可能向全链条传导扩散，例如养殖环节的生物安全管控不到位，可能导致整个生产计划中断，给集团造成重大经济损失。这些特性决定了农牧集团内控体系必须从“静态制度设计”转向“动态在线穿透”，方能实现全级次、全业务、全系统的有效管控。

（二）农牧集团下属子公司内控现存共性问题

结合A集团历年内部审计发现的问题，下属子公司在内控执行层面主要存在三类共性问题。一是内控标准执行存在“上下脱节”问题。集团层面发布的内控制度在子公司落地时，经常出现“选择性执行”的情况，部分基层业务环节存在“制度规定与实际操作两张皮”的现象，例如生物资产死淘确认流程流于形式，实际操作中并未严格执行现场核验要求。二是风险识别存在滞后

性，以被动应对为主。多数子公司的内控监督依赖事后审计，缺乏常态化的风险前置预警机制，往往在问题已经造成实际损失后才介入整改，2023至2025年A集团下属子公司披露的内控缺陷中，超过60%是在造成实际损失后才被发现。三是系统数据贯通不足，存在信息孤岛。采购、养殖、财务等关键业务系统尚未实现有效联通，同一事项在不同系统中的编码和口径不一致，难以支撑全量数据的实时采集与横向关联，这是制约穿透式监管落地的深层瓶颈。此外，部分子公司内控岗位人员配置不足，缺乏针对农牧行业特色业务的专项培训，对生物资产核算、养殖流程管控等特殊场景的风险识别能力不足。

二、农牧集团内控有效性评价体系框架构建

（一）体系建设核心目标

立足内部审计第三道防线的独立监督定位，并适应穿透式监管的新要求，内控有效性评价体系建设需紧扣三个核心目标。实质有效性，即评价标准需充分贴合农牧行业业务实际，能够真正识别出高风险问题，避免评价工作陷入形式主义。闭环管控，即构建从问题识别、反馈、整改到回头看的全流程管控机制，确保所有内控缺陷能够见底清零。以评促改，即通过评价结果的刚性应用，倒逼业务部门优化流程、完善制度，实现“评价—整改—提升”的良性循环。此外，需新增“穿透有效性”目标，按照15号文提出的“全级次覆盖、全业务在线、全系统联通、全要素监管”要求，重点衡量内控体系是否具备支撑智能化穿透式监管的能力^[9]。

（二）体系设计核心原则

评价体系设计需遵循五项基本原则。行业适配性原则，将农牧行业特有的生物资产管理、合作养殖户运营、投入品管控等业务场景纳入核心评价范围，避免直接套用通用制造业评价标准。独立性原则，评价工作由集团审计部牵头组织开展，保持第三道防线的独立性，确保评价结果客观公正，不受业务部门干预。可操作性原则，评价标准、实施流程、操作工具需明确具体，子公司可直接对照开展自查工作，无需额外进行复杂的适配转化。成本效益原则，评价工作聚焦高风险领域，优先覆盖饲料私转、生产数据造假、养殖户补贴发放等核心风险点，合理控制评价实施成本。穿透性原则，按照2号文要求，将组织穿透、业务穿透、数据穿透作为评价体系设计的重要维度，重点考察制度是否支持逐级穿透和责任回溯、系统是否联通、数据是否可横向关联和纵向追溯^[2]。

（三）体系核心构成

农牧集团内控有效性评价体系由六大核心模块构成。统一评价标准模块，制定覆盖全业务、全流程、全场景的评价标准，明确关键控制点、评价方法、缺陷分级认定规则，形成可量化、可核验、可追溯的评价依据，针对农牧行业特色业务单独设置专项评价指标，例如生物资产死淘确认流程合规性、合作养殖户保证金足额缴纳比例等。分层评价实施模块，采用“不定期全面评价、重点专项评价、动态实时监测”相结合的评价模式，对全级次分子公司、高风险业务条线开展穿透式监督评价，实现事前预警、事中监控、事后评价全链条覆盖。缺陷闭环管控模块，构建“评价—反馈—整改—回头看”全流程闭环管理机制，对内控缺陷按一般、较大、重大分类管理，明确责任主体、整改时限与验收标准，全程跟踪整改进度。评价结果运用模块，将内控有效性评价结果纳入子公司经营绩效考核、年度审计计划制定、干部履职评价的核心依据，以评价结果倒逼业务与内控部门完善制度、优化流程。三道防线协同模块，建立审计部门与业务执行部门、内控统筹部门常态化协同工作机制，明确职责边界、信息共享、问题共判、整改共推的规则，形成“前端风险防控、中端流程管控、后端监督评价”的一体化内控治理格局。穿透能力评估模块，这是适应穿透式监管新增的核心模块，重点评估六个维度：组织上能否穿透——总部对各级单位的责任链条是否清晰；制度上能否穿透——章程、授权、合同是否支持逐级穿透和责任回溯；流程上能否穿透——关键业务是否真正在线流转，是否存在线下绕行；系统上能否穿透——采购、资金、财务等系统是否打通，是否存在信息孤岛；数据上能否穿透——数据口径是否统一、主数据是否干净、字段是否可关联；监督上能否穿透——异常能否自动识别、线索能否自动派单、整改能否自动跟踪。

三、内控有效性评价的实施路径——以 A 集团下属子公司自查实践为例

国内头部畜禽养殖集团下属分子公司数量超过150家，2026年集团审计部牵头组织开展下属子公司内控有效性评价工作，采

用“集团统一制定方案—子公司自主开展自查—结果统一汇总分析—问题跟踪整改闭环”的实施模式。

（一）前期准备阶段：明确方案与工具

集团层面首先制定并下发《内控及经营管理审计自查实施方案》，明确三项核心内容。清晰界定自查范围，时间覆盖近3个完整会计年度，异常交易可向前追溯；业务采用“2+2+N”模式，将财务内控、工程管理作为必查核心，同时覆盖销售收款、养殖户管理、采购供应、资产管理、合同管理等流程，重点聚焦禽肉销售、合作养殖户管理等高风险领域。标准化自查工具，统一设计分层调研提纲和标准化工作底稿，采用“模块分级”设计逻辑，针对行政、财务、生产、客服、工程五大核心业务模块分别梳理三级检查要点：一级要点对应模块整体管控有效性，二级要点对应核心流程合规性，三级要点对应关键岗位操作规范性，每个要点明确检查方法与记录规范，生产模块专门设置生物资产死淘确认合规性检查项。同时统一设置标准化数据采集字段，涵盖业务流水号、交易时间、关键岗位人员、审批节点、佐证材料索引五大核心维度，明确不同业务模块最低抽样量要求。自查结论采用5级分级标准，从1级完全符合到5级严重不合规。明确实施节奏，整体自查周期共8周，分为前期准备1周、现场实施4周、成果提交2周、整改跟踪1周。此外，按照15号文要求，新增“穿透能力评估”专项工具，围绕组织、制度、流程、系统、数据、监督六个维度设计评估问卷和测试清单，用于判断各子公司信息系统的穿透性是否满足监管要求^[3]。

（二）自查实施阶段：分层推进穿透核查

子公司收到实施方案后，成立专项自查工作小组，按照以下步骤有序开展工作。调研问卷填报，首先组织管理层和各业务部门负责人完成问卷填报，管理层说明近三年经营成果、核心优势、主要挑战与合规风险情况，业务部门梳理本部门KPI完成、流程执行、现存风险点等内容，通过问卷快速识别子公司经营层面的异常点与高风险领域。核心流程测试，对照标准化工作底稿开展穿行测试、抽样检查和数据分析，生产养殖户管理模块重点检查死淘确认合规性、保证金足额缴纳、饲料转料管理，财务模块重点检查销售回款流程、成本核算准确性，测试过程详细记录步骤、问题、金额与证据索引，确保问题可追溯、可核验。针对生物安全、工程结算、账款管理三类特殊业务设置专项管控要点，要求样本全覆盖、证据全留痕、核查全穿透。问题初步分级，按照缺陷分级认定规则对发现问题进行分类：一般缺陷指单个流程执行不规范、影响范围局限于单个岗位、未造成直接经济损失且无合规风险的流程瑕疵；较大缺陷指核心流程存在管控漏洞、影响范围覆盖单个业务模块、可能造成10万元以下经济损失或存在潜在合规风险的问题；重大缺陷指多个核心流程管控失效、影响范围覆盖多个业务模块、已造成10万元以上经济损失或存在明确合规风险的问题。所有缺陷同步记录问题描述、影响范围、潜在风险、责任主体四项核心信息，为后续整改提供清晰依据。新增穿透监管测试环节，重点测试系统是否实现数据实时采集和横向关联、重点领域是否存在穿透缺陷、异常交易能否被系统自动识别预警，以及预警后是否形成派单、核查、反馈、整改

的完整闭环。

（三）成果汇总与整改阶段：闭环管控确保落地

子公司完成自查后，将工作底稿、调研问卷、相关证据统一提交至集团审计部，集团安排专人审核，对问题描述不清、证据不充分的要求补充完善，保障自查成果质量。针对自查发现的问题实行分级分类管控，重大问题由集团审计直接跟踪督办，纳入后续现场审计重点范围，深入核查问题根源；较大问题由集团审计与子公司共同跟踪整改；一般问题由子公司自行整改，整改结果在后续审计中复核。按照“谁主管谁负责”原则落实整改责任，涉事业务单位为整改责任主体，单位负责人为第一责任人；针对制度设计层面缺陷，由集团对应职能部门牵头优化制度流程。针对穿透能力评估中发现的系统联通不足、数据口径不一致等深层次问题，由集团信息部门和业务管理部门联合制定整改方案，推动制度数字化改造和数据治理工程，将制度要求从“纸面”翻译为系统可执行的规则模型，确保整改不仅修复个案问题，更从根源上消除穿透盲区。

（四）评价结果运用：强化刚性约束

集团层面将内控有效性评价结果与三项核心管理机制直接挂钩。与绩效考核挂钩，内控评价得分占子公司管理层年度绩效考核权重的15%，评价结果为“不合格”的子公司管理层年度考核不得评为优秀等次。与审计计划挂钩，内控评价发现问题较多、风险等级较高的子公司，作为下一年度审计重点关注对象，增加现场审计频次与资源投入。与流程优化挂钩，对自查发现的共性问题进行汇总分析，定期向集团业务管理部门反馈，推动制度流程的统一优化，例如针对多个子公司自查发现的死淘确认流程漏洞，集团养禽业务中心统一优化了死淘确认的现场核验要求与视频留痕规定，从集团层面完善了管控规则。同时，按照15号文要求，将穿透能力评估等级纳入资源配置和经营授权的重要参考依据，穿透能力不达标的子公司在关键业务授权上予以必要限制，以此倒逼各层级加快系统联通和数据治理进程^[3]。

四、实施保障与实践启示

内控有效性评价体系的顺利落地，需建立三项核心保障机制。组织保障方面，由集团审计部牵头成立专项工作小组，销售业务部、养禽业务中心、财务中心等部门安排专人协同支持，明确各部门的职责边界，避免出现责任真空。文化保障方面，加强对子公司管理层与基层员工的内控合规培训，强化“内控创造价值”的管理理念，提升各层级人员配合自查工作的主动性，避免出现抵触情绪或隐瞒问题的情况。数据治理保障方面，按照2号文要求，建立统一的数据标准、采集和治理规则，推动跨业务、跨层级的数据传输与汇集应用，确保主数据统一、编码规则统一、数据口径统一，为穿透式监管提供坚实的数据底座^[2]。从A集团的自查实践经验来看，农牧集团开展内控有效性评价需重点把握三个核心要点：必须贴合行业特性设计评价体系，将生物资产管理、合作养殖户运营等农牧行业特色场景作为评价重点，才能真正识别高风险问题，避免评价工作流于形式；必须强化第三道防线的独立性，评价工作由审计部门独立牵头开展，不受业务部门干预，才能确保评价结果客观公正，真正发挥监督作用；必须重视整改闭环的落地执行，明确整改责任、强化跟踪验收、推动根源性优化，真正实现以评促改、以评促管的目标。同时，应深刻认识到内控评价标准已从“制度是否健全”转向“能否支撑穿透式监管”，未来需加快规则模型建设和制度数字化改造，将内控规则从“人看人记”升级为“系统自动识别、自动预警、自动派单、自动跟踪”，真正实现让风险藏不住、让问题跑不掉的目标。本研究案例主要来源于单一农牧集团的实践经验，不同发展阶段、业务结构的农牧企业可根据实际情况对评价体系进行适配调整。未来随着农牧行业数智化转型深入，可进一步探索物联网、大数据等技术在内控评价中的应用场景，提升风险识别的实时性与精准性。

参考文献

[1] 农业农村部. 农业产业化龙头企业合规经营指引 [EB/OL]. (2025-01-15). http://www.moa.gov.cn/gk/tzgg_1/tzgg/202501/t20250115_6412873.htm.
[2] 国务院国资委. 关于加强中央企业穿透式监管的指导意见（试行）（国资发监督〔2026〕2号）[Z]. 2026.
[3] 国务院国资委办公厅. 关于做好2026年中央企业内部控制体系建设与监督工作有关事项的通知（国资厅监督〔2026〕15号）[Z]. 2026.
[4] COSO. 内部控制——整合框架（农业行业适配指南补充说明）[M]. 2025.
[5] Smith D, Johnson R, Lee K. Internal control effectiveness in agricultural enterprises: A global empirical study[J]. Journal of Agricultural Economics, 2024, 75(3): 456-472.
[6] 李军, 张华. 集团化农牧企业内控评价的分层分级指标体系研究 [J]. 会计研究, 2025(4): 89-96.
[7] 王佳, 赵明, 陈磊. 农牧企业内控有效性评价指标体系构建——基于12家头部企业的实证分析 [J]. 中国农业会计, 2026(2): 28-33.
[8] 农牧行业协会. 2025年中国农牧行业内控白皮书 [R]. 2025.

新课标下构建数学师范生教学技能训练培养体系与实践

盛婷婷

内蒙古民族大学 数学科学学院, 内蒙古 通辽 028000

DOI:10.61369/ASDS.2026060004

摘 要 : 本文聚焦数学与应用数学专业师范生核心素养与实践能力的培养, 直面以往师范生培养中普遍存在的训练内容聚焦性不足、多停留在基础技能重复训练层面, 实施过程缺乏精准化设计、粗放低效且针对性不强, 培养模式固化陈旧、难以适配新时代基础教育改革需求, 以及理论教学与一线教学实践脱节、技能训练与岗位需求衔接不紧密等突出问题, 依托学校完善的师范技能实训硬件资源, 系统构建并实施了“实训赋能 – 赛教融合 – 闭环提质”三位一体的师范生培养体系。通过遵循 OBE 教学理念, 围绕“师德高尚、学识扎实、技能突出”的培养目标, 坚持“目标导向、问题导向、实践导向、产出导向”的原则, 形成了全程贯通、校地协同、循环提升的培养新路径, 有效破解了师范生教学技能训练的系统性不足与实效性不强等难题。

关 键 词 : 数学师范生; 教学技能训练

Construction and Practice of Teaching Skill Training System for Mathematics Normal Students Under the New Curriculum Standard

Sheng Tingting

College of Mathematical Sciences, Inner Mongolia Minzu University, Tongliao, Inner Mongolia 028000

Abstract : Focusing on the cultivation of core literacy and practical abilities of normal students majoring in Mathematics and Applied Mathematics under the new curriculum standard, this paper addresses the prominent problems existing in the current training of normal students. These problems include insufficient focus on training content, over-reliance on repetitive training of basic skills, lack of targeted and refined design in the implementation process, extensive and inefficient training mechanisms, rigid and outdated traditional training modes that cannot meet the demands of basic education reform in the new era, disconnection between theoretical teaching and front-line teaching practice, and poor alignment between skill training and actual job requirements. Based on the well-equipped practical training hardware resources for teacher training in the university, this paper systematically constructs and implements a three-in-one training system of "practical training empowerment, competition-teaching integration, and closed-loop quality improvement". Adhering to the OBE educational concept, centering on the training goal of "noble teacher ethics, solid professional knowledge and outstanding teaching skills", and following the principles of goal orientation, problem orientation, practice orientation and output orientation, this research constructs a new whole-process, school-local coordinated and cyclically improved training path. The research results effectively solve the practical difficulties such as insufficient systematicness and low effectiveness in the teaching skill training of mathematics normal students.

Keywords : mathematics normal students; teaching skills

引言

2021 年教育部出台《中学教育专业师范生教师职业能力标准（试行）》，将师范生能力标准化为师德践行、教学实践、综合育人、自主发展四大模块^[1]，其中教学实践能力作为连接理论知识与教学实操的核心纽带，直接决定师范生职后教学质量高低，更事关我国基础教育人才培养根基的稳固。2022 年教育部出台《义务教育数学课程标准（2022 版）》立足立德树人根本任务，确立以核心素养为导向的课程

基金项目：“新课标”背景下数学师范生教学技能培养策略研究（YB2022014）。

作者简介：盛婷婷（1984-），女，汉族，内蒙古民族大学数学科学学院教师，硕士，讲师，研究方向：数学教育。

目标体系，明确提出“三会”，即会用数学的眼光观察现实世界、会用数学的思维思考现实世界、会用数学的语言表达现实世界的育人总要求^[2]，强调四基、四能与数学核心素养深度融合，推动数学教育从知识本位向素养本位转型，对数学教师的教学设计、课堂实施、综合育人、评价反思等能力提出更高、更精准的时代要求。

本文提出的培养体系以实训赋能体系为基础支撑，以赛教融合模式为实践载体，以全程监测与闭环改进系统为质量保障，系统提升了师范生教学实践能力与综合素养。该培养体系实施以来，培养成效显著：师范生在各级教学技能竞赛中屡获佳绩，实习期间展现出过硬的专业实践能力，毕业生就业竞争力与职业发展质量持续领先。该模式成功塑造了“基础扎实、能力出众、发展可持续”的师范生培养新样态，形成了可复制、可推广的育人改革方案，为基础教育师资培养提供了有效范式。

一、目前数学师范技能训练存在的问题

在新课标落地与教师能力标准实施的双重背景下，传统数学师范技能训练模式的弊端日益凸显，主要集中在内容、模式、质量保障三个核心维度，成为制约人才培养质量提升的关键瓶颈^[3]。

（一）师范生技能训练“内容零散、实效不强”^[4]

训练内容仍停留在“三字一话”等基础技能，重形式、轻内涵。训练内容滞后于中学教学实践，导致师范生入职后“学的用不上，用的没学过”，无法快速适配一线教学。

（二）解决了培养模式“方法陈旧、动力不足”的问题

教学方法仍采用“课堂讲授+微格教学+模拟说课”的传统模式：微格教学多为“无生课堂”，缺乏真实课堂的生成性、灵活性训练。训练多以“任务式、考核式”驱动，师范生被动完成训练任务，缺乏自主探究、自主反思、自主提升的空间；导致师范生参与积极性弱、主动钻研意识不足，技能提升完全依赖外部要求，而非内在职业认同驱动。

（三）解决了质量保障“评价单一、改进乏力”的问题

训练仍以“期末一次性考核”为核心：仅通过一次试讲、说课、技能比赛判定训练效果，忽视训练过程。评价只看“最终展示”，不看“成长轨迹”，无法真实反映师范生的技能短板和进步幅度。评价主体仅为高校教师，无中学一线数学教师、学生、师范生同伴参与；高校教师评价侧重理论规范、形式完整，一线教学更看重课堂实效、学生接受度、实战能力，导致评价结果脱离教学实际，无法精准判断师范生是否适配中学教学需求。评价仅给出分数或等级，无针对性诊断、无具体改进建议、无跟踪指导：师范生知道“考核不合格”，但不知道“哪里差、怎么改”；且评价结果未与后续训练、实践、辅导联动，“评价—反馈—改进—再评价”的闭环机制缺失，技能训练只评不改、屡训屡错，质量持续无法提升。

二、解决以上教学问题的方法

为有效解决上述三个问题，经过多年的实践探索，数学专业形成了“实训赋能·赛教融合·闭环提质”的教学技能训练模式，秉承OBE教学理念，以系统化设计、协同化实施、精准化调控为核心，全方位提升数学师范生教学技能培养质量^[5]。具体解决教学问题的方法如下：

（一）构建素养导向的系统化训练体系，健全双导师协同育人培养机制

1. 对标新课标，把高中数学六大数学核心素养嵌入技能训练全程。
2. 打破传统零散训练模式，系统规划了从第一学期至第八学期的阶梯化训练项目，实现了“三习一训”（“三习”是指教育见习、教育实习、教育研习；“一训”是指教学技能训练）的有机贯通^[6]。同时施行“双导师”制，即高校教师与一线中学教师共同指导。

（二）赛教融合锤炼专业师能，职业认同激活内生发展动力

1. 通过建立分层分类的赛事机制，充分激活学生的内生学习动力。
2. 开展名师讲座、优秀校友分享、一线教师进课堂等活动强化学生的职业认同感。

（三）全景评价促进精准调控，反馈改进驱动闭环提质

1. 建立师范生个人成长电子档案，对标新课标与教师专业标准，制定可量化、可观测的评价指标。
2. 系统化收集来自校外指导教师、实习学校、用人单位等多方评价数据。

三、具体实施过程

“实训赋能—赛教融合—闭环提质”三位一体培养体系的实施，以全过程、全覆盖、协同化为原则，分模块、分阶段、分主体有序推进，确保各项举措落地生根、取得实效。

（一）实训赋能

1. 对标新课标，在以下几个方面将高中数学六大数学核心素养嵌入技能训练全程。

（1）教学设计训练：围绕大单元、主题式、问题链设计，强化数学抽象、逻辑推理、建模等素养落地。

（2）课堂技能训练：增加探究式教学、合作学习组织、数学思维引导等实战内容。

（3）补充新课标必备能力：作业设计、命题能力、错题分析、学习评价、数学文化渗透。

2. 重构“三习一训”的技能训练体系，在每一阶段，均配备高校学科教学论教师与一线中学骨干教师组成的“双导师”团队，共同参与训练项目的设计、实施与评价。依托学校师范技能

中心的标准化微格教室、板书训练室以及学院配备的专业录播教室、智慧教室等硬件设施，在双导师的协同指导下，为每个阶段的技能训练提供了强有力的情境化、专业化平台支撑与实践智慧注入。

（二）赛教融合

1. 具体构建了“院级—校级—省级—国家级”四级竞赛进阶体系，低年级侧重基础类赛事（如院级板书大赛、校级教学技能大赛、数学解题竞赛），依托师范技能中心三笔字练习教室、微格教室开展基础技能集训；高年级侧重高阶类赛事（如自治区教学技能大赛、全国师范生教学技能竞赛），利用学院录播教室进行模拟竞赛、复盘打磨，提升赛事应对能力。针对重点赛事，组建由经验丰富的指导教师构成的专项集训团队，制定个性化方案，通过案例分析、模拟竞赛、往届选手经验分享等多元化活动，进行精准、高效的赛前强化，将参赛压力切实转化为持续精进的专业能力，形成了“以赛促学、以赛促练、以赛促创”的良性循环^[7]。

2. 通过定期邀请基础教育数学名师开展专题讲座，分享新课标教学理念与职业成长历程；邀请优秀校友返校交流从教经验、职场发展心得；常态化安排一线中学数学教师走进课堂，现身示范授课、分享真实教学案例，帮助师范生树立教育理想，增强从教使命感与职业认同感。

（三）闭环提质

1. 利用师范技能中心平台和实习实训平台对学生“三习一训”的全过程进行自动化采集与量化分析。师范技能中心平台全程记录学生的在各项技能训练时的照片、视频及指导教师评价，且每次评价必须给出诊断意见+改进建议+改进任务，拒绝“凭感觉打分”，构建“评价—反馈—改进—再评价”闭环质量机制，对薄弱学生实行一对一帮扶、重复训练，确保人人过关。实习实训平台记录学生竞赛表现、实习全景等多维度信息。依托两大平台数据联动，自动生成个人专业成长画像，精准研判技能短板，实现训练过程可追溯、能力水平可量化、成长路径可追踪，为个性化培养与精准化育人提供数据支撑。

2. 定期召开人才培养专题研讨会，收集多方的评价与需求，经线下专题研讨与深度分析后，将其中具有普遍性和典型性的问题反馈，精准对应至相关课程体系与训练环节，为教学方案的持续优化、培养目标的动态调整提供客观依据，从而形成螺旋式质量提升闭环，推动人才培养质量持续优化。

四、创新之处与推广应用

（一）创新之处

本培养体系突破传统数学师范教育的局限，在模式、机制、评价三个维度实现创新突破，形成独具特色的育人范式。

1. 模式创新：首创“能力进阶、双师协同”的立体化实训赋能体系

针对技能训练零散化问题，本成果突破传统实训模式，创新构建了“四学年阶梯递进、三习一训全程贯通”的系统化训练体系。该体系的核心创新在于：一是时序上的系统性重构，将训练项

目精准嵌入每学期课程，实现理论学习与实践训练的同步推进；二是空间上的双师协同，依托学校 UGS 合作模式，组建以“高校理论导师+中学实践导师”双导师共同体，贯穿训练全程，在师范技能中心、智慧录播教室等平台上共同指导学生完成从教学设计到课堂反思的全流程实训，实现了教育理论与一线教学智慧的深度融合与即时转化。

2. 机制创新：构建“以赛导学、分层激发”的赛教融合动力转化机制

为破解培养模式陈旧、学生动力不足的难题，本成果创建了“竞赛牵引、教创互嵌”的赛教融合新机制。其创新性体现在：一是构建了“院—校—省—国”四级竞赛进阶体系，将竞赛标准与要求反向导入日常教学与训练，实现“赛”与“教”在内容与目标上的精准融合；二是建立了“基础普及—高阶挑战”的分层参与与专项辅导机制，通过个性化备赛方案与多元化强化活动，将外在的竞赛压力有效转化为学生持续精进、自主创新的内在驱动力，形成了“人人参与、层层选拔、赛学相长”的良性育人生态^[8]。

3. 评价创新：建立“数据驱动、循证改进”的全景化质量闭环调控机制

针对质量保障中评价单一、改进乏力的问题，本成果革新了传统评价方式，建立了基于全过程数据采集与循证分析的质量持续改进机制。其创新突破在于：一是评价维度的全景化，通过师范生个人成长电子档案，系统整合课程学习、技能实训、竞赛表现、实习评价等多源数据，实现成长轨迹的可视化与量化分析^[9]；二是改进机制的闭环化与循证化，通过线下专题研讨深度解析来自实习学校、用人单位等的外部评价数据，并将共性结论精准反馈至课程与训练环节的调整中，形成了“监测—诊断—反馈—优化”的螺旋式质量提升闭环，实现了人才培养质量的动态、精准、持续改进^[10]。

（二）推广应用效果

通过“实训赋能—赛教融合—闭环提质”三位一体培养体系的系统实施，在提升师范生培养质量与扩大改革影响力方面取得了显著成效。

1. 师范生培养质量显著提升，关键指标成效突出

训练模式的实施直接带来了师范生教学技能与实践能力的全面提高，相关量化与质性评价指标均表现优异：竞赛成绩屡创新高、实习表现广受好评、就业竞争力持续领先，形成了良好的社会声誉闭环。

2. 改革模式产生辐射效应，校际交流合作深入开展

基于共同的师范教育改革目标，与多所省属师范类高校建立了稳定的互访交流机制。双方已成功互派多批师范生，举办联合教学技能展示与交流研讨会。通过板书展示、微格教学互评等活动，两校学生同台切磋、互相学习，不仅拓宽了学生的视野，也促进了两校在培养模式、评价标准上的深度对话与共同提升，形成了跨区域协同育人的良好范例。

五、需要进一步深化和完善之处

尽管本培养体系取得了显著成效，但面对教育数字化转型、

人工智能赋能教育、新课标持续深化的新形势，仍存在一些需要优化完善的薄弱环节，未来将从以下三方面持续深化改革。

（一）实训赋能的课程适配性需升级

训练课程没有设置生成式人工智能内容，缺乏 AI 与数学教学融合的专项模块，需进一步开发适应数学教学技能训练的生成式人工智能工具助力师范生适配数字化教学场景。

（二）赛教融合的广度与深度待拓展

现有竞赛体系集中于核心教学技能，对跨学科融合、信息化高阶应用（如 AI 辅助教学、智慧课堂设计）等新型赛事覆盖不足。同时，部分低年级学生参与度较低，需加大动员力度，增加赛事的趣味性 with 参与感。

（三）闭环提质的数字化与精准度需提升

个人成长电子档案的数据采集以人工记录、平台上传为主，自动化采集覆盖率不足。数据多围绕技能表现，对师范生师德素养、沟通能力等软素质的量化评价缺失。需丰富评价指标维度，实现更精准的个性化诊断。指导教师数量有限，还需扩大指导教师团队，以满足小班化精准化指导。

六、结语

新时代基础教育改革对数学师范生培养提出了更高要求，构建适配新课标、对接教师能力标准、服务一线教学的师范技能培养体系，是地方高校数学师范教育改革的核心任务。本研究构建的“实训赋能－赛教融合－闭环提质”三位一体培养体系，以问题为导向、以实践为核心、以质量为目标，通过系统化训练、协同化育人、闭环化保障，有效破解了传统数学师范技能训练的痛点难题，全面提升了数学师范生的教学实践能力与综合素养。

未来，我们将持续立足基础教育需求，紧跟教育改革趋势，不断优化完善培养体系，深化教育数字化转型，强化校地协同育人，致力于培养更多师德高尚、学识扎实、技能突出、适应新时代义务教育数学教育发展的高素质专业化教师。

参考文献

[1] 中华人民共和国教育部. 教师厅 [2021]2 号《中学教育专业师范生教师职业能力标准（试行）》[S]. 教育部办公厅 (http://www.moe.gov.cn/srcsite/A10/s6991/202104/t20210412_525943.html).

[2] 中华人民共和国教育部.《义务教育数学课程标准（2022 版）》[S]. 北京师范大学出版社.

[3] 王思义. 数学师范生教学技能调查研究 [J]. 现代交际, 2021 (23):36-38.

[4] 徐辉, 孙平, 王洪兰. 师范生教学技能训练的现实要求、实施困境及改进策略 [J]. 郑州师范教育, 2025, 14 (06):55-58.

[5] 张秋平, 姚素梅, 张艺, 李晓严. 基于 OBE 理念的师范生教学技能训练模式的构建与实施 [J]. 教育信息化论坛. 2024 (12):84-86.

[6] 王光明, 冯虹, 康玥媛. 新理念数学教学技能训练 [M]. 北京大学出版社. 2021.08.

[7] 熊显萍, 黄明春. 民族师范院校师范类专业教学体系构建与实践——以兴义民族师范学院数学专业为例 [J]. 兴义民族师范学院学报, 2022(06):52-57.

[8] 黄德兴, 李元欢, 杨海斌. 师范生教学技能“责任田”训练模式的实践探索 [J]. 教育教学论坛. 2025 (37):115-118.

[9] 黄利华, 刘忠蓉. 专业认证下高校教法课程学习评价体系改革与实践——以小学数学课堂教学技能训练课程为例 [J]. 科教导刊. 2025 (03):54-56.

[10] 彭晓琼. 师范生教学技能训练与考核 [J]. 玉林师范学院学报. 2023, 44 (05):130-133.

教育数学思想下中学极限概念的教学衔接研究

胡志安^{1,2}, 徐洪焱¹, 李弘³

1. 宿迁学院数理学院, 江苏 宿迁 223800

2. 青海师范大学数学与统计学院, 青海 西宁 810016

3. 赣南师范大学科研处, 江西 赣州 341000

DOI:10.61369/ASDS.2026060005

摘 要 : 极限概念是连接初等数学与高等数学的关键纽带, 其教学衔接质量直接影响学生数学思维的深化与知识体系的进阶。本研究针对当前初高中与大学微积分课程中极限教学存在的断层问题, 结合对城乡普通高中学生学习现状的调研数据, 指出教学改进的迫切性。基于张景中院士提出的教育数学思想, 研究尝试从教学理念与内容呈现两方面进行重构: 一方面, 以刘徽“割圆术”、“一尺之捶”等经典极限思想案例为认知起点, 降低形式化理解的难度; 另一方面, 依托自然数列、不等式等学生熟悉的知识载体, 设计符合学生认知节奏的教学路径。在此基础上, 论文进一步开发出可供课堂使用的具体教学案例, 旨在为强化中学阶段的极限概念教学提供具有操作性的参考, 促进知识体系从直观到严谨的自然过渡。

关 键 词 : 教育数学; 极限; 衔接教学; 中学数学

Grounded in Educational Mathematical Thinking: Research on the Cognitive Progression and Teaching Integration of the Concept of Limits in Middle School

Hu Zhi'an^{1,2}, Xu Hongyan¹, Li Hong³

1. School of Mathematics and Physics, Suqian University, Suqian, Jiangsu 223800

2. School of Mathematics and Statistics, Qinghai Normal University, Xining, Qinghai 810016

3. Office of Research, Gannan Normal University, Ganzhou, Jiangxi 341000

Abstract : The concept of limits is a key link between elementary mathematics and advanced mathematics, and the quality of its teaching connection directly affects the deepening of students' mathematical thinking and the advancement of their knowledge system. This study aims to address the gap in extreme teaching in calculus courses between middle and high schools, as well as universities. Based on research data on the learning status of urban and rural high school students, it points out the urgency of teaching improvement. Based on the educational mathematics ideas proposed by Academician Zhang Jingzhong, this study attempts to reconstruct teaching concepts and content presentation from two aspects: on the one hand, using typical extreme thinking cases such as Liu Hui's "cutting circle technique" and "one foot hammer" as cognitive starting points to reduce the difficulty of formal understanding; On the other hand, relying on familiar knowledge carriers such as natural sequences and inequalities, design teaching paths that conform to students' cognitive rhythms. On this basis, the paper further develops specific teaching cases that can be used in the classroom, aiming to provide operational references for strengthening the teaching of extreme concepts in middle school, and promote the natural transition of the knowledge system from intuitive to rigorous.

Keywords : educational mathematics; limit; bridging teaching; middle school mathematics

引言

极限思想是现代数学学科发展的重要理论基础, 贯穿中学数学和高等数学的知识体系, 它不仅是导数、微分、积分等核心知识的必

基金项目: 研究生创新基金项目(2025LPCX26); 中国高等教育学会高等教育科学研究规划课题重点项目(24SX0206)。

作者简介:

胡志安, 安徽池州人, 硕士, 研究方向: 中学数学教育;

徐洪焱, 江西乐平人, 教授, 研究方向: 复分析理论及其应用;

李弘, 江西赣州人, 副教授, 研究方向: 心理统计与测量, 教育心理学。

要前置知识，学习意义重大^[1]。在数学史上，极限思想的应用同样丰富。我国魏晋数学家刘徽就提出过“割圆术”，使用不断增加圆内接正多边形边数的方法以正多边形面积持续逼近圆面积从而较为准确地推出圆周率的近似值。在古代文学上，《庄子》中有“一尺之棰，日取其半，万世不竭”这种用形象通俗的语言诠释极限思想的例子。在现代中学数学教学中极限不仅是必修知识的重要组成部分，还是学生衔接后续高等数学学习的关键桥梁^[2]。

然而当前中学数学教材中对极限概念的教学仍然存在一定程度上的概念表述模糊化、内容碎片化和思维引导片面化的问题^[3]。

基于当前极限概念教学中的一些难点和不足，我国中科院院士张景中教授在其建构的教育数学理论框架下，创造性地提出以学生更加熟悉的自然数列、不等式等知识点为核心的极限新定义，绕开以往抽象难懂的逻辑表述从而符合学生常规认知水平，不失为重构中学极限概念教学的新思路^[4]。

一、极限概念在初高等数学中教学的衔接现状

（一）初高等数学极限教学衔接的不足

当前，中学数学教材中对极限概念的教学存在显著的模糊和一笔带过的现象，仿佛极限概念只是一个为研究函数导数和函数性质的附庸。首先具体表现为概念表述模糊化，往往以“无限趋近”这类直观描述代替严谨定义，逻辑性不强，学生认知常常停留在字面；其次是内容碎片化，与数学其他知识板块联系不够紧密，人教版选择性必修二里直接提出极限与导数概念割裂感强，且极限的“ $\varepsilon - \delta$ ”语言表述艰难晦涩，显著区别于以往中学数学语言，知识脉络不够清晰；最后是思维引导的片面化，忽视学生无限与逼近思维逻辑的培养，侧重公式记忆与习题练习的倾向严重，不注重对极限概念本质的理解^[5]。

从实际学习场景来看，中学极限教学是直接为高考服务的，表现为聚焦公式记忆和函数习题练习，忽视与高等数学关联知识的铺垫与衔接^[6]。例如中学导数教学时仅靠极限概念的直观定义推导出各类函数的导数公式，对极限与导数的内部逻辑关系介绍不够深入，极限的工具属性严重，而这一重使用轻理解的倾向与高等数学培养严谨逻辑思维能力的取向严重脱钩，无疑加剧了极限概念的衔接矛盾。

（二）城乡普高学生极限概念的学习现状调查

为进一步验证中学极限教学改进的可行性与必要性，本研究以江苏省宿迁市城乡各一所普通高中学生为调查对象，共发放问卷540份，有效回收问卷509份（城区普高300份、乡镇普高240份），具体调查结果如下表：

表：中学生极限概念学习现状调查汇总表

调查维度	具体调查内容	整体情况（占比/数值）	城区普高情况	乡镇普高情况
基础调查信息	发放问卷数量	540份	300份	240份
	有效回收问卷数量	509份	291份	218份
概念理解	能准确阐述极限本质	23.1%	29.1%	17.2%
	仅能通过“无线趋近”描述	68.5%	61.8%	75.3%
	完全无法理解极限含义	8.4%	7.5%	9.1%
衔接认知	不清楚极限与微积分的联系	79.3%	75.6%	83.0%
	了解极限是大学微积分的基础	12.5%	15.3%	9.8%
学习困惑	认为极限概念抽象难懂	65.7%	61.0%	70.3%
	无法掌握极限与相关知识的应用	58.2%	54.0%	62.3%

问卷调查结果显示：对于概念理解层面，仅23.1%的学生能准确阐述极限本质，68.5%的学生仅能通过“无限趋近”描述极限，8.4%的学生完全无法理解极限含义，乡镇普高学生对极限概念准确认知的占比（17.2%）低于城区普高（29.1%）；对于衔接认知层面，79.3%的学生不清楚中学极限与高等数学微积分的关联，仅12.5%的学生了解中学极限是大学微积分的基础；对于学习困惑层面，65.7%的学生认为极限概念抽象难懂，58.2%的学生表示无法掌握极限与关联知识的融合应用，整体来看乡镇普高学生困惑占比略高于城区普高。

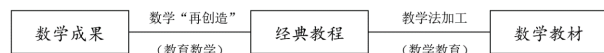
由调查结果可以发现当前高中生极限学习存在概念认知模糊、衔接认知缺乏、知识联系和应用能力不足等问题；且无论是城区学生还是乡镇学生都受到现行极限教学不足的影响，中学极

限教学的改进具有一定的必要性。同时学生对极限知识的学习需求是明确的，且需要更加科学的教学引导，因此在进行中学极限教学时结合张景中院士的教育数学思想优化教学方式具有现实可行性^[7]。

二、教育数学思想指导下的中学极限概念教学策略

（一）教育数学思想

张景中院士提出的教育数学思想核心是“改造数学知识结构，使其更适合教学与学习”，主张摒弃数学知识里冗余、抽象的表述，提炼核心逻辑与本质属性，使数学知识更加符合学生的认识水平从而达到重新建构知识体系的目的^[8]。



图：教育数学示意图

在极限概念的教学，教育数学思想打破“直观描述”的教学逻辑，以贴近中学生认知、兼顾严谨性与易懂性为目标，重构极限概念的表述，一方面可以弥补现行教学过程中的严谨性不足，另一方面也避免了使用高等数学教材中魏尔斯特拉斯提出的“ $\varepsilon - \delta$ ”或“ $\varepsilon - N$ ”的高度抽象极限语言^[9]。

（二）教育数学思想下的极限概念

张景中院士提出的中学极限概念教学，以自然数列、不等式这些学生头脑里已有的知识为核心联结，简化概念推导逻辑，降低认知难度。具体教学方式如下：

定义1（无穷大数列的定义）设 a_n 是一数列，如果有无界不减数列 D_n ，使得 $|a_n| > D_n$ 对所有的 n 都成立，则称 a_n 为无穷大数列。

定义2（无穷小数列的定义）设 a_n 是一数列，如果有无界不减数列 D_n ，使得 $|a_n| < D_n$ 对所有的 n 都成立，则称 a_n 为无穷小数列。

于是极限概念就瓜熟蒂落了；

定义3（极限的定义）设 a_n 是一数列，如果有实数 A ，使得数列 $\{a_n - A\}$ 是无穷小数列，则称数列 a_n 以 A 为极限。

像这样在学生已学过的概念基础上平易自然地引入极限概念，没有逻辑上的跳跃，学生不但容易接受新知识，还巩固了旧知识^[10]。

（三）教育数学思想下的极限概念教学的核心原则与策略

通过查阅大量文献和资料，可以大致将教育数学思想指导下的极限概念教学归结为以下三大核心原则^[11]：

1. 认知适配原则。贴合高中生抽象思维发展阶段，以已学知识为基础重构极限概念，避免超认知水平的抽象表述，保障学生能逐步理解概念本质；

2. 严谨性与易懂性平衡原则。既保留极限概念的逻辑严谨性，又通过简化表述、具象载体降低学习难度，杜绝“重直观轻严谨”或“重严谨轻易懂”的片面教学；

3. 知识衔接原则。教学中强化极限与中学关联知识、高等数学前置知识的关联，帮助学生构建完整的知识脉络，为后续学习铺垫基础。

同时可以在极限概念教学中融入三大核心教学策略^[12]：

1. 具象化导入策略。可以结合刘徽的割圆术、“一尺之棰”等经典案例，将抽象极限概念转化为具象场景，帮助学生快速建立初步认知；

2. 循序渐进教学策略。按“自然数列感知→不等式定义推导→概念应用深化”的步骤开展教学，逐步提升知识难度，符合学生认知发展规律；

3. 互动探究教学策略。通过设计数列分析、不等式应用等探究任务，引导学生自主分析、合作讨论，主动构建极限概念认知，培养逻辑推理能力，用科学的原则与实用的策略帮助学生掌

握极限的概念。

三、教育数学思想下的极限概念的教学设计

下面介绍教育数学思想下的极限概念的教学设计。

（一）教学目标

1. 以自然数列和不等式为知识联结点，构建极限概念的直观理解与严谨表述，掌握极限的核心内涵。

2. 经历极限概念从具体到抽象的形成过程，体会教育数学“从初等数学自然延伸至高等数学”的思想，提升数学抽象、逻辑推理素养。

（二）教学过程

教学活动1：复习联结，唤醒旧知

问题1. 同学们，我们已经学习了自然数列（1,2,3, …… n ），请大家思考，自然数列的项有什么变化趋势？再回忆一下不等式的性质，比如“若 $a < b, c > 0$ ，则 $ac < bc$ ”，这些知识能为我们今天的学习带来什么启发吗？

师生活动：教师引导学生回顾自然数列的无限性（项可以无限增大），以及不等式的基本运算性质。学生分组讨论后分享观点，教师总结：自然数列的“无限”和不等式的“精确刻画”，是理解“极限”重要工具。

设计意图：从学生已熟悉的自然数列、不等式知识切入，契合中学阶段的认知基础，降低概念学习的陌生感。

教学活动2：情境探究，初步感知

问题2. 假设小胡在直线上跑步，第一秒跑1米，第二秒跑 $\frac{1}{2}$ 米，第三秒跑 $\frac{1}{4}$ 米，第四秒跑 $\frac{1}{8}$ 米……按照这个规律，小胡跑的总路程会超过2米吗？你能结合自然数列和不等式，说明其中的道理吗？

师生活动：教师呈现问题后，学生先独立思考，再进行小组交流。有学生可能会列出总路程的表达式 $S_n = 1 + \frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^{n-1}}$ ，教师引导学生利用数列求和公式得出 $S_n = 2 - \frac{1}{2^{n-1}}$ 。接着教师追问：“对于任意给定的正数 c 能否找到一个自然数 N ，使得当 $n > N$ 时， $|S_n - 2| < c$ 恒成立？”学生通过赋值计算（如 $c = 0.1$ 时，解 $\frac{1}{2^{n-1}} < 0.1$ ，得 $n > 4$ ），体会“通过自然数列和不等式，精确刻画总路程与2的接近程度”。

设计意图：以“跑步路程”的生活情境为载体，将抽象的极限概念具象化，符合中学生的具象思维特点，体现“初等知识解决高等问题雏形”的衔接逻辑。

教学活动3：抽象概括，形成概念

问题3. 我们把刚才的例子推广到一般情况：对于数列 $\{a_n\}$ 和常数 A ，若对于任意给定的正数 c 都存在自然数 N ，使得当 $n > N$ 时， $|a_n - A| < c$ 成立，就称数列 $\{a_n\}$ 的极限是 A 。请大家结合自然数列（ N 的作用）和不等式（ $|a_n - A| < c$ 的刻画），说说这个定义的核心意思。

设计意图：先抽象出形式化定义，再引导学生用通俗语言解读，实现“严谨定义”与“易懂理解”的平衡；让学生经历“具象→抽象→解读”的过程，适配中学阶段从具象思维向抽象思维

过渡的认知特点。

教学活动4：应用拓展，联系生活

问题4. 生活中也有极限的影子，比如“圆的内接正多边形面积无限接近圆的面积”。请大家尝试用极限定义的语言，描述这个过程（提示：设正 n 边形面积为 S_n ，圆面积为 S ）。

师生活动：将班级学生分组并开始尝试进行数学语言表述，最终引导出“对于任意 $c>0$ ，存在自然数 N ，当 $n>N$ 时， $|S_n - S|<c$ ”。教师总结这体现了张景中先生强调的“数学概念源于生活又高于生活，通过初等知识的联结点（自然数列、不等式），能严谨解释生活中具体的“无限过程”。

设计意图：从生活中的“圆面积逼近”现象切入，让学生感受极限概念的现实意义，契合其“从生活到数学”的认知路径；再次强化“自然数列、不等式”作为初等数学与高等数学联结点的作用，体现知识的连贯性。

四、教学反思

基于教育数学思想的极限概念教学设计，可以将相对抽象的极限概念回归到学生已掌握的自然数列、不等式等初等数学知识，能够有效实现学生认知水平和能力实践的平稳过渡^[13]。相较于传统教学中直观描述有余而严谨性不足从而导致学生在学习中认知超载的问题，本研究设计通过经典案例具象导入、旧知联结自然过渡、探究任务驱动建构，既在一定程度上保留了数学概念的逻辑严谨性，又贴合高中生由具体到抽象、由静态到动态的思维发展规律。

在教学设计过程中，学生能够借助生活情境与几何直观来理解极限概念“无限逼近”的内涵，用不等式语言刻画极限过程，初步建立起极限的严谨认知。但反思整个研究设计过程，仍存在可优化空间：一是对不同层次学生（尤其是城乡基础差异学生）的个性化引导仍可加强；二是对于动态可视化工具的融入深度不足，对极限的“动态过程”的直观呈现可进一步提升；三是课堂探究时间分配需更灵活，以保证更多学生完整参与概念建构的全过程。

总体而言，教育数学指导下的教学设计将极限这一抽象概念

“返璞归真”为学生熟悉的自然数列增长和不等式距离刻画，既体现了数学学科的严谨性，又契合中学阶段的认知水平，充分体现了张景中院士教育数学思想中“让数学平易近人、让概念自然生长”的核心理念^[14]，同时发展了学生的数学抽象、逻辑推理等核心素养。

五、研究意义

从教学实践来看，该研究针对初高等数学极限教学断层问题，结合城乡学生调研数据，精准定位概念模糊、衔接缺失等痛点，为中学极限教学提供可落地的优化方案，通过教育数学思想重构极限概念呈现方式，降低严谨定义的认知门槛，有助于提升课堂教学效率，增强学生数学学习信心，为导数、微分、积分等后续内容奠定稳固的概念基础与思维基础。

从学科发展来看，研究以极限概念为切入点，丰富了教育数学思想在中学数学核心概念教学中的应用案例，为初等数学与高等数学衔接教学提供新思路。它建构了一个层层递进的概念教学体系，对推动中学数学核心概念教学的科学化、系统化具有一定参考价值。

从基础教育人才培养来看，极限是学生从“有限数学”走向“无限数学”的重要节点，是培养数学抽象、逻辑推理、直观想象等核心素养的重要载体。优化极限概念的教学，能够帮助学生建立无限与逼近的辩证思维，培养理性精神与数学素养，为理工科、经济金融类专业人才培养筑牢数学根基。

六、展望

当下，国内外 AI 技术的快速发展，自然也引发了人们的新的深层思考，探索 AI 技术与该教学方案的融合，如开发智能题库，根据学生对数列、不等式的掌握情况，个性化推送极限概念练习题将是未来研究方向之一；同时将该思想延伸至函数、几何等其他数学知识点，形成更系统的教育数学教学体系也是未来研究的一个可能方向。

参考文献

- [1] 付夕联，张玉峰. 极限概念教学的系统分析 [J]. 数学教育学报, 2013, 22(01): 83-88.
- [2] 周晨. 初等数学与高等数学衔接问题的研究 [D]. 南京师范大学, 2018.
- [3] 杨欢. 基于“中大学”衔接的高中导数教学策略研究 [D]. 西南大学, 2024.
- [4] 张晓东. 教育数学的理解与应用——访张景中院士 [J]. 高中数学教与学, 2022, (16): 1-4+28.
- [5] 许艳红. 高中生对极限概念理解的研究 [D]. 东北师范大学, 2015.
- [6] 王思义. 高中数学与高等数学在极限教学方面的衔接 [J]. 考试周刊, 2009, (32): 75.
- [7] 易斌，郝圆. 张景中院士教育数学思想与微课融合的应用探索 [J]. 数学通讯, 2024, (13): 10-14.
- [8] 张景中. 什么是“教育数学” [J]. 高等数学研究, 2004(6).
- [9] 陈怡菁. 基于数学核心素养的概念课教学设计——以“数列的极限”为例 [J]. 上海中学数学, 2021, (11): 34-38.
- [10] 赵怀广，杨晓琴. 张景中数学实验教学中培养学生高阶思维实践研究 [J]. 智力, 2024, (02): 45-48.
- [11] 范琪. 数学家张景中：用“教育数学”为“数学教育”做减法 [J]. 教育家, 2023, (16): 23-26.
- [12] 刘若晴. 基于张景中教育数学思想的初中数学新授知识类微课优化策略及应用研究 [D]. 广西师范大学, 2022.
- [13] 胡晓，肖继红，李世伦. 非系统极限定义的实验 [J]. 江西科学, 2018, 36(04): 582-587.
- [14] 张景中，余孟孟. 让数学变容易：张景中院士谈人生与教育 [J]. 新课程评论, 2025, (10): 7-14.

人工智能与线性代数课程深度融合的教学改革研究

——基于实证数据的成效验证

吴嘉润, 赵娜娜*, 苑孟理想, 唐子航, 刘新红

北京石油化工学院, 北京 102617

DOI:10.61369/ASDS.2026060006

摘 要 : 为探索人工智能技术与线性代数课程深度融合的有效路径, 本研究以某高校2024级511名学生为研究对象, 其中实验班101人实施AI融合教学干预(包含可视化工具辅助、AI智能答疑与自适应练习), 其余学生构成全级对照基准。通过期末统一考试成绩, 运用独立样本t检验、Mann-WhitneyU检验、Kolmogorov-Smirnov双样本检验及卡方检验等多元统计方法验证教学成效。结果表明: 实验班平均分(78.71 ± 15.18)显著高于全级对照组(70.94 ± 18.62), 差异具有极显著统计学意义($t=3.939, p<0.001$, Cohen's $d=0.458$); 实验班优秀率(40.6%)较对照组(16.4%)提升24.2个百分点, 不及格率由27.2%降至16.8%。各题型成绩整体优于对照组, 提升最明显的是综合应用类题目。研究表明, AI辅助教学能够在一定程度上改善线性代数课程的学习成效, 也为后续教学改革提供了实证依据。

关 键 词 : 人工智能; 线性代数; 教学改革; 实证研究; 效应量分析

Research on Teaching Reform of Deep Integration of Artificial Intelligence and Linear Algebra Curriculum—Effectiveness Verification Based on Empirical Data

Wu Jiarun, Zhao Nana*, Yuan Menglixiang, Tang Zihang, Liu Xinhong

Beijing Institute of Petrochemical Technology, Beijing 102617

Abstract : To explore effective pathways for the deep integration of artificial intelligence (AI) technology with the linear algebra curriculum, this study enrolled 511 students from the 2024 cohort of a Chinese university. Among them, 101 students in the experimental class received AI-enhanced instruction—including visualization tools, intelligent Q&A systems, and adaptive practice platforms—while the remaining students served as the control group. Final examination scores were analyzed using independent-samples t-test, Mann-Whitney U test, two-sample Kolmogorov-Smirnov test, and chi-square test. Results show that the experimental class achieved a significantly higher mean score (78.71 ± 15.18) compared with the control group (70.94 ± 18.62) ($t=3.939, p<0.001$, Cohen's $d=0.458$). The excellent rate (≥ 90 points) of the experimental class (40.6%) exceeded that of the control group (16.4%) by 24.2 percentage points, while the failure rate dropped from 27.2% to 16.8%. Performance was better across all question types, especially in comprehensive application problems. These findings suggest that AI-assisted teaching can improve learning outcomes in linear algebra and offer a practical reference for teaching reform.

Keywords : artificial intelligence; linear algebra; teaching reform; empirical research; effect size analysis

引言

线性代数作为理工科专业的核心公共基础课程, 以其高度抽象性与强逻辑依赖性著称, 历来是高等数学教学改革的难点与焦点。传统教学模式以板书演算为主, 难以有效帮助学生建立矩阵变换、特征值等抽象概念的几何直觉, 学生普遍反映知识碎片化严重、“知其然而不知其所以然”。

随着人工智能技术进入教育场景, 课程教学的组织方式也在发生变化。教育部相关政策多次强调, 要推进AI与教学深度融合, 探

基金项目: 国家级大学生创新创业训练计划(2026J00115); 北京市高等教育学会课题(MS2025130); 北京市大学生创新创业训练计划(2026J00091); 北京市教委科研计划一般项目(KM202410017004)。

作者简介:

吴嘉润, 本科生, 研究方向: 数据挖掘与个性化推荐算法;

苑孟理想, 本科生, 研究方向: 人工智能与智能数据分析;

唐子航, 本科生, 研究方向: 数理统计与数据建模;

刘新红, 副教授, 博士, 研究方向: 应用数学。

通讯作者: 赵娜娜, 讲师, 博士, 研究方向: 概率论与数理统计。

素更加注重学生体验和学习效果的智慧教育模式^[1]。从已有研究看，知识图谱、自适应平台等工具已经在高等数学教学中出现了不少应用^[2-4]，但落到线性代数课程上，系统性的实证研究仍然不多，尤其是同时使用对照组设计、多种统计检验和效应量分析的研究更少。

基于这一背景，本研究以某高校2024级511名学生为样本，构建准实验研究框架，在统一考试评价体系下比较实验班（n=101）与全级对照基准的成绩表现，并结合参数与非参数方法检验差异是否稳定。我们希望通过真实教学数据，尽量回答一个更具体的问题：AI融合教学到底能否在这门课程中带来可观察、可验证的改进。

一、AI 融合教学模式设计

（一）设计理念

本研究把 AI 定位为教学辅助工具，而不是替代教师讲授的主体。它的作用不是直接给出结论，而是通过结构化呈现、即时反馈和个性化练习，帮助学生降低理解门槛、形成直观认识，并逐步提高知识迁移能力。这个思路主要借鉴了建构主义学习理论和认知负荷理论：前者强调学生要在有意义的情境中主动建构知识，后者提醒教学设计要尽量减少无效信息对工作记忆的占用^[5、6]。

据此，本研究将教学目标概括为三点：第一，帮助学生把矩阵运算、行列式、向量空间、线性变换、特征值与特征向量等核心概念学得更扎实；第二，提高综合应用题的解题质量，尤其是跨知识点的迁移能力；第三，根据学习过程中的反馈及时调整教学安排，让课堂更有针对性。

（二）AI 融入教学方式

本研究将 AI 嵌入教学过程，主要分为三个层次，彼此之间是递进关系。

第一层是可视化辅助理解。借助可视化工具，把矩阵变换、特征向量方向等抽象内容转化成动态几何演示。比如在讲矩阵乘法时，直接展示它对二维向量的作用效果，让学生把代数运算和几何变化联系起来。这样的呈现方式通常更容易帮助学生建立直观印象，也能缓解抽象概念带来的理解压力^[7]。

第二层是 AI 答疑和个性化练习。课后如果学生在某一步推导上卡住，可以借助 AI 工具查看分步解释，而不是只拿到一个最终答案。与此同时，系统会根据学生过往答题情况推送不同难度和不同侧重点的练习题，重点补足薄弱环节。

第三层是基于数据的教学调整。教师定期查看平台汇总的作业和测验结果，了解哪些知识点是全班普遍薄弱的，再据此调整课堂讲解顺序和训练重点。

（三）教学实施步骤

教学实施分为课前、课中、课后三个环节，具体如下：

课前，教师在每章开始前推送少量基础测试题，用来了解学生对核心知识点的掌握情况。AI 自动生成班级学情报告，教师据此判断本节课应该重点讲哪些内容；学生则可以结合自己的报告提前预习，把时间更多放在薄弱点上。

课中，教师结合动态可视化界面讲解知识之间的联系，并根据 AI 统计出来的高频错误点进行重点突破。课堂练习后，系统会即时反馈学生对各知识点的掌握情况，教师可以据此继续调整节奏。

课后，AI 根据课堂测验数据为每位学生生成差异化补救方案，推送相关微课和同类习题。与此同时，教学中适当加入数学史和跨专业应用案例，如计算机图形学、数据降维等，有助于提高学生兴趣，也能让他们看到课程内容的实际用途。

二、研究设计与方法

（一）研究对象与数据

本研究采用准实验设计，以某高校一个年级的理工科学生为对象。实验班共 101 人，接受为期一学期的 AI 融合教学干预；全级 511 人（包含实验班）作为参照基准，其中非实验班学生采用传统教学方式。两组学生使用相同教材和教学大纲，参加同一期末统一考试，评分标准一致，因此数据具有较好的可比性。

期末考试满分 100 分，题型结构如下：选择题 30 分（10 题 × 3 分），填空题 20 分（10 题 × 2 分），大题 50 分（第 3 至第 8 大题，前 4 题满分 10 分，后 2 题满分 5 分）。

（二）研究假设

本研究提出以下三项假设：

1. 接受 AI 融合教学的实验班期末总分显著高于全级对照基准。
2. 实验班在选择、填空和大题三个题型上的得分率均优于对照组，且综合应用题的提升最明显。
3. 教学干预的综合效应量达到中等及以上水平（Cohen's $d \geq 0.3$ ）。

（三）统计分析方法

为尽量提高结论的稳健性，本文从参数检验和非参数检验两个角度同时展开分析。

- ① 独立样本 t 检验——检验两组总分均值差异，适用于样本量较大、近似正态的情形；
- ② Mann-Whitney U 检验——非参数稳健性验证，不依赖正态分布假设，应对数据偏态情况；
- ③ Kolmogorov-Smirnov 双样本检验——检验两组成绩累积分布函数的整体形态差异；
- ④ χ^2 检验——比较优秀率、不及格率等离散指标的组间差异；
- ⑤ Cohen's d 效应量——量化干预的实质意义（0.2= 小效应，0.5= 中效应，0.8= 大效应）^[8]；
- ⑥ 各题型分组 t 检验 + 森林图——细粒度分析各题型干预效果及置信区间。

三、实证分析结果

（一）总体成绩比较

实验班平均分为78.71分（SD=15.18），全级对照组平均分为70.94分（SD=18.62），组间差值为7.77分。独立样本t检验结果显示，差异具有极显著统计学意义（ $t=3.939$ ， $df=610$ ， $p<0.001$ ）。由于Shapiro-Wilk检验显示两组数据均存在一定程度的非正态性，同时采用Mann-WhitneyU检验进行稳健性验证，结论一致（ $U=31815$ ， $p<0.001$ ）。效应量Cohen's $d=0.458$ ，达到中等效应水平，表明教学干预具有实质性的成效，而非仅具统计显著性（见表1）。

表1：描述统计指标对比

统计指标	全级对照组 (n=511)	实验班 (n=101)	差值
均值 ± 标准差	70.94 ± 18.62	78.71 ± 15.18	+7.77
中位数 (IQR)	74 (59 - 86)	80 (67 - 91)	+6
优秀率 (≥90分)	16.4%	40.6%	+24.2%
不及格率 (<60分)	27.2%	16.8%	-10.4%
最高分 / 最低分	100/15	100/19	—

注：优秀率与不及格率组间差异均经 χ^2 检验验证，均具有统计显著性（ $p<0.05$ ）。

成绩分布的频率直方图与核密度估计（图1）直观展示了两组的分布形态差异：实验班分布整体右移，峰值区域向高分段偏移，分布更为集中（标准差较对照组小3.44分）；对照组分布则呈现明显的左偏特征，低分段占比更高。

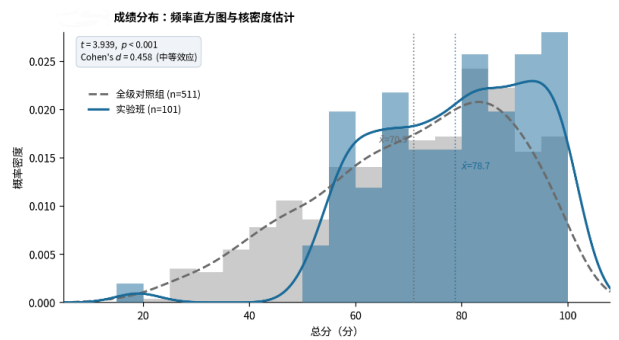


图1：成绩分布对比（频率直方图与核密度估计）

Figure 1 Score Distribution Comparison (Histogram and Kernel Density Estimation)

（二）成绩分布与等级结构

从等级结构分析（图2），实验班在优秀（≥90分）档次的占比为40.6%，较全级对照组（16.4%）提升24.2个百分点，差异经 χ^2 检验高度显著（ $\chi^2=26.74$ ， $p<0.001$ ）；“不及格”（<60分）档次由对照组的27.2%降至实验班的16.8%（ $\chi^2=5.21$ ， $p=0.022$ ）。值得注意的是，实验班的“良好”档次（75-89分）占比（30.7%）略低于对照组（31.9%），与之相对的是优秀率的大幅提升，提示AI融合教学对高分段学生的拉升效果尤为突出。

累积分布函数（CDF）分析进一步说明了两组之间的差别。Kolmogorov-Smirnov 双样本检验显示，两组CDF存在显著整

体偏移（ $D=0.208$ ， $p<0.001$ ）。从实际图形上看，实验班在各个分数节点上的累积百分比都更低，说明其整体成绩表现优于对照组，而不是只在某一小段分数上占优。

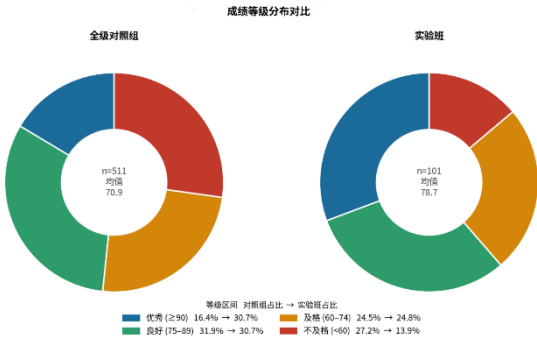


图2：成绩等级分布对比

Figure 2: Grade Distribution Comparison

（三）各题型得分率分析

按题型细看，实验班在8个题型上的得分率均高于对照组，说明AI融合教学对不同知识点都有一定帮助。

选择题和填空题方面，实验班得分率分别为84.1%和86.6%，高于对照组的79.1%和82.4%，两项差异均达到统计显著（ $p<0.05$ ）。相比之下，这类基础题的提升幅度不算特别大（分别提升5.0和4.2个百分点），可能是因为这部分内容更依赖日常练习和熟练度积累。

大题，也就是综合应用题，提升最明显。其中第7题提升23.5个百分点（ $p<0.001$ ），第8题提升13.3个百分点（ $p<0.001$ ），第4题提升12.1个百分点（ $p<0.001$ ）。这几道题都需要把多个知识模块连起来使用，例如线性变换、特征值分解和方程组求解等，说明AI融合教学对知识整合和迁移的帮助更突出。

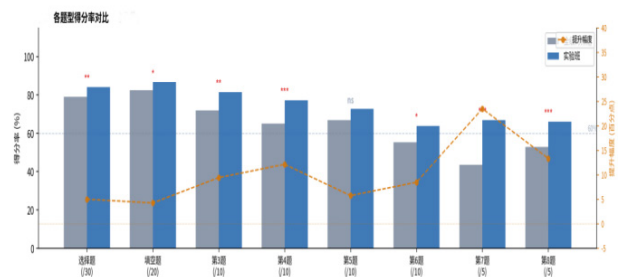


图3：各题型得分率对比与提升幅度

Figure 3: Section Score Rate Comparison and Improvement Margin

（四）效应量异质性分析

为进一步量化各题型的干预效果，计算各题型的Cohen's d 值并绘制森林图（图4）。结果显示，效应量存在显著的异质性：第7题效应量最大（ $d \approx 0.95$ ），远超其他题型；第4、8题次之（ d 分别约为0.60和0.55）；选择题和第5题效应量相对较小（ $d<0.3$ ），且第5题95%置信区间包含0值，提示该题型的组间差异在统计上尚不稳健。综合效应量 $d=0.458$ （95%CI不含0），对应中等效应量水平，证实研究假设H3成立。

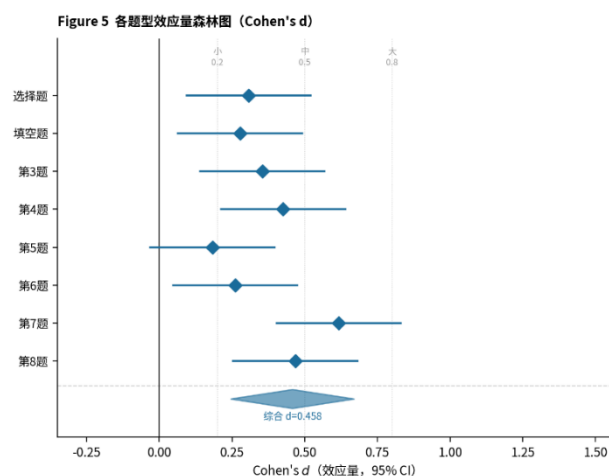


图4：各题型效应量森林图（Cohen's d, 95% 置信区间）

Figure4: ForestPlotofEffectSizesbyQuestionType(Cohen's d,95%CI)

（五）综合统计检验汇总

表2汇总了本研究所采用的全部统计检验结果，可见各方法均指向一致结论：AI 融合教学干预对线性代数成绩的提升效果具有极显著统计学意义，且在参数与非参数两类检验框架下均稳健成立，有效排除了结论依赖于分布假设的风险。

表2：假设检验综合结果汇总

检验方法	统计量	自由度 / 样本量	p 值	结论
独立样本 t 检验	$t=3.939$	$df=610$	<0.001	***
Mann-WhitneyU	$U=31815$	$n1=101, n2=511$	<0.001	***
K-S 双样本检验	$D=0.208$	—	<0.001	***
优秀率 χ^2 检验	$\chi^2=26.74$	$df=1$	<0.001	***
不及格率 χ^2 检验	$\chi^2=5.21$	$df=1$	0.022	*
综合效应量	$Cohen'sd=0.458$	中等效应	—	—

注：*** 表示 $p<0.001$ ，* 表示 $p<0.05$ ；Cohen's d 效应量解释标准：0.2= 小效应，0.5= 中效应，0.8= 大效应。

四、讨论

（一）结果解释

总体上看，本文结果与已有 AI 辅助教学研究的结论大体一致，即 AI 干预通常能够带来中等水平的正向效果^[9, 10]，但本文进一步看到，这种效果并不是平均铺开在所有题型上的，而是集中体现在综合应用题上。换句话说，AI 融合教学的优势更多在于帮助学生把零散知识串起来、用起来，而不仅仅是把记忆题做得更快。这一点也与 Mayer 关于多媒体学习的观点相吻合：恰当的视

觉呈现和符号呈现，有助于学生进入更深层的认知加工^[7]。

从成绩结构上看，实验班不及格率从27.2%降到16.8%，优秀率从16.4%升到40.6%，这两项变化是同时发生的。也就是说，AI 融合教学一方面帮助了原本基础薄弱的学生，另一方面也让成绩较好的学生有了进一步提升的空间，这和差异化教学中“分层支持”的思路比较一致。

（二）研究局限性

本研究也有以下几点局限，结论在解读时需要保持谨慎：

① 本研究是准实验设计，不是随机分组，因此实验班和对照组之间可能本来就存在学习动机等方面的差异，这一点很难完全排除。后续研究可以考虑随机分班，或者借助倾向得分匹配（PSM）尽量平衡组间差异。

② 本文主要使用期末考试成绩作为结果变量，没有充分覆盖学习过程中的变化，比如元认知策略、课堂投入度等。后续如果能加入过程性评价，结论会更完整。

③ 样本来源只有一所学校，外部推广时还需要更多学校、更多专业类型的验证，才能判断结论是否具有更强的普适性。

④ 教师因素没有完全控制。实验班和对照组如果由不同教师授课，教师风格本身也可能影响结果，这一部分在当前框架下不容易拆分出来。

五、结束语

本研究基于511名学生的真实考试数据，借助多种统计方法验证了人工智能融合教学在线性代数课程中的实际效果。研究结果支持三项核心假设：实验班总分显著优于全级对照基准（H1成立）；各题型得分率整体提升，综合应用题最为突出（H2成立）；综合效应量达到中等水平（H3成立）。

总体来看，以 AI 辅助教学为主线的融合教学模式，确实能够帮助学生更好地理解线性代数中的核心概念，也能在一定程度上提升他们解决综合题的能力。它对不同层次学生都有积极作用，说明这种模式不只适合“学得快”的学生，也能对基础较弱的学生起到补充和支持作用。本文所采用的准实验设计和统计分析路径，也可以为同类课程改革研究提供一个可借鉴的样本。后续研究可继续从随机化设计、长期追踪和多校验证三个方向推进，以进一步增强因果推断的说服力。

参考文献

- [1] 教育部等九部门关于加快推进教育数字化的意见 [J]. 中华人民共和国教育部公报, 2025, (04): 16-20.
- [2] 杨文霞, 王卫华, 何朗, 等. 知识图谱赋能智慧教育的研究与实践——以武汉理工大学“线性代数”课程为例 [J]. 高等教育教育研究, 2023 (6): 111-117.
- [3] 刘媛媛, 高小伟. 基于 AI 的混合式教学模式优化——以“线性代数”课程为例 [J]. 大学教育, 2025 (32): 72-75.
- [4] 涂建华, 肖碧怡, 姜广峰. 构建微积分知识图谱助推一流课程建设 [J]. 中国大学教学, 2020 (11): 33-37.
- [5] 龙珠, 陈祥伟, 董波. 工科数学实验教学体系融合思政的研究与探索 [J]. 实验室研究与探索, 2023, 42 (11): 216-219.
- [6] Sweller, J. Cognitive load theory, learning difficulty, and instructional design [J]. Learning and Instruction, 1994, 4(4): 295-312.
- [7] Mayer, R.E. Multimedia Learning (3rd ed.) [M]. Cambridge University Press, 2020.
- [8] Cohen, J. Statistical Power Analysis for the Behavioral Sciences (2nd ed.) [M]. Lawrence Erlbaum Associates, 1988.
- [9] Kulik, J.A., & Fletcher, J.D. Effectiveness of intelligent tutoring systems: A meta-analytic review [J]. Review of Educational Research, 2016, 86(1): 42-78.
- [10] Hattie, J. Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement [M]. Routledge, 2009.

基于组合赋权云模型的自动驾驶安全综合评价研究

詹媛媛¹, 周金明^{2*}

1. 安徽信息工程学院 通识教育与外国语学院, 安徽 芜湖 241000

2. 安徽工程大学 数理与金融学院, 安徽 芜湖 241000

DOI:10.61369/ASDS.2026060007

摘 要 : 自动驾驶安全已成为汽车制造、科技与半导体等行业亟待突破的关键议题。为建立科学、全面的自动驾驶安全评价体系, 文章构建了涵盖“感知系统、决策与规划、系统安全性与可靠性、人机交互、数据与伦理”五个维度的综合评价指标体系, 综合运用层次分析法(AHP)确定主观权重、CRITIC法确定客观权重, 并借助博弈论实现主客观权重的优化融合, 从而克服了单一赋权方法的局限性。进一步引入云模型理论, 将定性评价中的模糊性与随机性转化为定量分析。最后, 通过典型自动驾驶系统的案例进行实证分析, 验证了该方法的有效性与科学性, 为自动驾驶系统的安全水平提供了稳定、科学的综合评价。

关 键 词 : 自动驾驶安全; 博弈论; 组合赋权; 云模型

Combined Weighting and Cloud Model Based Comprehensive Safety Evaluation of Autonomous Driving

Zhan Yuanyuan¹, Zhou Jinming^{2*}

1.School of General Education and Foreign Languages, Anhui Institute of Information Technology, Wuhu, Anhui 241000

2.School of Mathematics-Physics and Finance, Anhui Polytechnic University, Wuhu, Anhui 241000

Abstract : Autonomous driving safety has become a critical issue requiring urgent breakthroughs in industries such as automotive manufacturing, technology, and semiconductors. To establish a scientific and comprehensive safety evaluation system for autonomous driving, this paper constructs a comprehensive evaluation index system covering five dimensions: "perception system, decision-making and planning, system safety and reliability, human-machine interaction, data and ethics." The Analytic Hierarchy Process (AHP) is used to determine subjective weights, and the CRITIC method is employed to calculate objective weights. By leveraging game theory, an optimal integration of subjective and objective weights is achieved, thereby overcoming the limitations of single-weighting methods. Furthermore, cloud model theory is introduced to transform the fuzziness and randomness in qualitative evaluation into quantitative analysis. Finally, empirical analysis is conducted using a typical autonomous driving system as a case study, verifying the effectiveness and scientific validity of the proposed method. This approach provides a stable and scientific comprehensive evaluation of the safety level of autonomous driving systems.

Keywords : autonomous driving safety; game theory; combination weighting; cloud model

引言

在现代科技持续突破的背景下, 自动驾驶技术早已从科幻走进现实, 逐步落地为真实可应用的技术形态。通过高精度传感器、计算机视觉与人工智能等核心技术, 自动驾驶系统可实现对周边行驶环境的实时感知, 及时完成动态决策并控制车辆运动来

实现自动驾驶^[1]。目前, 根据《汽车驾驶自动化分级》国家标准, 自动驾驶从L0到L5分为6个等级^[2], 每个等级的具体划分如表1所示:

基金项目: 国家社会科学基金(20BTJ048); 安徽省“四新”研究与改革实践项目(2024SX083); 安徽工程大学中青年拔尖人才培养计划(S022021054, S022022012); 研究生教育教学改革研究项目(2023YZL057); 安徽省高等学校科学研究项目(自然科学类)(2024AH050643); 安徽省高等学校科学研究项目(自然科学类)(2025AHGXZK30969)。

作者简介: 詹媛媛, 安徽信息工程学院通识教育与外国语学院, 助教, 研究方向为云模型理论与方法应用。

通讯作者: 周金明, 安徽工程大学数理与金融学院, 教授, 研究生导师, 研究方向为复杂模糊决策理论。

表1：自动驾驶等级表
Table 1: Autonomous driving level

智驾等级	自动化程度
L0: 无自动化	完全由驾驶员操作，系统仅提供警告（如盲区提醒，车辆偏离警告）
L1: 辅助驾驶	系统可以控制车辆的一个功能（如加速或者转向），但驾驶员需全程监控并负责驾驶
L2: 部分自动驾驶	系统可同时控制加速、减速和转向等功能，但驾驶员必须保持警觉并随时接管控制
L3: 有条件自动驾驶	在特定条件下，系统可完全接管驾驶任务，驾驶员无需持续监控，但需要在系统提醒时接管车辆
L4: 高度自动驾驶	系统在特定环境下（如高速公路或特定区域）能够完全自主驾驶，无需驾驶员干预
L5: 完全自动驾驶	系统在所有道路和环境条件下完全自主驾驶，无需驾驶员干预或控制

然而，随着自动驾驶等级不断提升，对自动驾驶技术的安全性进行评估至关重要^[3]。例如，在极端天气、施工路段、非标准道路标识等复杂环境下，系统可能会出现失效或识别错误；驾驶员因过度依赖自动驾驶而放松警惕，导致驾驶技能退化，失去对紧急情况判断的能力。除此之外，自动驾驶会增加数据隐私泄露和被滥用的风险^[4]。因此，安全问题是自动驾驶所面临的不可逾越的核心问题。建立一套科学、全面、量化的安全评价体系，不仅是技术研发迭代的“指挥棒”，也是政府进行监管认证、公众建立技术信任的“基石”。

目前，对自动驾驶安全的评估多采用单一方法（传统的实车测试或单一赋权方式），导致评价视角不够全面、主客观权重失衡、风险模糊性难以充分量化等问题^[5]。为此，本文尝试构建一个覆盖技术、功能、人机交互与伦理层面的多层次自动驾驶安全综合评价体系。主要创新包括：第一，系统整合从感知、决策到人机交互、数据伦理五个维度的指标；第二，提出一种基于博弈论组合赋权的方法，结合层次分析法的主观经验和 CRITIC 法的客观数据，以获取更合理的综合权重；第三，引入云模型理论，量化评估过程中的模糊性和随机性^[6]，从而实现对安全等级更直观、更稳定的判断。该研究不仅有助于突破单一评价方法的局限，也拓展了多属性决策方法在自动驾驶安全领域的应用，为高阶自动驾驶系统中模糊性安全风险的量化提供了新的思路和方法。

一、构建评价指标体系

自动驾驶（也称为无人驾驶）是指通过人工智能、传感器、高精地图和控制系统等技术，让汽车在没有人类驾驶员干预的情况下，自主完成感知环境、规划路径和操控行驶的过程。自动驾驶评价指标体系的构建是一个复杂且系统的工程，为了减少人为失误交通事故的发生；通过优化行驶路线，缓解拥堵，提升道路通行能力；以及平稳的驾驶模式可降低车辆尾气排放，能进一步推动绿色交通发展^[7]。

本文旨在全面、科学、客观地评估影响自动驾驶的安全性和可靠性等关键因素，分别从技术性能、环境感知、系统可靠性以及人机交互等多个维度构建影响车辆自动驾驶安全的指标体系，如图1所示。

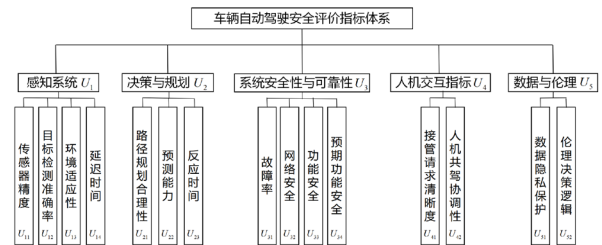


图1：安全评价指标体系

Figure 1: System of safety evaluation index

基于上述总结，构建无人驾驶指标体系是评估和优化自动驾驶系统性能的核心环节，围绕安全、可靠、合规及用户体验，最终形成感知系统、决策与规划、系统安全性与可靠性、人机交互^[8]和数据与伦理5个一级指标；传感器精度、目标检测准确率、环境适应性、延迟时间、路径规划合理性、预测能力、反应时间、故障率、网络安全、功能安全、预期功能安全、接管请求清晰度、人机共驾协调性、数据隐私保护和伦理决策逻辑15个二级指标多维度的车辆自动驾驶安全评价指标体系。

二、研究方法和模型

本文围绕自动驾驶系统的性能评估展开研究，首先依据科学性、针对性、可操作性等原则选取核心评价指标，构建多维度评估体系；进而分别采用层次分析法（AHP）挖掘专家主观判断信息以确定主观权重，借助 CRITIC 法从指标数据本身的波动性与关联性出发计算客观权重；最后引入博弈论思想对主客观权重进行组合优化，形成兼顾主观经验与客观数据特征的综合权重，为自动驾驶系统的性能量化评估提供更为可靠的方法论支持，以期和技术迭代与应用落地提供理论参考。

（一）AHP 法

层次分析法（AHP 法）是一种将复杂决策问题分解为目标层、准则层和方案层，通过两两比较确定权重，最终得出最优方案的主观权重综合决策分析方法^[9]。采用 AHP 法计算评价指标权重的主要步骤如下：

1. 构建层次结构

明确决策目标（如“选择对车辆自动驾驶安全影响最大的评价指标”），梳理评价体系内各因素的关联，搭建层级递进的结

构模型。

2. 构造判断矩阵

多个专家将各个指标相互比较后,使用1~9标度法(如表2所示)标记两个评价指标的相对重要程度,构造比较矩阵:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}. \quad (1)$$

3. 计算权重与一致性检验

(1) 计算权重:通过“和积法”计算各元素的权重(即同一层次中各元素对上层基准的重要程度占比),每一个因素对于其准则的相对权重计算公式为:

$$\omega_i = \frac{1}{n} \frac{\sum_{j=1}^n a_{ij}}{\sum_{k=1}^n a_{kj}}, \quad i = 1, 2, \dots, n \quad (2)$$

式中: ω_i 为因素 i 的指标权重值。

(2) 一致性检验:在构造判断矩阵时,专家对多个指标的重要性的判断容易受主观影响,指标进行风险评估时,为了确保权重计算结果的科学性和可信度,保证判断矩阵是一致的,所以需要对其进行一致性检验,计算公式如下:

$$CR = \frac{CI}{RI} \quad (3)$$

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (4)$$

式中: CR 为一致性比率,若 $CR < 0.1$,说明判断矩阵逻辑一致,权重有效;若 $CR \geq 0.1$,需重新调整两两比较的结果。 CI 为一致性指标,式(4)中, $\lambda_{\max}$ 为判断矩阵的最大特征值, n 为矩阵阶数。 RI 为平均随机一致性指标,可通过标准对照表获取^[10]。

表2: 1~9标度法的含义
Table2: 1~9 Scale method and meaning^[9]

标度	含义
1	两因素相比,因素 <i>i</i> 同因素 <i>j</i> 同样重要
3	两因素相比,因素 <i>i</i> 比因素 <i>j</i> 稍微重要
5	两因素相比,因素 <i>i</i> 比因素 <i>j</i> 显著重要
7	两因素相比,因素 <i>i</i> 比因素 <i>j</i> 非常重要
9	两因素相比,因素 <i>i</i> 比因素 <i>j</i> 极其重要
2, 4, 6, 8	上述两相邻判断的中值
倒数	<i>i</i> 和 <i>j</i> 相比,如果标度为3,则 <i>i</i> 和 <i>j</i> 相比就是1/3

(二) CRITIC 法

CRITIC 法是一种客观赋权法,用于多指标评价体系中确定各指标的权重,核心是通过指标的对比值(数据差异程度)和冲突性(指标间相关程度)综合计算权重,对比值越大、冲突性越强,则指标权重越大^[11]。CRITIC 法是完全基于数据客观计算,同时考虑指标的数据差异与信息重叠,避免了单一赋权的偏差,能更精确反映各指标的实际贡献度,其具体步骤如下^[12]:

1. 数据标准化

构建数据矩阵 $X = (x_{ij})_{m \times n}$, 其中 x_{ij} 表示第 i 个样本第 j 项评价指

标的数值 ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$)。为消除因量纲不同对评价结果造成的影响,对数据进行正向化或逆向化处理,标准化处理后的数据为 x'_{ij} 。

正向化处理:

$$x'_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}}. \quad (5)$$

逆向化处理:

$$x'_{ij} = \frac{x_{\max} - x_{ij}}{x_{\max} - x_{\min}}. \quad (6)$$

2. 计算指标对比度

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^m (x_{ij} - \bar{x}_j)^2}{n-1}}, \quad \bar{x}_j = \frac{1}{n} \sum_{i=1}^m x_{ij}. \quad (7)$$

其中 σ_j 表示 j 项评价指标的标准差。从指标特性来看,标准差的大小与数据差异度正相关:标准差越大,意味着该指标的样本数据离散程度越高,越能反映出更多的信息,说明评价力度也就越强,故而在权重设置中需要给予其更大的比重。

3. 计算指标冲突性

$$C_j = \sum_{i=1}^m (1 - r_{ij}). \quad (8)$$

$$I_j = \sigma_j \times C_j. \quad (9)$$

式中: C_j 表示指标冲突性; r_{ij} 表示评价指标之间的相关系数, $1 - r_{ij}$ 越大,指标 j 和 j 的冲突性越强,信息重叠越少; I_j 为指标信息量。

4. 计算权重

$$\omega_j = \frac{I_j}{\sum_{j=1}^n I_j}. \quad (10)$$

(三) 博弈论组合赋权

博弈论组合赋权法是一种基于博弈论思想,对主观赋权法和客观赋权法得到的权重向量进行协调优化,最终得到综合权重的方法^[13]。通过权重之间的“博弈均衡”,实现主客观权重的最优组合。具体步骤如下^[14]:

1. 确定基础权重向量集

分别采用 AHP 和 CRITIC 计算出同一评价指标体系下的初始权重的线性组合:

$$W_k = [\omega_{k1}, \omega_{k2}, \dots, \omega_{kn}], \quad k = 1, 2, \dots, L \quad (11)$$

式中: W_k 为第 k 种方法确定的权重集合, n 为评价指标的数量, L 为确定权重的方法数量。

2. 构建组合权重模型

设组合权重向量为 ω , 将其表示为初始权重的线性组合:

$$\omega = \sum_{k=1}^L \alpha_k W_k^T \quad (12)$$

ω 为组合权重向量； α_k 为对应的线性组合系数， $\alpha_k > 0$ ； W_k^T 为向量集的转置矩阵。

3. 建立博弈均衡目标函数

为使组合权重与各初始权重的偏差最小，构建极小化偏差的目标函数：

$$\min \left\| \sum_{k=1}^L \alpha_k W_k^T - W_i^T \right\|_2, \quad i, k = 1, 2, \dots, L \quad (13)$$

式中： $\sum_{k=1}^L \alpha_k W_k^T$ 表示为某种权重的线性组合； W_i^T 为第 i 种方法确定的权重向量转置； $\|\cdot\|_2$ 为范数。将式 (13) 根据微分的性质展开，得到线性方程组 (14)。

$$\begin{bmatrix} \omega_1 \omega_1^T & \omega_1 \omega_2^T & \cdots & \omega_1 \omega_L^T \\ \omega_2 \omega_1^T & \omega_2 \omega_2^T & \cdots & \omega_2 \omega_L^T \\ \vdots & \vdots & \ddots & \vdots \\ \omega_L \omega_1^T & \omega_L \omega_2^T & \cdots & \omega_L \omega_L^T \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_L \end{bmatrix} = \begin{bmatrix} \omega_1 \omega_1^T \\ \omega_2 \omega_2^T \\ \vdots \\ \omega_L \omega_L^T \end{bmatrix} \quad (14)$$

求解得出最优组合系数 $\alpha^* = (\alpha_1, \alpha_2, \dots, \alpha_k)$ 。

4. 求解组合系数并归一化

在上述求得最优组合系数之后，进行归一化处理，确保 $\sum_{k=1}^L \alpha_k = 1$ 。

$$\alpha_k^* = \frac{|\alpha_k|}{\sum_{k=1}^L |\alpha_k|} \quad (15)$$

5. 计算最终综合权重

将归一化后的组合系数代入式 (12)，计算得到最终的博弈论组合权重向量^[14]。

$$\omega = \sum_{k=1}^L \alpha_k^* \omega_k^T \quad (16)$$

三、基于云模型的综合评价方法

云模型^[15]是一种实现定性概念与定量数值之间不确定性转换的数学模型，核心是融合模糊性与随机性，解决传统模糊理论忽略随机性、概率论难以处理模糊语义的问题。目前，云模型广泛应用于不确定性评价、决策分析、模式识别等领域，比如多属性指标重要性排序、风险评估、图像识别等场景。

云模型适用于自动驾驶安全性影响指标的评估，是因为自动驾驶的指标评价本身带有模糊性（“较安全”、“相对可靠”这类定性描述无明显边界）和随机性（同一指标在不同路况和不同天气下的表现存在波动），而云模型能同时刻画这两种不确定性，比传统的确定性方法更贴合实际。自动驾驶的影响指标体系复杂且存在耦合关系，云模型可通过多概念云的叠加、对比分析，只管呈现不同指标的影响程度差异，为多属性决策提供清晰的可视化依据。云模型计算步骤如下^[16]：

（一）确定评语集与标准云

首先根据构建的评价指标体系建立评语集，通过标准云将评语集进行量化，把评价等级划分为 h 个区间。其中，第 i 个子区间 $[x_i^{\min}, x_i^{\max}]$ 所对应的标准云数字特征为 (Ex_i, En_i, He_i) ，具体转换公式如下^[14]：

$$\begin{cases} Ex_i = \frac{x_i^{\min} + x_i^{\max}}{2} \\ En_i = \frac{x_i^{\max} - x_i^{\min}}{6} \\ He_i = \frac{En_i}{10} \end{cases} \quad (17)$$

式中， $x_i^{\min}$ 和 $x_i^{\max}$ 分别为该评价子区间中的最大值和最小值； k 为常数，根据评价指标的不确定性进行选取，一般取值在 0.001~0.1 之间。

（二）确定评价指标云及综合评价云^[17]

邀请自动驾驶行业技术专业性强、实践经验丰富、领域覆盖度广的 m 个专家，依据划分的风险评价等级对二级评价指标进行打分，汇总得到专家评分矩阵 Z 。

$$Z = \begin{bmatrix} x_{11} & \cdots & x_{1j} \\ \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ij} \end{bmatrix} \quad (18)$$

式中， x_{ij} 表示第 i 个专家对第 j 个二级指标的评价值。然后，利用逆向云发生器计算二级指标的云数字特征，公式如下：

$$\begin{cases} Ex_j = \frac{1}{m} \sum_{i=1}^m x_{ij} \\ En_j = \sqrt{\frac{\pi}{2}} \times \frac{1}{m} \sum_{i=1}^m |x_{ij} - Ex_j| \\ S_j^2 = \frac{1}{m-1} \sum_{i=1}^m (x_{ij} - Ex_j)^2 \\ He_j = \sqrt{|S_j^2 - En_j^2|} \end{cases} \quad (19)$$

（三）确定评价等级，得到评价结果

根据上述所得指标的云数字特征和指标权重，计算可得综合评价云的数字特征，计算公式如下^[18]：

$$\begin{cases} Ex = \sum_{j=1}^n \omega_j Ex_j \\ En = \sqrt{\sum_{j=1}^n \omega_j En_j^2} \\ He = \sqrt{\sum_{j=1}^n \omega_j He_j^2} \end{cases} \quad (20)$$

式中， n 为指标个数， ω_j 为第 j 个指标的权重， (Ex_j, En_j, He_j) 为指标云数字特征。

将综合云数字特征输入到正向云发生器得到综合评价云图，并于标准云图进行比较，即可确定评价等级。

四、实证分析

(一) 实例概况

以某 L3 级自动驾驶系统为研究对象^[19]，围绕其安全运行能力构建多维度评价体系，通过实地测试、数据采集与量化分析相结合的方式，对该 L3 级自动驾驶系统的安全性能进行全面测度，旨在验证所构建评价体系的科学性与适用性，为 L3 级自动驾驶系统的安全优化与行业标准制定提供实证依据。

邀请 10 位自动驾驶行业技术专业性强、实践经验丰富、领域覆盖度广的专家结合专业知识以及实际情况比较自动驾驶安全的影响因素，并对其影响程度进行打分。对数据进行处理后，运用组合赋权法得到各级指标的综合权重。

(二) 评价指标权重确定

1. 层次分析法计算主观权重

邀请专家采用 1-9 标度法进行两两比较，构造判断矩阵，并利用式(1)一式(4)计算得到一级评价指标的主观权重，如表3所示：

表3：一级指标判断矩阵表

Table 3: Judgment matrix of first-level indicators

	U_1	U_2	U_3	U_4	U_5	ω_i
U_1	1	3	1/2	2	2	0.254
U_2	1/3	1	1/3	1	2	0.139
U_3	2	3	1	3	2	0.363
U_4	1/2	1	1/3	1	1	0.122
U_5	1/2	1/2	1/2	1	1	0.122
CR_1	$CR=0.040<0.100$					

一致性检验结果： $CI = \frac{\lambda_{\max} - n}{n - 1} = \frac{5.178 - 5}{5 - 1} = 0.045$ 。

$n=5$ 矩阵的平均随机一致性指标^[10]： $RI=1.120$ 。

随机一致性比率： $CR = \frac{CI}{RI} = 0.040 < 0.100$ ，通过一致性检验。

同样得到二级指标感知系统 U_1 的计算结果，如表4所示：

表4：感知系统 U_1 判断矩阵表

Table 4: Judgment matrix of perception system U_1

	U_{11}	U_{12}	U_{13}	U_{14}	ω_i
U_{11}	1	2	1/3	3	0.233
U_{12}	1/2	1	1/4	2	0.140
U_{13}	3	4	1	5	0.542
U_{14}	1/3	1/2	1/5	1	0.085
CR_2	$CR=0.019<0.100$				

一致性检验结果： $CI = \frac{\lambda_{\max} - n}{n - 1} = \frac{4.051 - 4}{4 - 1} = 0.017$ 。

$n=4$ 矩阵的平均随机一致性指标： $RI=0.890$ 。

随机一致性比率： $CR = \frac{CI}{RI} = 0.019 < 0.100$ ，通过一致性检验。

二级指标决策与规划 U_2 的计算结果，如表5所示：

表5：决策与规划 U_2 判断矩阵表

Table 5: Judgment matrix of decision-making and planning U_2

	U_{21}	U_{22}	U_{23}	ω_i
U_{21}	1	3	5	0.633
U_{22}	1/3	1	3	0.260
U_{23}	1/5	1/3	1	0.106
CR_3	$CR=0.037<0.100$			

一致性检验结果： $CI = \frac{\lambda_{\max} - n}{n - 1} = \frac{3.038 - 3}{3 - 1} = 0.019$ 。

$n=3$ 矩阵的平均随机一致性指标： $RI=0.520$ 。

随机一致性比率： $CR = \frac{CI}{RI} = 0.037 < 0.100$ ，通过一致性检验。

二级指标系统安全性与可靠性 U_3 的计算结果，如表6所示：

表6：安全性与可靠性 U_3 判断矩阵表

Table 6: Judgment matrix of safety and reliability U_3

	U_{31}	U_{32}	U_{33}	U_{34}	ω_i
U_{31}	1	1/3	1/4	1/5	0.074
U_{32}	3	1	1/2	1/3	0.171
U_{33}	4	2	1	1/2	0.284
U_{34}	5	3	2	1	0.471
CR_4	$CR=0.01940<0.100$				

一致性检验结果： $CI = \frac{\lambda_{\max} - n}{n - 1} = \frac{4.051 - 4}{4 - 1} = 0.017$ 。

$n=4$ 矩阵的平均随机一致性指标： $RI=0.890$ 。

随机一致性比率： $CR = \frac{CI}{RI} = 0.019 < 0.100$ ，通过一致性检验。

二级指标人机交互 U_4 的计算结果，如表7所示：

表7：人机交互 U_4 判断矩阵表

Table 7: Judgment matrix of human-machine interaction U_4

	U_{41}	U_{42}	ω_l
U_{41}	1	3	0.750
U_{42}	1/3	1	0.250
CR_5	$CR=0<0.100$		

对于2阶判断矩阵 ($n=2$)，自动满足一致性。二级指标数据与伦理 U_5 的计算结果，如表8所示：

表8：数据与伦理 U_5 判断矩阵表

Table 8: Judgment matrix of data and ethics U_5

	U_{41}	U_{42}	ω_i
U_{41}	1	2	0.667
U_{42}	1/2	1	0.333
CR_6	$CR=0<0.100$		

层次分析法结果如表9所示：

表9：层次分析法权重计算结果汇总

Table 9: Summary of weight calculation results by AHP

一级指标	权重	二级指标	权重	最终权重
感知系统 (U_1)	0.254	传感器精度 (U_{11})	0.233	0.059
		目标检测准确率 (U_{12})	0.140	0.035
		环境适应性 (U_{13})	0.542	0.138
		延迟时间 (U_{14})	0.085	0.022
决策与规划 (U_2)	0.139	路径规划合理性 (U_{21})	0.633	0.088
		预测能力 (U_{22})	0.260	0.036
		反应时间 (U_{23})	0.106	0.015
		故障率 (U_{31})	0.074	0.027
系统安全性与可靠性 (U_3)	0.363	网络安全 (U_{32})	0.171	0.062
		功能安全 (U_{33})	0.284	0.103
		预期功能安全 (U_{34})	0.471	0.171
		接管请求清晰度 (U_{41})	0.750	0.092
人机交互指标 (U_4)	0.122	人机共驾协调性 (U_{42})	0.250	0.031
		数据隐私保护 (U_{51})	0.667	0.081
数据与伦理 (U_5)	0.122	伦理决策逻辑 (U_{52})	0.333	0.041

2.CRITIC 法计算客观权重

首先，对车辆自动驾驶的评价指标标准化处理。其中，自动驾驶车辆的15个评价指标中延迟时间、反应时间和故障率为负向指标，其余均为正向指标，最终得到数据的标准化处理结果如表10所示：

表10：数据标准化处理结果

Table 10: Results of data standardization processing

系统指标	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
U_{11}	1.000	0.811	0.378	0.135	0.622	0.541	0.730	0.000
U_{12}	0.795	1.000	0.409	0.114	0.705	0.568	0.727	0.000
U_{13}	1.000	0.524	0.755	0.204	0.633	0.449	0.673	0.000
U_{14}	1.000	0.814	0.571	0.214	0.671	0.429	0.614	0.000
U_{21}	1.000	0.367	0.929	0.173	0.561	0.500	0.755	0.000
U_{22}	0.929	1.000	0.571	0.262	0.726	0.536	0.786	0.000
U_{23}	1.000	0.643	0.714	0.214	0.571	0.429	0.786	0.000
U_{31}	0.162	1.000	0.723	0.000	0.523	0.823	0.392	0.038
U_{32}	1.000	0.303	0.545	0.894	0.197	0.000	0.652	0.818
U_{33}	1.000	0.000	0.233	0.933	0.433	0.067	0.667	0.800
U_{34}	1.000	0.416	0.650	0.153	0.504	0.299	0.693	0.000
U_{41}	0.924	0.456	1.000	0.177	0.608	0.418	0.835	0.000
U_{42}	0.955	0.466	1.000	0.216	0.625	0.523	0.830	0.000
U_{51}	1.000	0.000	0.560	0.980	0.430	0.300	0.770	0.890
U_{52}	0.787	0.080	0.413	1.000	0.227	0.000	0.600	0.893

将标准化后的数据用 CRITIC 法计算得到各指标的客观权重和标准差^[20-21]。计算结果如表11所示：

表11：指标层 CRITIC 权重计算结果

Table 11: CRITIC weight calculation results of the index layer

项	指标变异性 σ_j	指标冲突性 C_j	信息量 I_j	权重 ω_i
U_{11}	0.340	7.182	2.441	0.050
U_{12}	0.344	8.614	2.966	0.061
U_{13}	0.315	6.011	1.896	0.039
U_{14}	0.321	6.683	2.144	0.044
U_{21}	0.352	6.444	2.268	0.047
U_{22}	0.338	7.491	2.529	0.052
U_{23}	0.322	5.890	1.894	0.039
U_{31}	0.374	13.914	5.205	0.107
U_{32}	0.357	13.003	4.640	0.095
U_{33}	0.390	13.747	5.366	0.110
U_{34}	0.320	5.703	1.826	0.038
U_{41}	0.358	6.531	2.335	0.048
U_{42}	0.352	6.583	2.316	0.048
U_{51}	0.358	14.003	5.009	0.103
U_{52}	0.379	15.438	5.851	0.120

3. 博弈论组合赋权

上述计算中，AHP 法计算的主观权重和 CRITIC 法计算的客观权重，结果明显存在偏差。为了有效平衡主观偏好的片面性与客观数据的偶然性，采用博弈论确定主客观权重组合的最优组合系数，计算最终得出纳什均衡解： $a_1=0.6245$ 和 $a_2=0.3755$ 。最终确定各指标的综合权重如表12所示：

表12：一级指标、二级指标的综合权重

Table 12: Comprehensive weights of first-level and second-level indicators

一级指标	主观权重	客观权重	综合权重	二级指标	主观权重	客观权重	综合权重
感知系统 (U_1)	0.254	0.194	0.231	传感器精度 (U_{11})	0.059	0.050	0.056
				目标检测准确率 (U_{12})	0.035	0.061	0.045
				环境适应性 (U_{13})	0.138	0.039	0.101
				延迟时间 (U_{14})	0.022	0.044	0.030
决策与规划 (U_2)	0.139	0.137	0.138	路径规划合理性 (U_{21})	0.088	0.047	0.072
				预测能力 (U_{22})	0.036	0.052	0.042
				反应时间 (U_{23})	0.015	0.039	0.024
				故障率 (U_{31})	0.027	0.107	0.057
系统安全性与可靠性 (U_3)	0.363	0.350	0.358	网络安全 (U_{32})	0.062	0.095	0.075
				功能安全 (U_{33})	0.103	0.110	0.106
				预期功能安全 (U_{34})	0.171	0.038	0.121
				接管请求清晰度 (U_{41})	0.092	0.048	0.075
人机交互指标 (U_4)	0.122	0.096	0.112	人机共驾协调性 (U_{42})	0.031	0.048	0.037
				数据隐私保护 (U_{51})	0.081	0.103	0.089
数据与伦理 (U_5)	0.122	0.223	0.160	伦理决策逻辑 (U_{52})	0.041	0.120	0.071

(三) 综合评价云模型

1. 标准云图绘制及区间确定^[22]

根据指标的相对重要性，假设有效论域的范围是 [60,100]，将

评价等级分为四个等级，具体划分如表13所示，其中分值越高，代表该指标对自动驾驶安全的重要性越大。根据评价标准等级的特征参数，绘制得到标准云图，如图2所示：

表 13: 评价标准等级划分

Table 13: Classification of evaluation standard grades

评语集	对应区间	特征参数
低重要性 (LI)	[60,70)	(65,1.667 0.100)
中等重要性 (MI)	[70,80)	(75,1.667 0.100)
高重要性 (HI)	[80,90)	(85,1.667 0.100)
极高重要性 (VHI)	[90,100)	(95,1.667 0.100)

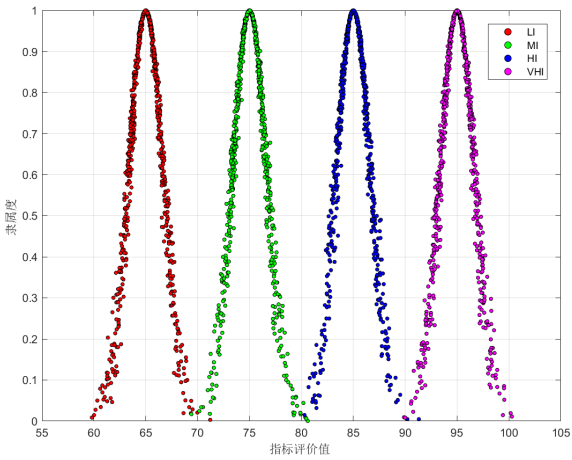


图 2: 标准云图

Figure 2: Graph of standard cloud

2. 评价指标云

为了更加直观的呈现影响自动驾驶安全指标的评价等级, 根据计算所得一级指标的数字特征^[14], 生成综合评价云如图3-7所示:

表 14: 一级指标数字特征

Table 14: Numerical characteristics of first-level indicators

一级指标	数字特征	重要性等级
U_1	(85.446,3.028,0.774)	偏高重要性
U_2	(87.343,3.373,0.900)	偏高重要性
U_3	(92.028,3.292,0.910)	偏极高重要性
U_4	(82.088,3.468,0.946)	高重要性与中等重要性之间
U_5	(82.238,3.356,0.892)	高重要性与中等重要性之间

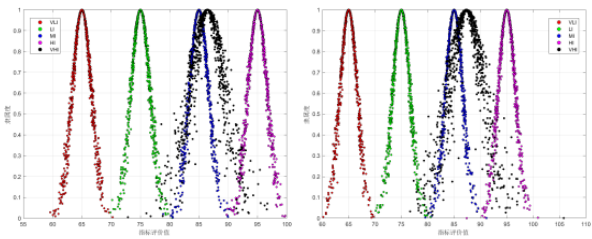


图 3: 感知系统评价云图

Figure 3: Evaluation cloud graph of perception system

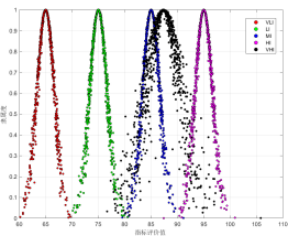


图 4: 决策与规划评价云图

Figure 4: Evaluation cloud graph of decision-making and planning

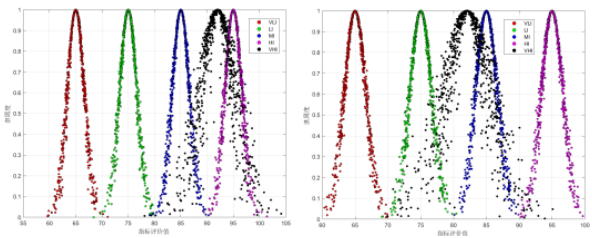


图 5: 系统安全性与可靠性评价云图

Figure 5: Evaluation cloud graph of system safety and reliability

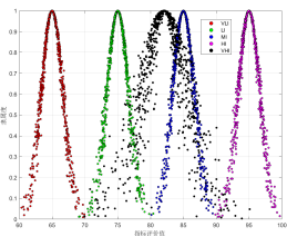


图 6: 人机交互评价云图

Figure 6: Evaluation cloud graph of human-machine interaction

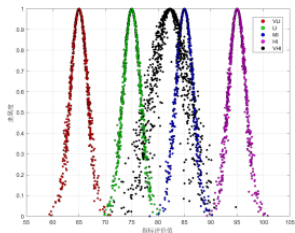


图 7: 数据与伦理评价云图

Figure 7: Evaluation cloud graph of data and ethics

由一级指标数字特征以及评价云图可知, 5个一级指标对自动驾驶系统的重要性排序如下: $U_3 \succ U_2 \succ U_1 \succ U_5 \succ U_4$ 。说明系统安全性与可靠性对自动驾驶影响最大, 该指标主要反映了系统故障率、网络安全、功能安全和预期功能安全。实际情况中, 系统安全性与可靠性同样在自动驾驶中重要性最大, 属于自动驾驶的安全基石。因为该指标的核心就是保障驾驶安全, 避免交通事故, 获取公众信任。因此, 研发自动驾驶系统时, 应注重增强对该项指标的研发力度与要求。

决策与规划和感知系统对自动驾驶的影响几乎相同, 都具有高重要性。当汽车自动驾驶时, 感知能力的大小直接定义了自动驾驶的运行设计域, 说明感知是决策的基础, 决策是对执行命令的输出。一旦感知失效, 系统将面临灾难的安全隐患。数据与伦理和人机交互指标在自动驾驶中的重要性大致相同。数据与伦理定义了法律和道德边界, 使自动驾驶系统在紧急情况下的决策涉及到生命价值的权衡。当自动驾驶由条件自动驾驶升级到完全自动驾驶时, 人机交互指标影响性相对于其它四个指标时, 重要性就会削弱。

五、结束语

文章通过融合 AHP、CRITIC 与博弈论的组合赋权法, 并引入云模型, 构建了一套科学平衡主客观信息的自动驾驶性能评价体系。主要特点如下:

(一) 围绕影响自动驾驶性能的核心指标展开系统评价

首先, 采用层次分析法 (AHP) 对各指标进行主观赋权, 反映专家经验与行业判断; 其次, 引入 CRITIC 法计算客观权重, 体现数据本身的变异性与冲突性; 最后, 基于博弈论组合赋权方法, 确定主客观权重融合的最优系数, 实现了主客观信息的平衡与权重分配的合理化。在实证分析阶段, 引入云模型以刻画评价过程中的模糊性与随机性, 实现定性认知与定量数值之间的映射, 并通过云图的可视化表达, 直观呈现评价结果的分布特性与稳定程度, 为自动驾驶系统性能的综合评估提供了新的方法论支撑。

(二) 构建的评价体系与技术路径具备较强的现实应用价值, 可直接服务于自动驾驶车辆的研发测试、准入评估及运维优化

通过系统量化感知系统可靠性、决策逻辑合理性、网络安全防护水平等关键性能指标的安全水平, 可为整车企业提供精准的技术提升方向; 为监管机构提供统一、量化的测评工具, 助力

破解当前行业面临的“测试资源不足、评估能力不匹配”等瓶颈问题。同时，该体系还可为消费者提供直观易懂的安全性能参考，提升公众对自动驾驶技术的认知与信任，从而推动自动驾驶产业向安全、高效、可持续的方向发展。

（三）尽管自动驾驶技术已取得长足进展，L5级完全自动驾驶仍面临诸多深层次挑战

首先是如何有效应对长尾分布的现实场景，包括极端天气条件、突发交通事件以及复杂的非结构化道路环境，这些都对感知

系统的鲁棒性与决策系统的泛化能力提出了更高要求。其次，如何在风险不可避免的情境下设计优先级避让策略，使系统能够遵循“最小伤害原则”进行自主决策，不仅涉及技术层面的路径规划与控制策略，更深层牵涉法律规制、道德伦理、社会接受等多维复杂议题。这需要政府、科研机构与产业界形成合力，从法规建设、技术攻关与伦理框架等多方面协同推进，为完全自动驾驶时代的真正到来奠定基础。

参考文献

[1] 朱冰, 范天昕, 赵文博等. 自动驾驶汽车连续测试场景复杂度评估方法 [J]. 吉林大学学报 (工学版), 2025, 55(02): 456-467.

[2] 全国汽车标准化技术委员会 (SAC/TC 114). 汽车驾驶自动化分级: GB/T 40429-2021[S]. 中国标准出版社, 2021.

[3] 王春雨. 汽车自动驾驶技术安全性多维度评估研究 [J]. 汽车测试报告, 2025, (14): 82-84.

[4] 刘权. 自动驾驶风险及其包容审慎监管 [J]. 河北学刊, 2025, 45(04): 25-33.

[5] 赵维伟. F 公司自动驾驶汽车开发项目安全风险管理研究 [D]. 吉林大学, 2025.

[6] ZHOU Jinming, BALEZENTIS T, STREIMIKIENE D. Normalized Weighted Bonferroni Harmonic Mean-Based Intuitionistic Fuzzy Operators and Their Application to the Sustainable Selection of Search and Rescue Robots[J]. Symmetry, 2019, 11(2): 218-241.

[7] 王媛媛. 考虑不确定性的无人驾驶汽车轨迹规划算法研究 [D]. 吉林大学, 2021.

[8] WAYTZ A, HEAFNER J, EPLEY N. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle[J] Journal of Experimental Social Psychology, 2014, 52: 113-117.

[9] 杨戟, 杨树清, 范颖等. 基于正态云组合赋权-TOPSIS-灰色关联法的加工中心可靠性分配 [J]. 机床与液压, 2025, 53(10): 30-39.

[10] 吴靖. Y 公司智能驾驶项目风险评价及策略研究 [D]. 安徽理工大学, 2025.

[11] 周晓晖, 王萱, 黄兰. 电动汽车制造企业供应链韧性评-价研究-基于组合赋权-云模型 [J]. 中国物价, 2024, (07): 47-51+57.

[12] 李姗姗, 周弘毅. 基于 CRITIC-云模型的石化行业节能减排技术评价 [J]. 河北科技大学学报, 2025, 46(01): 108-118.

[13] 邓乐乐, 郭生练, 王俊等. 基于博弈论-云模型的湖北省汉江中下游地区水资源承载力评价 [J]. 水资源保护, 2025, 41(02): 200-208.

[14] 郭国强, 牛智奇, 李亚超等. 综合最优赋权和云模型的导引头方案优选方法 [J]. 系统工程与电子技术, 2025, 47(09): 3011-3021.

[15] LI Deyi. Artificial Intelligence with Uncertainty[M]. Beijing: National Defense Industry Press, 2005. 171-177.

[16] ZHOU Jinming, ZHAN Yuanyuan, CHEN Sibao. Multi-Dimensional Cloud Model-Based Assessment and Its Application to the Risk of Supply Chain Financial Companies[J]. International Journal of Fuzzy System Applications (IJFSA), 2024, 13(1): 1-29.

[17] 王金花. 基于云模型的 LR 高速公路 5 标段工程施工安全风险评价 [D]. 江西理工大学, 2024.

[18] 苏为华, 周金明. 基于云理论的统计信息质量评估方法研究 [J]. 统计研究, 2018, 35(04): 86-93.

[19] 林庆峰, 王兆杰, 鲁光泉. L3 级自动驾驶汽车的接管安全性评价模型 [J]. 汽车工程, 2019, 41(11): 1258-1264.

[20] 石古乐, 张玉召, 常全盛等. 基于云模型的动力锂电池铁路运输风险评估 [J/OL]. 铁道标准设计, 1-10.

[21] 刘熙. 基于博弈-可拓云模型的石家庄市轨道交通与常规公交换乘服务评价研究 [D]. 石家庄铁道大学, 2023.

[22] PENG Bo, ZHOU Jinming, PENG Dinghong. Cloud model based approach to group decision making under uncertain pure linguistic information[J]. Journal of Intelligent & Fuzzy Systems, 2017, 32(3): 1959-1968.

双模型校验驱动的 C 语言智能出题系统研究

江玉珍, 朱映辉

韩山师范学院 计算机与信息工程学院, 广东 潮州 521041

DOI:10.61369/ASDS.2026060008

摘 要 : 针对 C 语言教学试题需求量大、传统出题效率低、单一 AI 生成质量不稳定等问题, 本文基于 COZE 智能体平台设计并实现了一款智能出题系统。系统创新性地采用双 LLM 校验机制, 实现生成与审核分离, 结合知识库检索与 AI 生成的混合题源策略。实验结果表明, 系统出题效率较传统模式有明显提升, 且有效地保障了试题质量与多样性。

关 键 词 : C 语言; 智能体; 智能出题; 双 LLM 校验

Research on Intelligent C Language Question Generation System Driven by Dual-LLM Verification

Jiang Yuzhen, Zhu Yinghui

School of Computer and Information Engineering, Hanshan Normal University, Chaozhou, Guangdong 521041

Abstract : To address the issues of high demand for C language teaching test questions, low efficiency of traditional question generation, and unstable quality of single-AI generation, this paper designs and implements an intelligent question generation system based on the COZE agent platform. The system innovatively adopts a dual-LLM verification mechanism to separate question generation from review, and combines a hybrid question source strategy of knowledge base retrieval and AI generation. Experimental results show that the system significantly improves question generation efficiency compared with traditional methods and effectively ensures the quality and diversity of test questions.

Keywords : C language; intelligent agent; intelligent question generation; dual-LLM verification

引言

C 语言程序设计是计算机专业的核心基础课, 涵盖了程序设计思维、语法规则、算法实现等知识领域、也是数据结构、算法设计、操作系统等多门专业课的基础先修课, 同时还承担着转专业考试、专升本考试等重要考核任务, 是计算机专业人才培养中重要的一门核心课程, 其教学质量直接影响着学生的专业素养及相关课程的通过率^[1]。C 语言课程内容多, 逻辑强, 对实践及技能要求较高, 其在教学、作业、考试、实训、竞赛等环节中均需要大量不同类型、知识点及难度的试题^[2]。传统出题模式主要依赖教师手动完成, 其存在三方面缺点: 一是效率低下, 教师需花费大量时间梳理知识点、设计题干, 难以应对高频次需求; 二是质量参差, 手动出题易出现表述模糊、难度失衡; 三是重复率高, 受限于个人知识储备, 试题同质化严重。近年来, 大语言模型 (LLM) 在内容生成领域展现出强大能力, 其在自动化命题方面的潜力已得到初步验证^[3,4], 这为智能出题提供了新思路。现有的 LLM 均能快速生成试题, 然而实践中发现, 单一 LLM 可能会生成与现有试题相似的题目, 也容易生成可读性差的试题^[5], 且其缺乏独立质检环节, 试题可用性难以得到保证。相关研究也指出, 大语言模型在辅助命题时仍存在不可控因素, 人工审核不可或缺^[6]。对此, 一些系统引入协同优化机制以提高 AI 应用上的准确性和实用^[5,7]。本文基于 COZE 智能体平台, 设计并实现了一个基于双 LLM 校验机制的 C 语言智能出题系统, 实现生成与审核分离, 确保试题质量可控。

一、相关技术基础

COZE 是字节跳动推出的一站式 AI 智能体开发与应用平台, 支持可视化 workflow 配置、知识库管理、大模型调用、插件扩展、

多渠道发布等, 满足深度定制需求, 用户无需编写复杂代码即可在该平台上快速搭建领域应用。现有的大语言模型大多已具备强大的自然语言理解与生成能力, 既可用于创造性试题生成, 也可用于对新试题的独立审核。因此, 本系统引入两个模型分任两种

基金项目: 2025 韩山师范学院教学质量与教学改革工程项目 (粤韩师教 [2025] 126 号 -59); 广东省示范性现代产业学院 -- 软件与智能物联产业学院; 2023 广东省教育厅科研项目 “重点领域专项” (2023ZDZX1014)。韩山师范学院 2025 年度教授、博士启动项目 (第二批) (QD2025202)。

作者简介:

江玉珍 (1977-), 女, 韩山师范学院计算机与信息工程学院副教授, 硕士, 研究方向: 深度学习、图形图像处理; 邮箱: jiangyz@hstc.edu.cn;

朱映辉 (1977-), 男, 韩山师范学院计算机与信息工程学院教授, 硕士, 研究方向: 人工智能、对抗样本。

角色，实现出题与审核分离，旨在提升出题质量。

二、系统设计

（一）总体架构

系统采用分层架构：设置知识库层用于存储已有优质试题；搭建两个工作流层分别实现题库检索（search_question）与大模型新题生成（generate_test）两种功能；交互层解析用户需求并输出结果。系统定位为资深 C 语言出题人，参考教材是谭浩强著的《C 语言程序设计（第5版）》，试题指定为 C99 及以后版本，支持6类题型（选择、填空、判断、看程序写结果、程序完形填空、编程题），所有题目分5级难度，覆盖语法、数组、指针、函数、结构体等25个知识点。

（二）知识库设计

知识库包含题型、知识点、难度、题干、答案5个字段，原有试题以 Excel 格式批量导入新建知识库 C_test，支持以知识点、题型、难度为检索关键词，能确保精准匹配，也支持模糊检索。COZE 知识库支持后续的补充及更改。

（三）相关工作流

1.search_question 工作流：接收知识点、题型（选用）、难度（选用）及题量4个参数，在知识库中进行精确检索，检索结果直接展示并存入变量 {{current_set}}，供后续去重校验使用。{{current_set}} 代表了本次生成的题目集。由于入库试题均已经过人工审核，检索结果可直接使用。search_question 工作流的架构图如图1所示。

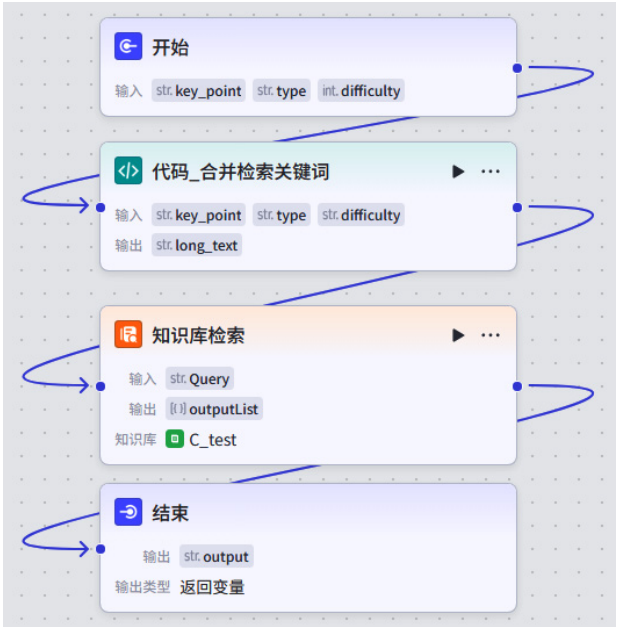


图1：search_question 工作流

2.generate_test 工作流：采用双 LLM 校验机制，两个大模型使用不同系列模型，一个作为出题 LLM，另一个作为审核 LLM。该工作流首先调用出题 LLM，根据参数输出符合要求的

试题，随后调用审核模型对刚生成的新题进行0–100分的独立评分，并将分数存入变量 {{score}} 中。评分维度主要包括质量维度（代码正确性、难度匹配度、表述清晰度）和创新维度（与 {{current_set}} 的相似度）。审核 LLM 的评分标准如表1所示，评分 ≥ 80 分视为合格，合格题目将以对话内容输出并更新至变量 {{current_set}} 之中；评分 <80 则触发循环重新生成试题。generate_test 工作流的架构图如图2所示。多模型协同驱动的思路在教育智能体领域已有初步探索^[9]，本文在此基础上，将这一思想引入 C 语言出题场景，通过“出题 – 挑错”分离机制，确保每道试题经过独立质检，避免单一模型的自我确认偏差。

表1：审核 LLM 的评分标准

维度	子指标	分值	扣分条件
题目正确性	代码语法正确、逻辑正确、边界处理正确等	40 分	有明显错误导致该题不可用的全扣（该维度得0分）；多余表述等不影响题目使用每处扣10分
创新性	与已有试题相似度	20 分	相似度 >30% 扣10分；>50% 全扣（该维度得0分）
表述清晰度	无歧义、格式规范	20 分	每处模糊、造成理解困难扣10分
难度匹配度	与指定级别相符	20 分	偏差1级扣10分

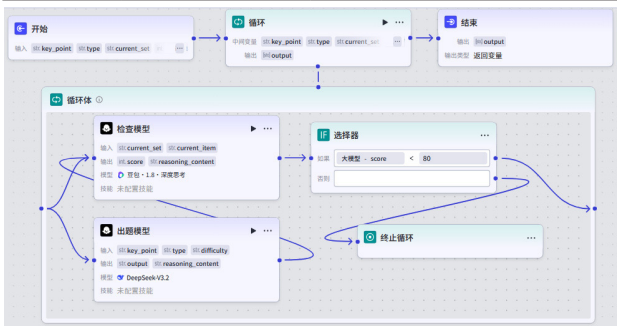


图2：generate_test 工作流

（四）自动召回与弹性补偿机制

选择智能体作为出题系统的优势是支持自然语言指令并能自动、灵活地处理。在知识库检索应用中，COZE 智能体提供了两种对用户非常友好的机制：自动召回机制与弹性补偿机制。

1. 自动召回机制是一种默认开启的“自动查知识库”功能。当需要做知识库检索时，系统首先启动“自动召回”对知识库进行简单相似度匹配，这是一种粗糙又快速的检索方法。只有当“自动召回”不能获得有效知识库反馈时，才启动 search_question 工作流进行精细检索。对于数据庞大的题库而言，“自动召回”大多情况下均能匹配到试题，即“自动召回”的响应成功率高，因此能有效地提高系统对知识库的检索效率。

2. 弹性补偿机制是一种动态资源调度策略，用于解决知识库检索结果不足时的资源缺口问题。对于出题系统而言，当某类题知识库检索出的数量不能满足需求时，系统会自动计算缺口。例如，需求5道题库试题和5道 AI 新题，但系统在知识库中仅找到3道，缺口为2道，则智能体将自动将差额转由 AI 生成补充，即

search_question workflows停止后会自动启动 generate_test workflows生成7道新题，使总题数仍为10道。弹性补偿机制实现了检索与生成的动态协同，保障了用户的特定用题需求。

三、实验与分析

以下通过3类实验数据验证该出题系统的实用性和有效性。

实验一：不同出题方式的综合评估

该实验按全知识库、全 AI 新题、知识库 +AI 新题，弹性补偿的知识库 +AI 新题4种方式测试出题效果，并从出题效率、试题质量（按表1的评分标准）两个维度对系统进行综合评估。测试计算机为 Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz，内存8.00 GB。COZE 智能体的知识库预先导入了524道优质 C 语言试题，涵盖6类题型、1至5级难度及各类知识点。针对10道指定类型题的生成任务，经16次试验，4种生成方式的平均出题时间和质量分均值如表2所示。

表2：不同出题方式的综合评估		
出题方式	生成时间	质量分均值
全知识库（含自动召回）	9.5s	97.1
全 AI 生题	83.7s	93.5
知识库 +AI 生题（各50%）	72.1s	96.8
弹性补偿的知识库 +AI 生题	91.8s	94.3

实验二：双 LLM 低分题分析

为验证双 LLM 系统的特性，在 AI 生题过程中提取部分 <80 的题目（即被弃用的题目），并结合审核 LLM 的评分进行分析。以下为两道错题的代表案例。

案例1：以下对二维数组 a 的初始化语句中，错误的是（ ）

- A. int a[2][3] = {{1,2}, {3,4}};
- B. int a[][3] = {1,2,3,4,5,6};
- C. int a[2][3] = {1,2,3,4,5,6,7};
- D. int a[2][] = {1,2,3,4,5,6};

答案 D

提取审核 LLM 分析：该题为数组类难度2级单项选择题，score 值为55，低分原因是题目存在两个错误选项（C 和 D），C 选项初始化值个数超过数组元素个数，编译无法通过；D 选项二维数组定义时第二维长度不可缺省。题目所给答案为 D，但 C 选项也明显错误。这类题目在出题上存在明显问题，不可使用。

案例2：分析以下程序，最终输出结果是（ ）

```
#include
int main()
{   int arr[5] = {1, 2, 3, 4, 5};
    int *p = arr;
    *(p++) += 12;
    *++p = *(p + 2);
```

```
printf( "%d %d\n" , arr[2], p[2]);
return 0;
}
```

A.33 B.54 C.35 D.55

答案 D

提取审核 LLM 分析：该题为指针类难度3级单项选择题，score 值为72，低分原因是 “*++p = *(p + 2);” 中，p 被多次修改和多次使用，存在未定义行为（UB），不同编译器结果可能不同。C 语言标准规定：在同一表达式中，如果对一个变量既有修改又有读取，且读取不是为了计算新值，则行为未定义。虽然在大多数编译器中可能得到 D 选项的结果，但这种一行代码完成过多操作的方式存在刻意复杂化嫌疑，实际工程中不推荐这种写法，不建议作为试题使用。

以上案例说明审核 LLM 能细致地发现出题 LLM 所给题目中不易觉察的错误，能较合理地对题目进行评分，并有效拦截不可用题目。

实验三：单 LLM 与双 LLM 出题效果比较

为验证双 LLM 相较于单 LLM 在出题上的优势，本实验参照 generate_test workflows架构模式，分别创建了无审核的单 LLM 出题工作流和自审核的单 LLM 出题工作流。三组方案各生成100道题目，每组方案的题目要求相同，合格阈值仍按表1评分标准执行，所得数据如表3所示。

表3：单 LLM 与双 LLM 出题效果对比			
方案	合格率	质量分均值	重生成次数
单 LLM（无审核）	89%	85.3	0
单 LLM（自审）	94%	91.0	8
双 LLM（本文方案）	100%	95.1	14

上述数据说明，单 LLM 方案无论是否带有自审机制，其生题时均有一定概率产出不可用但容易被忽视的题目，即单 LLM 存在审核缺乏或审核不力的问题。即使带有自审机制的单 LLM，其自我审核仍存在一定的不可靠性，其根源在于目前大多数 AI 存在“确认偏误”，即模型倾向于认可自己生成的内容。而引入独立的审核模型后，就能明显规避这一问题。这也是本文在出题系统工作流中引入不同类双 LLM 的原因：利用不同模型间的能力互补，提升出题与审核效果。

四、结论与展望

本文设计了基于双 LLM 校验机制的 C 语言出题智能体，该系统通过生成与审核分离、设置质量门禁、同批次去重等策略，有效解决了传统出题模式的效率与质量问题。在该系统中，生成与审核分离机制确保了两个 LLM 各司其职——生成模型专注于创造性内容产出，审核模型则严格把控质量，二者形成良性互补；质量门禁的设置通过多维度评分体系对试题进行筛选，只有达到预

设标准的题目才能存入当前题目集，从而保证了输出内容的可靠性；同批次去重策略则避免了相似或重复试题的冗余出现，提升了题库的整体质量与多样性。

未来工作主要包括：细化审核评分维度，引入区分度等教育测量学指标，构建更加科学完善的试题评价体系；扩展知识库规

模与覆盖范围，提升系统的通用性与专业深度；支持更复杂的用户需求，如智能批量组卷、试卷难度梯度调控等高级功能，进一步提升系统的智能化水平与用户体验，使其成为真正赋能教育场景的智能化出题助手。

参考文献

[1] 刘阳, 陈峰. C 语言程序设计课程教学改革探索 [J]. 安徽工业大学学报 (社会科学版), 2025, 42(03): 55-57.

[2] 赵雪梅, 张宏, 王如刚, 等. 互联网 + 教学模式下 C 程序设计题库平台建设 [J]. 电脑知识与技术, 2023, 19(12): 177-180.

[3] 陈大建, 胡杰辉. 基于大语言模型的自动化命题研究 —— 以英语阅读理解试题为例 [J]. 外语教学, 2025, 46(02): 40-47.

[4] 杨志明, 徐庆树, 祁长生. ChatGPT 命题潜力的实证研究 [J]. 教育测量与评价, 2023(04): 3-13.

[5] 景宏伟, 徐娟. 三类大语言模型自动出题的效能测评 —— 以 HSK6 级阅读理解题为例 [J]. 考试研究, 2026, 22(01): 44-53.

[6] 李峰, 郭嘉悦, 胡新雨, 等. 大语言模型辅助情境化命题模式探索 —— 以创造性思维测评为例 [J]. 中国考试, 2025(09): 76-86.

[7] 师曼飞, 钱玉航, 杨淑琪, 等. 知识库与多模型协同驱动的循证护理知识问答智能体的构建与评价研究 [J]. 中华护理教育, 2025, 22(09): 1036-1042.

[8] 刘嘉, 李楠, 梁晓娜, 等. 基于多智能体系统的中文学术问答与应用 [J]. 昆明理工大学学报 (自然科学版), 2025, 50(03): 58-69.

融合 BERT 语义表征与残差特征增强的 O2O 食品安全风险监测研究

金思彤, 李建波*

江苏师范大学数学与统计学院, 江苏 徐州 221116

DOI:10.61369/ASDS.2026060009

摘 要 : 随着 O2O 餐饮模式的快速发展, 用户评论已成为识别食品安全隐患的关键数据源。然而, O2O 评论存在明显的口语化、碎片化与语义模糊等特点, 传统机器学习及单一深度学习模型在微调过程中难以兼顾全局语义理解与局部细粒度特征的精准提取, 易造成食品安全风险信号漏检。针对上述挑战, 本文提出一种融合 BERT 预训练语言模型与残差网络的食品安全风险识别模型。该模型首先利用 BERT 获取文本深层双向语义表示, 构建全局特征基准; 随后通过堆叠残差块对隐层状态序列进行多尺度局部特征挖掘, 借助跳跃连接缓解深层网络微调时的梯度退化问题, 增强模型对“异物”、“发臭”、“致病”等风险关键词的感知能力; 最后采用特征拼接策略融合全局与局部特征向量, 实现风险评论的自动判别。基于 CCF BDCI 提供的 10000 条真实 O2O 评论数据集进行实验。结果表明: 本文提出的 BERT + Residual Block 模型在验证集上 F1 值最高可达 0.9257, 整体 F1 值为 0.9096, 相较于 CNN + BERT 和 BiLSTM + BERT 模型更具稳定性; 测试集全量准确率达到 97.5%, 五折交叉验证的标准差仅为 0.0062, 展现出较强泛化能力与鲁棒性; 模型在风险类别的 F1 分数达到 0.92, 可有效降低智慧监管场景下的风险漏检率。本文验证了残差结构在提升预训练模型特征提取能力的有效性, 为 O2O 场景下食品安全数字化监测与预警提供一种高效技术方案。未来研究将进一步引入外部知识图谱注入与多模态融合技术, 增强模型对复杂反讽语义的识别能力。

关 键 词 : 食品安全监测; O2O 评论分析; BERT; 残差网络; 文本分类

Research on O2O Food Safety Risk Monitoring Based on BERT Semantic Representation and Residual Feature Enhancement

Jin Sitong, Li Jianbo*

School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou, Jiangsu 221116

Abstract : With the rapid development of the O2O catering model, user reviews have become a key data source for identifying food safety hazards. However, O2O reviews have obvious characteristics of colloquialism, fragmentation, and semantic ambiguity. Traditional machine learning and single deep learning models are difficult to balance global semantic understanding and accurate extraction of local fine-grained features during fine-tuning, which easily leads to missed detection of food safety risk signals. To address the above challenges, this paper proposes a food safety risk identification model integrating the BERT pre-trained language model and residual network. The model first uses BERT to obtain deep bidirectional semantic representations of text and construct a global feature benchmark; then, it conducts multi-scale local feature mining on the hidden state sequence through stacked residual blocks, alleviates the gradient degradation problem during fine-tuning of deep networks with skip connections, and enhances the model's ability to perceive risk keywords such as "foreign objects", "odor", and "pathogenic"; finally, it adopts a feature concatenation strategy to fuse global and local feature vectors to realize automatic discrimination of risky reviews. Experiments were conducted based on 10,000 real O2O review datasets provided by CCF BDCI. The results show that the BERT + Residual Block model proposed in this paper achieves a maximum F1 score of 0.9257 on the validation set and an overall F1 score of 0.9096, which is more stable than the CNN+BERT and BiLSTM+BERT models; the full-scale accuracy of the test set reaches 97.5%, and the standard deviation of five-fold cross-validation is only 0.0062, showing strong generalization ability and robustness; the F1 score of the model in risk categories reaches 0.92, which can effectively reduce the

基金项目: 江苏省研究生科研训练项目 (SJCX25_1499); 江苏省学位与研究生教育教学改革课题 (JGKT25_C075); 江苏师范大学研究生课程思政教学研究示范中心项目。

作者简介: 金思彤, 江苏师范大学应用统计2023级硕士研究生, 研究方向为医学统计; Email:19895248863@163.com。

通讯作者: 李建波, 江苏师范大学数学与统计学院教授, 统计学博士, 研究领域包括生存分析、应用统计、统计学习等; Email: lijianbo@jsnu.edu.cn。

risk missed detection rate in intelligent supervision scenarios. This paper verifies the effectiveness of the residual structure in improving the feature extraction ability of pre-trained models, and provides an efficient technical solution for digital monitoring and early warning of food safety in O2O scenarios. Future research will further introduce external knowledge graph injection and multi-modal fusion technologies to enhance the model's ability to identify complex ironic semantics.

Keywords : food safety monitoring; O2O review analysis; BERT; residual block; text classification

引言

随着4G/5G网络的广泛覆盖和智能终端的普及,O2O(Online To Offline)商业模式已深刻改变消费者的餐饮消费习惯。外卖平台、到店团购等服务的快速兴起,使消费者能便捷获取线上优惠并享受线下服务^[1]。然而,商户数量的激增与准入门槛参差不齐,给食品安全监管带来了前所未有的挑战。食品配送过程中的温控不当、包装污染、商家卫生条件不达标等安全隐患,仅依靠传统的抽样检测和投诉举报机制,难以实现全面、高效的监管覆盖^[2]。用户评论作为消费者消费体验的直接反馈,蕴含着丰富的食品质量、卫生状况、配送服务等相关信息,已成为食品安全风险监测领域的重要数据源^[3]。

如何从海量非结构化的评论文本中快速、准确地识别出与食品安全相关的风险信号,成为自然语言处理与智慧监管交叉领域的研究热点。情感分析技术是处理此类问题的核心手段。早期研究多依赖于传统的机器学习算法,如支持向量机(SVM)、朴素贝叶斯等,但这类方法高度依赖人工特征工程,难以有效处理文本中的复杂语义歧义和上下文关联信息^[4]。随着深度学习的发展,卷积神经网络(CNN)和循环神经网络(RNN)被广泛应用于文本分类,它们能够自动提取局部特征和序列信息,显著提升文本分类性能^[5,6]。

2018年,Google提出的BERT(Bidirectional Encoder Representations from Transformers)模型,凭借双向Transformer架构,在超大规模语料上完成预训练,实现深层语义表示的突破,刷新了多项自然语言处理任务的性能记录^[7]。BERT通过自注意力机制捕捉词语间的全局依赖关系,其输出的[CLS]向量可作为整个句子的语义表征,直接应用于下游分类任务。然而,在食品安全这一垂直领域中,单一BERT模型在微调过程中,往往难以充分捕捉特定语境下的细粒度风险特征,例如“有虫子”、“发臭”等具有明确风险指向的局部关键短语。此外,随着网络深度的增加,模型在训练过程中还可能出现梯度不稳定、语义信息丢失等问题,影响模型性能。

残差网络(Residual Network)由He等人于2016年提出,通过引入跳跃连接(skip connection)有效解决了深层网络训练中的梯度消失和性能退化问题^[8]。近年来,残差结构被逐步应用于文本分类领域,用于增强模型的局部特征提取能力和训练稳定性。将残差模块与预训练语言模型相结合,有望在保留BERT模型全局语义理解优势的基础上,进一步强化模型对文本中关键语义片段的感知能力,从而提升风险识别准确性。

针对上述研究痛点与挑战,本文提出一种融合BERT与残差块(Residual Block)的食品安全评论监测模型。该模型首先利用BERT模型获取评论文本的深层语义表示,将其作为全局特征;随后通过堆叠的残差块,对BERT输出的隐层状态进行细粒度局部特征提取;最后将全局特征与局部特征进行融合,输入分类层完成风险评论判别。本文的主要贡献如下:(1)提出一种新颖的BERT+Residual Block混合架构,详细阐述其设计原理、结构组成及参数配置方案;(2)在公开的O2O食品安全评论数据集上开展系统的实验对比,验证所提模型的有效性与优越性;(3)通过参数敏感性分析、五折交叉验证及统计检验,全面评估残差块数量等超参数对模型性能的影响;(4)为O2O环境下的食品安全风险预警工作,提供一种高效、可靠的数据驱动型技术解决方案。

本文后续结构安排如下:第一章介绍数据来源与研究方法,包括数据集详细描述、模型评价指标、BERT与残差网络的核心原理,以及本文所提模型的具体构建细节;第二章汇报实验结果与分析,涵盖基线模型对比、超参数影响分析、交叉验证结果及显著性检验;第三章将本文方法与多种经典文本分类方法进行系统对比,进一步凸显所提模型的优势;第四章总结全文研究成果,并对未来研究方向进行展望。

一、数据来源与研究方法

(一)数据来源

本文采用的数据集来源于2019年中国计算机学会(CCF)大数据与计算智能大赛(CCF BDCI)“O2O店铺食品安全评论识别”赛题。该数据集由O2O电商平台真实用户评论构成,具有文本非结构化程度高、口语化特征突出、垂直领域专业性强等特

点,能够客观反映消费者在餐饮消费过程中对食品安全风险的真实反馈情况。

1. 数据规模与分布特征

该数据集共包含已标注样本10000条,每条评论均依据是否包含食品安全隐患(如食物变质、异物、致病反应等)标注为二元标签。经统计,正例样本(与食品安全风险相关)共4212条,占总样本量的42.12%;负例样本(与食品安全风险无关)共5788

条，占总样本量的57.88%。正负样本比例约为1:1.37，类别分布相对均衡，可有效消除极端类别不平衡可能导致的模型判别偏置问题。部分评论数据示例见表1。样本细分领域涵盖食品加工质量、配送卫生环境及商户经营资质等多个维度，为构建多场景下的食品安全风险识别模型提供了坚实的数据支撑。

表1：部分评论数据示例

ID	评论文本	标签
01f91684-d6a2-4641-b620-f0394b91ea24	里面有虫子 太恶心了 吐死了	1
020bf65e-e09a-4036-84a0-f4991112482d	炒苕尖太难吃了又老基本上没有吃	1
023d2dcc-45b1-4140-920f-0a5c3163f0b6	奶茶不错，鸡蛋仔要现做，需要等，脆脆的，很好吃	0
ffcc4330-2b02-485b-a3bb-e1c7d42baaae	今天的怎么是臭的?! 都是老客户了，以后不会再来!!	1
fff44f7a-1cef-4b7a-a2e5-1d2a111cb70d	速度快，价格便宜，味道很好。很不错	0

2. 文本长度统计分析

为了科学设定模型的输入维度，本文对10000条样本的字符长度进行了探索性统计。

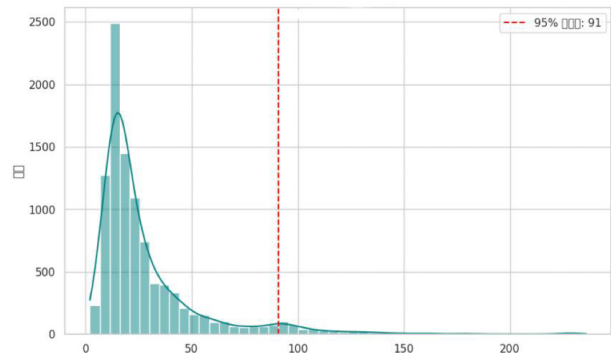


图1：集中趋势与分布形态

如图1所示，评论长度呈现显著的右偏分布（Right-skewed Distribution）。样本的平均长度为47.3个字符，中位数为38个字符，最长评论为186个字符，最短评论为4个字符。统计结果显示，长度在128个字符以内的样本占比超过95%。基于此，本文将模型输入的最大序列长度统一设定为128，该策略既能确保绝大部分评论的语义信息完整保留，又能有效控制计算开销，提升模型训练效率。

3. 数据预处理与样本划分

为确保模型评估的稳健性与科研结果的可重复性，本文执行了严谨的预处理流程，并采用科学的样本划分策略。样本划分方面，采用分层随机抽样方法，以8:2的比例将原始10000条数据划分为训练集（8000条）与独立测试集（2000条）。该划分方式可确保训练集与测试集的类别分布均与全量数据保持一致，有效避免样本划分偏差对模型评估结果的影响。利用BERT自带的WordPiece分词器^[9]进行Token化处理。

该分词机制通过子词拆分策略，有效缓解了中文文本中未登录词（Out-of-Vocabulary, OOV）的识别难题。针对前文设定的128个字符最大序列长度，对不足128位的序列进行补齐

（Padding）处理，同时生成对应的注意力掩码（attention mask）矩阵，引导模型在训练过程中精准聚焦于有效语义片段，忽略冗余的填充信息，进一步提升模型训练的精准度与效率。

（二）研究方法

1. 评价指标的定义

在二分类任务中，常用的评价指标包括准确率（Accuracy）、精确率（Precision）、召回率（Recall）和F1分数。考虑到食品安全监测任务中，“漏报”（将风险样本识别为正常）的代价远高于“误报”，且数据集中正负样本存在一定比例差异，本文选取F1分数作为核心评价指标。F1分数是精确率与召回率的调和平均，能够综合衡量模型在不平衡数据集下的判别稳健性，更贴合食品安全监测实际需求。

相关指标基于混淆矩阵定义如下：（1）TP（True Positive）：实际为正且被正确预测为正的样本数。（2）FP（False Positive）：实际为负但被错误预测为正的样本数。（3）FN（False Negative）：实际为正但被错误预测为负的样本数。

精确率、召回率和F1分数的具体计算公式为：

$$Precision = \frac{TP}{TP + FP}, \quad (1)$$

$$Recall = \frac{TP}{TP + FN}, \quad (2)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \quad (3)$$

2.BERT 模型原理

BERT 是基于Transformer编码器架构的深度预训练语言模型。该模型通过在海量无标注语料上交替执行掩码语言模型（Masked Language Model, MLM）与下一句预测（Next Sentence Prediction, NSP）两项任务，能够学习到兼顾左右上下文信息的深层双向语义表征，有效解决了传统单向语言模型难以捕捉上下文关联的问题。其架构核心由多头自注意力机制（Multi-Head Self-Attention）与逐位置前馈神经网络（Position-wise Feed-Forward Network, FFN）构成。

自注意力机制是BERT模型的核心，允许模型动态计算序列内各Token间的语义关联度，实现对文本全局语义的精准捕捉。其计算公式为：

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (4)$$

其中， Q (Query)、 V (Value)、 K (Key) 矩阵分别由输入嵌入矩阵与对应的权重矩阵相乘得到， d_k 为键向量的维度。缩放点积（Scaled Dot-Product）不仅衡量了序列中位置间的相关性，还通过根号项缓解了因向量维度过高导致的Softmax函数饱和、梯度失效问题，保障了深层网络的训练稳定性。

为使模型能够从多个语义子空间（如语法结构、指代关系、语义情感等）提取文本特征，BERT采用多头注意力机制，将多个单头注意力的输出进行拼接后，通过线性变换得到最终的注意

力特征，其计算公式为：

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O, \quad (5)$$

其中， $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$ 。此外，此外，Transformer 编码器的每一层均集成了残差连接与层归一化（Layer Normalization）操作，进一步缓解了深层网络训练中的梯度消失问题，确保模型训练的稳定性。

本文采用 BERT-base 模型作为基础预训练模型，该模型由 12 层 Transformer 编码器堆叠而成，隐藏层维度为 768，注意力头数为 12。在处理分类任务时，通常提取序列起始位置的特殊标记 CLS 对应的向量 $h_{cls} \in \mathbb{R}^{768}$ 作为全文的语义总结向量，将其输入下游分类层，实现对评论样本的风险判别。

3. 基线模型原理

为客观验证本文提出的 BERT + Residual Block 模型的优越性，研究选取了以下五种具有代表性的基线模型进行性能对标，涵盖了从传统统计学习到前沿深度学习的技术演进路径，确保对比实验的全面性与合理性：

（1）SVM + TF-IDF：传统机器学习基准。通过 TF-IDF 统计词频 - 逆文档频率特征，构建高维稀疏特征向量，利用支持向量机寻找最大间隔超平面进行分类^[10]。

（2）XGBoost + TF-IDF：集成学习基准。采用梯度提升决策树框架处理统计特征，利用其高效的迭代能力和鲁棒性处理非结构化文本数据^[11]。

（3）Vanilla BERT：纯预训练模型基准。直接提取预训练模型 BERT-base-chinese 的首位标记 CLS 向量作为全句语义总结，仅接全连接层完成分类^[12]。

（4）CNN + BERT：局部特征增强基准。在 BERT 编码层后堆叠一维卷积神经网络，通过卷积核捕获文本中关键的局部 n-gram 特征模式^[13]。

（5）BiLSTM + BERT：序列建模基准。在 BERT 编码层后引入双向长短期记忆网络，利用其门控机制捕获文本的长距离双向上下文依赖关系^[14]。

4. 残差网络原理

残差网络（Residual Network, ResNet）由 He 等人提出，主要用于克服深层神经网络在训练过程中出现的梯度消失与性能退化问题^[15]。ResNet 的核心思想是引入跳跃连接（shortcut connection），将网络目标从学习原始映射 $H(x)$ 转变为学习残差映射 $F(x)=H(x)-x$ ，有效降低深层网络的优化难度。基本残差块的形式定义为：

$$y = F(x, \{W_i\}) + x, \quad (6)$$

其中， x 为残差块的输入， $F(x)$ 代表包含卷积层、批量归一化（BN）及非线性激活（ReLU）的复合变换。跳跃连接通过保留恒等映射，确保梯度在反向传播过程中能够跨层直接流向浅层，有效缓解深层网络的优化难度，避免梯度消失问题。

对于多层堆叠的残差网络，从浅层 l 到深层 L 的前向传播可表示为：

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i). \quad (7)$$

对应的反向传播过程中，损失函数对 x_l 的梯度为：

$$\frac{\partial L}{\partial x_l} = \frac{\partial L}{\partial x_L} \left(1 + \frac{\partial}{\partial x_l} \sum_{i=l}^{L-1} F(x_i, W_i) \right). \quad (8)$$

由于公式中存在常数项“1”，即便残差映射的偏导数较小，总梯度也不会趋于消失，从而保证了超深层网络训练的可行性。

在本文模型中，残差网络将 BERT 输出的隐层序列 $H \in \mathbb{R}^{n \times d}$ 视为具备局部语义信息的特征图，通过一维卷积残差块提取文本中连续的 n-gram 特征，并利用跳跃连接融合 BERT 模型学习到的全局语义背景，实现全局与局部多尺度特征的协同感知，显著提升食品安全评论的风险识别性能，有效解决单一 BERT 模型局部特征提取不足的问题。

5. BERT + Residual Block 模型构建

为精准识别 O2O 平台评论中的食品安全风险，本文提出一种融合 BERT 语义表征与残差网络特征增强的文本分类架构。该模型利用 BERT 捕捉文本的宏观语境语义，通过残差结构深层次挖掘局部敏感风险特征（如“异物”、“腹泻”等关键风险信号），最终通过多层感知器实现特征汇聚与风险评论分类，有效弥补单一模型在全局与局部特征兼顾上的不足。

整体模型结构如图 2 所示，主要涵盖五大模块：文本输入表示层、BERT 语义编码层、Residual Block 局部特征提取层、特征融合层及分类层，各模块协同作用，实现食品安全风险评论的精准判别。

（1）文本输入表示层

在深度学习文本分类任务中，将自然语言文本转化为计算机可识别的高维连续向量是首要环节。传统 TF-IDF 或词袋模型仅能捕捉词频信息，难以有效捕捉词序关系及深层语义关联，适配性较差。本文采用 BERT 专用的 WordPiece 分词技术，将原始序列 x 映射为 Token 序列。每个词元的输入表征 E_i 由三部分加和构成^[16]：

$$E_i = T_i + S_i + P_i, \quad (9)$$

其中， T_i 、 S_i 、 P_i 分别代表词嵌入（Token Embedding）、句子嵌入（Segment Embedding）及位置嵌入（Position Embedding）。这种多维度融合的表征方式，确保模型在处理口语化、碎片化的 O2O 评论文本时，能够有效建模文本位置信息与句子结构，为后续语义编码奠定基础。

（2）BERT 语义编码

将预处理后的文本 Token 序列输入 BERT 预训练模型进行语义编码，本文提取 BERT 模型最后一层的隐藏状态序列 $H \in \mathbb{R}^{768}$ （其中 $n=128$ ），该序列包含了评论文本丰富的双向上下文语义信息。同时，提取序列起始标记 [CLS] 对应的输出向量 pooler_output 作为整个句子的全局语义表征，记为 $h_{cls} \in \mathbb{R}^{768}$ 。

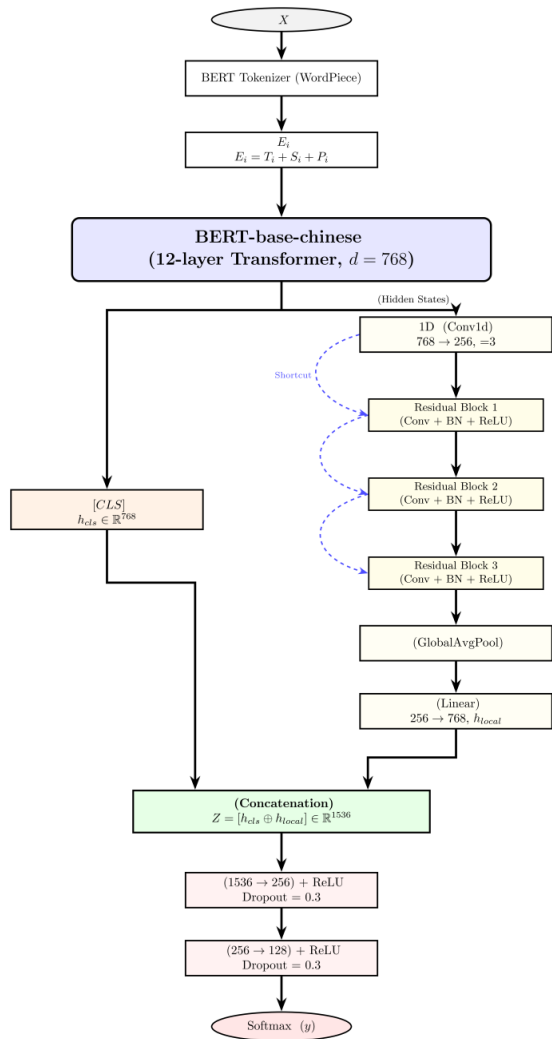


图2: BERT + Residual Block 模型

(3) Residual Block 局部特征提取层

为强化模型对评论文本中局部关键风险短语的识别能力, 弥补 BERT 模型在细粒度特征提取上的不足, 本文引入残差网络对 BERT 输出的隐藏层序列 H 进行进一步处理。首先, 利用一维卷积层将特征维度从 768 映射到 256, 以减少残差计算过程中的冗余:

$$H' = \text{Conv1d}(H) \in R^{n \times 256}. \quad (10)$$

随后, 将维度转换后的特征序列 H' 输入 k 个堆叠的残差块 (本文通过参数敏感性实验最终确定最优配置为 $k=4$)。每个残差块通过卷积操作与跳跃连接相结合的方式, 实现对局部语义信息的深层建模, 其表达式为:

$$H^{(l+1)} = H^{(l)} + F(H^{(l)}), \quad (11)$$

其中, $H^{(l)}$ 为第 l 层残差块的输入特征, $H^{(l+1)}$ 为输出特征。经过 k 个残差块处理后, 得到残差增强后的特征序列 $H_r \in R^{n \times 256}$ 进行全局平均池化, 并经由全连接层映射回 768 维, 得到局部特征向量 $h_{local} \in R^{768}$ 。

(4) 特征融合与分类层

为充分利用全局语义信息与局部风险特征, 本文采用向量拼接策略, 将 BERT 提取的全局语义向量 h_{cls} 与局部特征 h_{local} , 构建复合特征向量 Z :

$$Z = \text{concat}(h_{cls}, h_{local}) \in R^{1536}. \quad (12)$$

融合后的特征向量 Z 依次通过两个带有 Dropout 层的全连接层进行非线性转换, 缓解模型过拟合问题, 最终通过 Softmax 函数输出类别概率分布 y , 实现风险评论与非风险评论的二分类:

$$y = \text{softmax}(W_2 \times \text{ReLU}(W_1 Z + b_1) + b_2), \quad (13)$$

其中, W_1 和 b_1 为第一层权重和偏置, W_2 和 b_2 为第二次权重和偏置, y 为预测类别概率。

(5) 模型参数配置

训练过程采用 AdamW 优化器, 学习率设为 $3e-5$ (基于参数敏感性实验确定的最优值), 批次大小为 32, 训练轮数为 10。损失函数采用交叉熵损失 (Cross-Entropy Loss), 用于衡量模型预测概率与真实标签之间的差异, 其计算公式为:

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^2 y_{i,c} \log p_{i,c}, \quad (14)$$

其中 N 为样本数, $y_{i,c}$ 为真实标签的 one-hot 编码, $p_{i,c}$ 为模型预测概率。表 2 总结了 BERT + Residual Block 模型各层的详细参数配置, 确保实验的可重复性。

表 2: BERT + Residual Block 模型详细结构

模块	层类型	输入维度	输出维度	关键参数 / 说明
输入层	BERT Tokenizer	原始文本	(1, 128)	Max_len = 128
语义编码	BERT-base-chinese	(1, 128)	(128, 768)	12 层 Transformer
全局特征	[CLS] Pooling	(128, 768)	(768)	提取 CLS
局部特征	1D Convolution	(128, 768)	(128, 256)	核大小 = 3, 维度对齐
	Residual Blocks	(128, 256)	(128, 256)	堆叠数量 $k=4$
	GlobalAvgPool1d	(128, 256)	(256)	沿时间轴平均池化
	Linear Mapping	(256)	(768)	映射至 h_{local}
特征融合	Concatenation	(768)+(768)	(1536)	拼接全局与局部特征
分类头	Fully Connected 1	(1536)	(256)	ReLU + Dropout(0.3)
	Fully Connected 2	(256)	(128)	ReLU + Dropout(0.3)
	Output Layer	(128)	(2)	Softmax 分类输出

* 注: 模型核心超参数 (如 $k=4$ 及学习率) 均通过消融实验与敏感性分析验证。

二、研究结果

（一）实验环境与配置

为确保实验的稳定性、可重复性及基准对比的公平性，本文所有实验均基于 PyTorch 1.13.0 深度学习框架，在 Ubuntu 22.04 操作系统环境下开展。硬件核心配置为 Intel Xeon Platinum 8352V 处理器及 NVIDIA RTX 4090 GPU（显存 24GB），为深层模型训练提供充足的计算资源支撑。所有对比基线模型均在上述同一硬件环境下进行复现，消除硬件差异对实验结果的影响。

在模型训练层面，所有 BERT 相关模型均使用 bert-base-chinese 预训练权重，本文设定的核心参数如下：训练轮数设为 10，并引入早停策略（Early Stopping），以验证集 F1 分数作为权重保存的判别准则；批次大小（Batch Size）为 32；最大序列长度（Sequence Length）为 128。选用 AdamW 作为优化器。初始学习率设为 $2e-5$ ，权重衰减为 0.01，Dropout 比率为 0.3。学习率策略引入线性预热（Linear Warmup），在前 10% 的训练步数内平滑升温后进行线性衰减。

基线模型配置如下：（1）SVM + TF-IDF：设定 n-gram 范围为 (1,2)，特征维度上限为 5000，分类器选用线性核支持向量机（Linear SVC），惩罚参数 $C=1.0$ 。（2）XGBoost + TF-IDF：设定集成树数量为 100，损失函数采用对数损失（logloss）。（3）Vanilla BERT：直接提取预训练 bert-base-chinese 的 CLS 向量接全连接层。（4）CNN + BERT：提取 BERT 最后一层隐藏状态序列输入 CNN 层（维度从 768 映射至 256，卷积核尺寸为 3），通过全局最大池化（Max Pooling）提取显著局部特征模式。（5）BiLSTM + BERT：提取隐藏层序列输入双向长短期记忆网络（BiLSTM），设定隐藏层维度为 256（双向拼接后为 512），通过提取序列首位上下文向量实现特征映射与分类。

所有深度学习基准模型均在相同数据划分下独立运行 3 次循环，记录指标的平均值与标准差，以确保实验的可重复性与基准对比的公平性。

（二）基准模型对比分析

本文选取了传统机器学习方法及多种基于深度学习的 BERT 变体模型作为对照组。

表 3：基准模型对比实验结果

Model	Precision	Recall	F1-Score
SVM + TF-IDF	0.8861 ± 0.0000	0.6954 ± 0.0000	0.7792 ± 0.0000
XGBoost+ TF-IDF	0.9263 ± 0.0000	0.6656 ± 0.0000	0.7746 ± 0.0000
Vanilla BERT	0.9116 ± 0.0098	0.8974 ± 0.0047	0.9044 ± 0.0027
CNN + BERT	0.9129 ± 0.0282	0.8874 ± 0.0162	0.8994 ± 0.0073
BiLSTM + BERT	0.8889 ± 0.0058	0.9084 ± 0.0102	0.8985 ± 0.0022
Proposed (ResBlock + BERT)	0.9265 ± 0.0212	0.8940 ± 0.0162	0.9096 ± 0.0047

在测试集上对各模型进行性能评估，结果如表 3 所示（数值均为五次独立实验的均值及标准差）。本文模型采用经调优后的最优配置。

实验结果显示，所有基于 BERT 架构的深度学习模型在 F1 分数上均显著优于传统的机器学习方法（SVM 和 XGBoost）。

尽管 XGBoost 在精确率上表现出色（0.9263），但其召回率仅为 0.6656，导致 F1 分数大幅落后。这表明传统的 TF-IDF 特征工程难以捕捉 O2O 评论中复杂的语义关联和情感语境，在面对食品安全领域的隐晦风险表达时极易造成“漏报”。

在 BERT 变体模型中，Vanilla BERT 取得了 0.9044 的 F1 分数，而简单叠加 CNN 或 BiLSTM 层的模型性能反而略有波动（分别为 0.8994 和 0.8985）。这反映出在处理中等规模评论语料时，过于复杂的非残差结构可能会引入额外的噪声或面临优化困难，导致模型难以充分利用 BERT 预训练层输出的深层表征。

本文提出的 Proposed（ResBlock + BERT）模型在 F1 分数上达到了最优（0.9096）。该模型通过引入多层残差块并结合全局池化机制，有效解决了深度特征提取中的梯度稳定性问题。残差块的跳跃连接确保了 BERT 的原始全局语义与新提取的局部 n-gram 特征能够进行多尺度融合，从而在保持高精确率的同时，显著增强了对食品安全敏感信号的俘获能力。从标准差指标来看，本文模型在 F1 分数上的波动（ ± 0.0047 ）相对较小，表现出良好的训练鲁棒性。这证明了通过残差结构对局部特征进行深层挖掘，不仅提升了识别上限，也确保了模型在不同随机初始化下的性能稳定性。

（三）残差块数量影响分析

为探究残差块数量对模型性能的影响，我们在固定其他参数（卷积核大小 3，通道 256）的情况下，分别设置残差块数量 $k=1\sim5$ 进行实验，结果如表 4 所示。实验采用与本文模型相同的训练设置，在验证集上评估各配置的最佳 F1 分数。

表 4：不同残差块数量下的模型性能

残差块数量 (k)	验证集最佳 F1 分数
1	0.9022
2	0.9151
3	0.9063
4	0.9257
5	0.9094

由图 3 可见，随着 k 值由 1 增至 4，模型的特征抽象能力增强，F1 分数在 $k=4$ 时达到峰值 0.9257。当 k 继续增至 5 时，性能下降至 0.9094。原因可能在于网络过深导致过拟合，或参数过多使得优化困难。综上，本文确定 $k=4$ 作为最佳配置，实现了性能与复杂度的最优平衡。

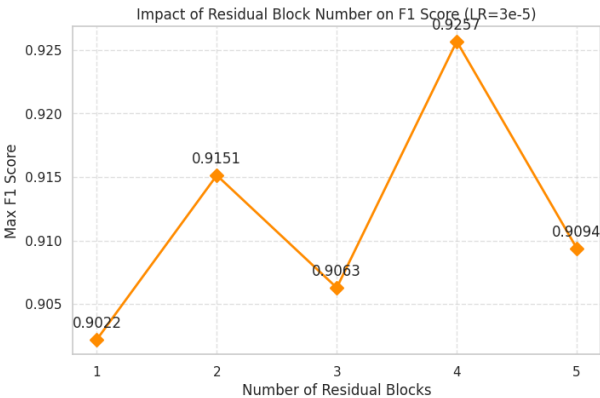


图 3：残差块数量对 F1 分数的影响折线图

（四）五折交叉验证与显著性检验

为了进一步验证模型在不同数据分布下的稳健性与泛化能力，避免因单次随机划分数据集带来的偶然性偏差，本研究在训练全集上进行了五折交叉验证，并与 CNN + BERT 基准模型进行了对比分析。五折交叉验证的详细对比结果如表5所示。

实验结果显示，本文模型在五折验证中的 F1 平均值达到 0.9000，略优于基准模型的 0.8981。更为关键的是，本文模型的标准差仅为 0.0062，远低于基准模型的 0.0177，这表明引入残差块结构后，模型对局部噪声数据的敏感度显著降低，在不同样本组合下均展现出极高的稳定性。

表5：五折交叉验证 F1 分数对比报告

Fold	CNN + BERT	Proposed Model
Fold 1	0.911290	0.901010
Fold 2	0.909091	0.907127
Fold 3	0.900000	0.888412
Fold 4	0.906445	0.902439
Fold 5	0.863544	0.901099
Mean	0.898074	0.900017
Std	0.017676	0.006217
Paired t-test: $t=0.2139, p=0.8411$		

特别是在 Fold 5 实验中，基准模型性能出现明显下滑（0.8635），而本文模型依然稳健地保持在 0.9011，这印证了多尺度特征融合机制在面对复杂语料时具有更强的纠错能力。最后，配对样本 t 检验结果显示 $t=0.2139, p=0.8411$ ，虽然在统计学意义上两模型平均性能处于同一量级，但本文模型极低的标准差意味着其在实际食品安全监测场景中具有更可预测、更可靠的风险识别表现，能有效降低监测系统的漏检波动风险。

（五）混淆矩阵与误分类分析

在 2000 条测试样本（1698 条非风险样本与 302 条风险样本）中，模型正确识别了 1950 条，整体准确率（Accuracy）达到 97.5%。详细分类指标如表6所示。

从实验结果可见，模型对非风险样本的判别极为精准（F1 = 0.99）。对于语境复杂的风险样本，F1 分数也达到了 0.92，显著降低了 O2O 食品安全监测中的虚假报警率。图4展示本文模型在测试集上预测分布情况。

表6：测试集全量分类报告

类别	精确率	召回率	F1 分数	支持数
非风险	0.98	0.99	0.99	1698
风险	0.94	0.89	0.92	302
宏观平均	0.96	0.94	0.95	2000
加权平均	0.97	0.97	0.97	2000

进一步分析图4中的误分类细节：总错误条数为 50 条，其中假正例（FP，误报）18 条，假负例（FN，漏报）32 条。这表明漏报是模型进一步优化的核心方向，说明在面对部分极隐晦的正例时，模型仍存在感知死角。

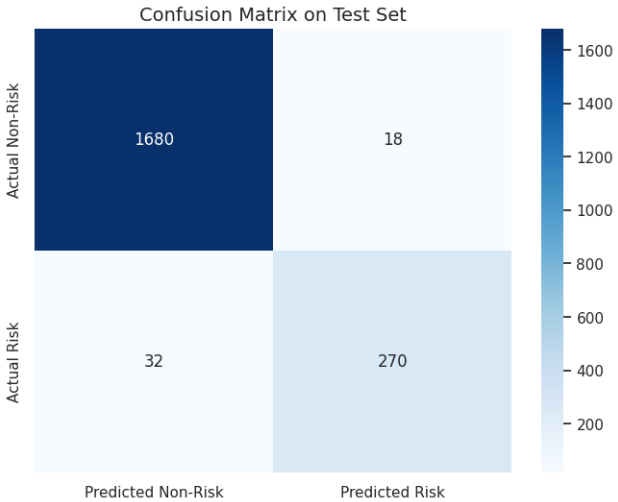


图4：测试集混淆矩阵

通过对误分类样本进行人工溯源，本文归纳出以下三类主要挑战：（1）反讽与语义偏转（占比约 40%）：如“这家的饭菜真‘干净’，吃完就拉肚子”，模型可能过度关注“干净”这一正向词汇，而忽略了后续的病理反馈。（2）极短文本与特征稀疏（占比约 30%）：如“避雷，再也不点了”，由于缺乏具体的食品卫生关键词（如“异物”、“变质”），模型难以将其与普通的口味不满区分开。（3）口语化与方言表达（占比约 20%）：如“肚子闹革命”、“拉稀摆带”等非规范表述，尽管有 BERT 的语义支撑，但在特定细分场景下的泛化能力仍有提升空间。

（六）超参数敏感性与训练动力学

为了进一步验证模型的鲁棒性并确定最佳实验配置，本研究针对学习率（Learning Rate）和批处理大小（Batch Size）进行了敏感性分析，并对模型的训练收敛过程进行了动力学监测。

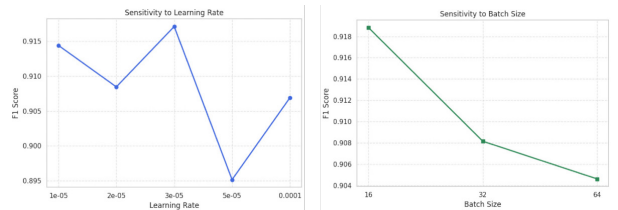


图5：模型学习率与批处理大小表现

超参数的选择直接影响 BERT 微调的稳定性与最终性能。图5展示了不同配置下模型在测试集上的表现：实验测试了从 1e-5 到 5e-5 的多个量级。结果表明，当学习率为 3e-5 时，模型达到了最优的 F1 分数（0.919）。过低的学习率会导致模型收敛于局部最优，而当学习率达到 5e-5 时，F1 分数出现明显下滑（0.898），这反映了过大的更新步长可能破坏 BERT 预训练权重的结构稳定性。如图6所示，模型性能随 Batch Size 的增大呈现下降趋势。在 Batch Size = 16 时，模型获得了最高的 F1 分数。这说明在处理 O2O 食品安全评论这类特征分布不均的数据集时，较小的批次有助于模型跳出局部鞍点，获得更好的泛化能力。

图6记录了模型在 10 个 Epoch 内的训练损失（Training

Loss) 与验证集 F1 分数 (Validation F1-Score) 的演化路径: 得益于残差连接对梯度流的优化, 训练损失在前 4 个 Epoch 内呈指数级下降, 并最终稳定在 0.02 以下。验证集 F1 分数在第 2 个 Epoch 即达到峰值 (0.915)。随后, 验证集表现出现了一定程度的波动, 这主要是由于网络评论中存在大量讽刺、口语化等高噪声样本, 导致判定边界产生细微偏移。通过引入早停机制, 本研究成功锁定了泛化性能最佳的模型权重, 有效规避了过拟合风险。

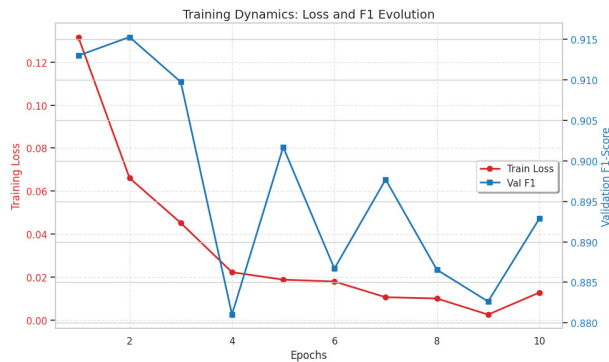


图6: 训练损失与验证 F1 曲线

三、方法对比与讨论

在本章中, 本文将所提出的 BERT+Residual Block 模型与前述基线模型及现有研究中的典型方法进行深度对比, 从特征提取能力、模型稳健性以及垂直领域适用性三个核心维度展开系统讨论, 全面验证所提模型的优越性与实用性。

(一) 特征建模能力的差异化分析

实验结果表明, 不同维度的特征处理方式对 O2O 食品安全监测效果的影响具有显著性差异, 具体分析如下:

1. 传统统计特征 vs. 深层语义表征: 以 SVM 和 XGBoost 为代表的传统机器学习方法, 主要依赖 TF-IDF 算法进行文本向量化处理。由于 O2O 平台评论文本具有极强的口语化和非规范性 (如“真‘干净’”、“肚子闹革命”), 这种基于词频统计的特征工程无法有效识别反讽语义或隐含的因果关系, 导致模型召回率普遍低于 0.7。从表 3 可见, SVM 和 XGBoost 的召回率分别仅为 0.6954 和 0.6656, F1 分数均不足 0.78。

相比之下, BERT 架构通过双向 Transformer 编码器捕捉文本全局语境, 显著提升了语义理解的深度与准确性, 使得模型 F1 分数直接从 0.77 区间跨越至 0.90 区间。这一跃升充分印证了预训练语言模型在理解复杂口语化表达、挖掘隐含语义信息上的核心优势。

2. 浅层结构 vs. 残差增强: CNN + BERT 和 BiLSTM + BERT 模型虽然尝试在 BERT 编码层基础上增加局部特征提取模块, 但从实验结果看, 其 F1 分数 (0.8994 和 0.8985) 反而略低于 Vanilla BERT (0.9044)。这一现象值得深思: 简单堆叠的 CNN 或 BiLSTM 在微调过程中, 易产生语义信息损失或梯度弥散问题, 甚至可能引入特征冗余, 导致模型性能不升反降。

本文模型引入的残差块通过跳跃连接实现了全局特征与局部

特征的“无损融合”。残差机制确保了 BERT 提取出的高层语义背景能够指导一维卷积层对“发臭”、“异物”、“腹泻”等风险关键词, 从而在保持高精确率 (0.9265) 的同时, 有效识别出隐晦的食品安全风险信号, 实现了性能的进一步提升。

(二) 模型稳健性与收敛特性的讨论

五折交叉验证与训练动力学监测结果进一步印证了本文所提模型的优越性, 具体表现为以下两方面:

1. 极低的标准差: 从表 5 可见, 本文模型在五折交叉验证中的标准差仅为 0.0062, 远低于 CNN + BERT 的 0.0177。这表明残差结构不仅能够提升模型的特征抽象能力, 还发挥了类似正则化的作用, 降低了模型对特定训练样本集的依赖性, 增强了模型的泛化能力。特别是在 Fold 5 中, 当 CNN + BERT 出现明显性能下滑 (F1 分数降至 0.8635) 时, 本文模型依然稳健地保持在 0.9011, 充分印证了多尺度特征融合机制在面对复杂、高噪声语料时具有更强的纠错能力和稳定性。

2. 收敛效率与过拟合控制: 图 6 的训练动力学曲线显示, 引入残差块后, 模型的训练损失下降速度更快且更平滑。这验证了残差学习通过将优化目标从 $H_{(x)}$ 转变为 $F_{(x)}=H_{(x)}-x$, 显著降低了深层网络的优化难度。验证集 F1 分数在第 2 轮训练即达到峰值 0.9153, 配合早停策略, 模型能够精准停留在泛化能力最强的阶段, 有效避免了因网络结构过深导致的过拟合问题。这一收敛速度远快于传统深度模型, 充分体现了残差连接对梯度流动的优化作用, 提升了模型训练的效率与稳定性。

(三) 垂直领域的实际应用价值

食品安全监测是一个对“漏报”高度敏感的领域, 漏报的风险样本可能直接威胁消费者身体健康。本文所提模型在测试集上实现了 0.92 的风险类别 F1 分数 (见表 6), 召回率达到 0.89, 相较于各类基线模型具有更强的工程实践意义, 具体体现在以下三点:

1. 多尺度特征感知: O2O 评论文本通常由“全局情感倾向 (如不满、愤慨)”和“局部事实描述 (如菜里有头发、吃完肚子疼)”组成。本文模型通过拼接 pooler_output 全局语义向量和经过残差处理的局部特征向量, 构建了多尺度特征空间。这种设计比单一维度的特征表征更适配食品安全监测这一复杂二分类任务, 能够同时捕捉用户情感倾向与具体风险信号, 提升风险识别的精准度。

2. 误报与漏报的平衡: 从混淆矩阵 (图 4) 可见, 本文模型在保持高精确率 (非风险样本 0.98、风险样本 0.94) 的同时, 将漏报数量控制在 32 条, 仅占总样本的 1.6%。尽管漏报问题仍是未来主要改进方向, 但相较于传统方法, 本文模型已显著降低了风险样本的遗漏概率, 为模型的实际部署提供了可靠的技术基础。

3. 局限性: 尽管模型整体表现优异, 但误分类样本分析也揭示了其对极致反讽表达的处理能力不足, 这类误分类样本占总错误样本的 40%。例如“这家的饭菜真‘干净’, 吃完就拉肚子”这类正向词汇, 而忽略后续的风险信号。这表明仅依靠文本内部特征进行语义识别仍存在上限, 未来研究可尝试引入外部知识图谱 (如食品卫生专业词库) 或多模态信息 (评论配图), 进一步

实现语义增强，提升模型对复杂隐晦表达的认识能力。

四、结论与展望

（一）研究结论

本文针对 O2O 餐饮平台食品安全监管的迫切需求，提出并实现了一种融合 BERT 与残差网络的食品安全评论识别模型。通过在真实 O2O 评论数据集上的系统实验与对比分析，得出以下核心结论：

1. 特征融合的有效性：实验结果证明，单一预训练模型在处理细粒度食品安全风险信号时存在明显局限。本文提出的 BERT + Residual Block 架构，通过残差连接实现了全局语义背景与局部敏感特征（如“异物”、“腹泻”等关键短语）的协同建模，其测试集 F1 分数达到 0.9096，验证集最高 F1 分数达 0.9257，显著优于传统机器学习模型及简单叠加的深度学习模型，充分验证了特征融合策略的有效性。

2. 残差结构的优越性：残差块的引入不仅提升了模型的特征抽象能力，更通过跳跃连接有效缓解了模型微调过程中的梯度不稳定问题。五折交叉验证结果显示，本文模型的标准差仅为 0.0062，展现出极强的稳健性与泛化能力，能够有效应对 O2O 评论文本高噪声、口语化、非规范的 data 分布特点。

3. 工程应用价值：模型在风险类样本的召回率上表现优异，

整体准确率达到 97.5%。通过参数敏感性分析，确定了模型的最优学习率（ $3e-5$ ）与残差块数量（ $k=4$ ），为智慧监管系统在海量用户评价中自动预警食品安全风险，提供了一套高效、稳定、可落地的技术方案，具有重要的工程实践价值。

（二）不足与展望

尽管本文模型在特征提取、稳健性、工程应用等多个维度取得了突破，但仍存在进一步优化的空间，未来研究方向主要包括以下三点：

1. 语义理解的深度优化：误分类分析表明，模型对“反讽”及“隐晦表达”的识别仍面临挑战。未来研究可尝试引入情感增强向量或外部食品安全知识图谱，通过知识注入（Knowledge Injection）技术，提升模型对特定行业语境下反向语义、隐晦风险信号的捕捉能力，进一步降低误分类率。

2. 多模态特征融合：目前模型的监测仅依赖文本信息，而 O2O 平台评论通常伴随用户上传的实拍图片。未来可探索视觉-文本双模态模型，利用图像中的异物、环境卫生等视觉信息，补齐文本描述的不足，实现多维度风险识别，进一步降低漏报率。

3. 模型实时性优化：为适应大规模在线监测场景的需求，后续可研究模型蒸馏（Distillation）技术，在保留残差结构性能优势的前提下，降低模型参数量与计算开销，实现模型的轻量化部署，提升风险监测的实时性。

参考文献

- [1] 宋超玉, 方美芳. O2O 模式在大学及周边的发展研究——以蚌埠高校为例 [J]. 山西农经, 2018, (06): 110-111. DOI: 10.16675/j.cnki.cn14-1065/f.2018.06.083.
- [2] Zhang Q, Wu A, Liu L, et al. Application of Machine Learning in Food Safety Risk Assessment [J]. Foods, 2025, 14(23): 4005.
- [3] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques [C]// Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. 2002: 79-86.
- [4] Kim Y. Convolutional Neural Networks for Sentence Classification [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014: 1746-1751.
- [5] Tang D, Qin B, Liu T. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification [C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2015: 1422-1432.
- [6] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [EB/OL]. (2018-10-11). <https://arxiv.org/abs/1810.04805>.
- [7] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 770-778.
- [8] 梁小林, 柳映堂, 梁盟, 等. 基于偏最小二乘支持向量机模型的个人信用评估研究 [J]. 湖南文理学院学报 (自然科学版), 2021, 33(04): 6-10.
- [9] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint arXiv:1810.04805, 2018.
- [10] CORTES C, VAPNIK V. Support-vector networks [J]. Machine Learning, 1995, 20(3): 273-297.
- [11] CHEN T, GUESTRIN C. XGBoost: A Scalable Tree Boosting System [C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 785-794.
- [12] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [J]. arXiv preprint arXiv:1810.04805, 2018.
- [13] KIM Y. Convolutional Neural Networks for Sentence Classification [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014: 1746-1751.
- [14] GRAVES A, SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures [J]. Neural Networks, 2005, 18(5-6): 602-610.
- [15] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 770-778.
- [16] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint arXiv:1810.04805, 2018.

基于代价敏感 XGBoost 模型的企业 ESG 评级预测

鲍金敏

合肥工业大学 数学学院, 安徽 合肥 230009

DOI:10.61369/ASDS.2026060010

摘 要 : 针对企业环境、社会与治理 (ESG) 评级分布不平衡导致的预测偏差, 以及现有模型对不同行业 ESG 行为解释力不足的问题, 本文提出一种基于代价敏感极限梯度提升 (XGBoost) 与沙普利加法解释 (SHAP) 框架结合的改进预测模型。我们引入代价敏感学习策略, 并结合合成少数类过采样技术 (SMOTE), 强化模型对少数类等级样本的捕获能力。同时量化财务指标的边际增益并利用 SHAP 算法剖析行业异质性。研究结果表明代价敏感优化显著提升了极端评级的预测精度, “落后者”召回率91.72%, “领导者”召回率92.24%; 财务基本面和行业信息的引入有效修正了历史评级的依赖性, 宏平均受试者工作特征曲线下面积 (Macro AUC) 高达97.81%; 驱动机制存在行业分化, 重污染行业呈现出“资产密集型合规”逻辑, 而非重污染行业则展现出“效率溢价与市场敏感”特征。本文为投资者精准锚定优质 ESG 标的、排查尾部风险提供了高效的决策工具。

关 键 词 : ESG 评级; 代价敏感 XGBoost 模型; SHAP; 行业异质性

Enterprise ESG Rating Prediction Based on Cost-Sensitive XGBoost Model

Bao Jinmin

School of Mathematics, Hefei University of Technology, Hefei, Anhui 230009

Abstract : Aiming at the prediction bias caused by the imbalance in the distribution of ESG ratings and the insufficient explanatory power of existing models for ESG behavior in different industries, this paper proposes an improved prediction model based on the combination of cost-sensitive limit gradient lifting (XGBoost) and Shapli addition explanation (SHAP) framework. We introduce a cost-sensitive learning strategy combined with synthetic minority oversampling (SMOTE) to enhance the model's ability to capture minority-class samples. At the same time, the marginal gain of financial indicators is quantified, and the SHAP algorithm is used to analyze the industry heterogeneity. The results show that cost-sensitive optimization significantly improves the prediction accuracy of extreme ratings, with a recall rate of 91.72% for "laggards" and 92.24% for "leaders". The introduction of financial fundamentals and industry information effectively corrects the dependence of historical ratings, and the macro average receiver operating characteristic area (Macro AUC) is as high as 97.81%. The driving mechanism is differentiated by industries, with heavy polluting industries showing the logic of "asset-intensive compliance", while non-heavy polluting industries showing the characteristics of "efficiency premium and market sensitivity". This article provides an efficient decision-making tool for investors to accurately anchor high-quality ESG targets and investigate tail risks.

Keywords : ESG rating; cost-sensitive XGBoost model; SHAP; Industry heterogeneity

引言

在全球经济逐步转向可持续发展、我国“双碳”目标持续推进的背景下, 环境、社会与治理 (ESG) 表现已不仅是监管部门推动企业绿色转型的重要抓手, 也在企业层面体现出较为明显的价值创造作用。从信号理论与代理理论视角看^[1], 管理层借助 ESG 评级向股东传递相关信息, ESG 表现较好的企业往往更容易获得投资者关注, 并对其融资方式和融资工具产生影响^[2]。在价值创造方面, 研究表明, ESG 评价与资产回报率、股本回报率等财务绩效指标呈正相关关系, 是推动企业价值提升的重要因素^[3, 4]。在具体应用中, 将 ESG 信息纳入制造业评估体系, 有助于增强模型对企业信用风险的识别与预警能力^[5]。不过, 现有 ESG 评级体系仍存在第三方评价结果滞后、信息披露成本较高以及评级过程透明度不足等问题, 不同评级机构之间还存在较为明显的分歧现象^[6]。在此背景下, 构建更为准确的 ESG 预测机制, 对引导企业绿色创新和促进高质量发展具有现实意义^[7, 8]。

作者简介: 鲍金敏, 合肥工业大学数学学院, 硕士研究生, 研究方向为 ESG 数据分析与机器学习。

为提高 ESG 表现的预判准确性,学术界逐步将研究重点放在机器学习方法的应用上。DeLucia 等人^[9]选用随机森林、支持向量机等模型,并基于汤森路透数据库实证分析,发现 ROA 与 ROE 对 ESG 指标具有较好的预测效果。Gupta 等人^[10]则通过线性回归与随机森林回归方法,进一步分析利润率对 ESG 评分的影响。D'amato 等人^[11]从财务报表角度出发,指出资产负债表项目对 ESG 评级具备显著解释力。随着算法不断发展,研究开始更多采用非线性模型以提升预测精度。CINI 等人^[12]利用随机森林分类模型,对道琼斯欧洲斯托克 600 指数公司进行分析,结果表明资产负债变量及系统性风险对评级类别具有较强解释能力,模型预测效果较为理想。Kartal 等人^[13]以土耳其上市公司 ESG 报告为样本,引入超级学习者算法,模型预测准确率达到 94.3%。TASKIND 等人^[14]则发现,历史 ESG 评分与评级本身即可作为未来 ESG 表现的有效预测依据,为复杂数据驱动方法提供了简洁替代路径。Zhang 等人^[15]结合 XGBoost 算法与 SHAP 解释方法,在提升评级预测准确性的同时,对各类指标的整体贡献进行了更为细致的分析。

尽管现有研究虽已取得一定进展,但仍存在若干不足。首先,在识别偏误方面,ESG 等级通常呈现明显的不平衡分布,传统模型多默认各类别地位相同,实际训练中往往降低了对少数类样本的识别能力,从而影响极端风险情形下的稳定性。其次,在评价维度上,尽管已有研究关注历史评级与财务表现之间的联系,但对其他基本面信息在评级预测中的边际作用缺乏清晰界定。再次,在解释层面,现有成果多停留在行业整体分析,对不同行业间决策逻辑差异的刻画仍不充分。

针对上述问题,本文构建基于改进 XGBoost 的企业 ESG 评级预测框架。具体而言:其一,在处理类别不平衡问题时,通过引入类别权重的代价敏感学习方法,在目标函数中设置差异化惩罚权重,从而增强模型对少数类别样本的识别能力,缓解由样本分布不均带来的偏置。其二,通过多维度消融实验,对其他基本面信息在 ESG 评级预测中的增量作用进行量化分析,为企业识别关键变量提供依据。其三,在结果解释方面,借助 SHAP 方法分析不同行业的驱动机制差异,进一步揭示重污染与非重污染企业在 ESG 表现上的不同逻辑。

本文围绕 ESG 评分的可解释预测问题,构建了较为系统的研究框架,并为后续相关研究提供参考依据。行文结构安排如下:第一部分介绍理论基础,第二部分说明模型与方法,第三部分呈现实证结果及模型表现,第四部分归纳主要结论并对未来研究进行展望。

一、基本方法

(一) XGBoost 算法

XGBoost^[16]是一种先进的集合学习算法,旨在提升梯度增强树的计算效率和预测性能,其通过迭代构建多棵决策树来拟合前一模型的残差。在 ESG 评级的预测背景下,XGBoost 算法旨在最小化由损失函数和正则化项两部分组成的目标函数 $L^{(t)}$ 。如下所示:

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (1)$$

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (2)$$

$l(\cdot)$ 为度量 ESG 真实评级 y_i 与预测评级 $\hat{y}_i$ 差异的损失函数; $\Omega(f_t)$ 为正则化项, γ 和 λ 是正则化项的系数,叶节点数量 T 及权重 w_j 用来控制模型复杂度。XGBoost 通过对目标函数进行二阶泰勒展开,利用一阶和二阶导数信息寻找最优分裂点,能够有效捕捉预测指标和评级之间的非线性关系。

(二) 代价敏感学习

ESG 评级样本在等级分布上往往呈现不平衡特征,多数类别的样本在传统分类模型训练过程中会起到主导作用,导致模型对少数类别样本的判别能力不足,产生系统性的分类偏置问题。为缓解上述问题,本文引入基于类别权重调整的代价敏感学习策略^[17],对模型训练过程中的不同类别样本赋予差异化的重要性权重。

具体来说,代价敏感学习通过在模型损失函数中引入类别权重向量,使模型在优化过程中对少数类别样本的误判承担更高的

损失惩罚,引导模型更加关注少数类别的判别边界学习。本文在此基础上结合 SMOTE^[18]过采样技术,构建“数据层重采样—算法层加权”的双重均衡机制。首先利用 SMOTE 方法对训练集中少数类别样本进行线性插值扩充,接着在 XGBoost 模型的目标函数中引入基于类别的权重向量,对不同等级样本在损失函数中的相对重要性进行调整。改进后的目标函数如(3)所示:

$$L_{cs}^{(t)} = \sum_{i=1}^n \mathbf{w}_{y_i} \cdot l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (4)$$

$\mathbf{w}_{y_i}$ 为类别权重向量。本文结合样本分布特征与实验调优结果对少数类等级赋予较高的惩罚权重,对多数类等级赋予基准权重。

(三) SHAP 算法

SHAP 算法利用 Shapley 值来解释机器学习模型,旨在为每个特征为每个样本分配一个贡献值。通过计算其对应 SHAP 值的绝对平均值来量化特征的显著性。通过特征值与其 SHAP 值的图形表示,可以辨别每个特征对模型预测的影响^[19]。该方法既可用于分析单一样本层面的预测机制,也能够整体层面揭示模型的特征依赖结构,有助于实现对模型预测行为的局部与全局理解。

设定样本数量为 n ,特征集合为 F ,特征维度为 M 。对于第 i 个样本,模型的预测输出可表示为各特征贡献值的加法和:

$$f(x_i) = \phi_0 + \sum_{j=1}^M \phi_{ij} \quad (5)$$

$f(x_i)$ 为解释模型, ϕ_0 为所有样本预测的平均值, ϕ_{ij} 代表第 i

个样本中第 j 个特征的贡献值。 ϕ_{ij} 的计算公式如下：

$$\phi_{ij} = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(M-|S|-1)!}{M!} [f_{S \cup \{j\}}(x_i) - f_S(x_i)] \quad (5)$$

S 是特征集合 F 中不包含特征 j 的子集， $f(S)$ 表示特征子集 S 下模型的预测输出。本文采用 “One-vs-Rest” 策略^[20]，分别计算每个目标类别对应的 SHAP 值， $I_j^{(k)}$ 表示 i 个样本中第 j 个特征对类别 k 的贡献值，则第 j 个特征在类别 k 下的全局重要性定义为：

$$I_j^{(k)} = \frac{1}{N} \sum_{i=1}^N |\phi_{ij}^{(k)}| \quad (6)$$

同时，本文将各类别下的全局重要性进行等权累加用来刻画特征在整体多分类任务中的综合影响力，定义总重要性为：

$$I_j = \sum_{k=1}^K I_j^{(k)} \quad (7)$$

其中 $K=3$ 表示评级类别数量。

二、模型与方法

（一）数据来源和变量说明

本文构建以历史 ESG 评级与财务指标为核心的预测指标体系。参考既有研究^[2, 11]所用框架，本文筛选与 ESG 表现关联较强的关键财务变量，以保证评估结果具备完整性与可靠性。整个指标体系划分为五个维度。就盈利与营运情况而言，选取净资产收益率、总资产净利润率、息税前利润以及人均创利，用于衡量资源配置效率。围绕资本结构与风险水平，引入资产负债率、利息保障倍数以及自由现金流，以反映财务稳健程度。针对市场表现与成长性，采用年换手率、市盈率、营业收入增长率和资本密集度，对市场关注程度与扩张能力进行刻画。在行业划分方面，依据证监会行业分类标准，将样本区分为“重污染行业”和“非重污染行业”，并设置虚拟变量，用于识别行业差异。考虑到历史评级具备延续特征^[14]，本文引入“上年综合得分”，以反映 ESG 表现的动态惯性。

本文数据来源于华证 ESG 评价体系与 CSMAR 数据库，相关指标定义见表1。

表1：指标定义

维度	变量名称	变量符号	变量定义与计算方法
历史评级特征	上年综合得分	Score{t-1}	企业在 t-1 年度的华证 ESG 综合得分
	净资产收益率	ROE	净利润 / 股东权益平均余额
	总资产净利润率	ROA	净利润 / 总资产平均余额
	息税前利润	EBIT	利润总额 + 利息支出
	人均创利	PPL	净利润 / 员工总数
	营业收入增长率	Growth	(本期营收 - 上期营收) / 上期营收
财务盈利能力	资本密集度	CI	总资产 / 营业收入
	年换手率	Turnover	统计区间内股票累计成交量 / 流通总股数
	资产负债率	Lev	总负债 / 总资产
	利息保障倍数	ICR	息税前利润 / 利息费用
	企业自由现金流	FCFF	息税前利润 - 税收 + 折旧摊销 - 营运资本变动 - 资本开支
	市盈率	PE	每股市价 / 每股收益
行业属性	重污染分组	Pollute	虚拟变量：若企业属于重污染行业取1，否则取0

（二）样本划分与不平衡处理策略

本研究以华证 ESG 评级为主要数据来源，实现了严格的数据收集和预处理程序。我们选取2014—2023年沪深 A 股上市公司企业 ESG 表现、财务指标及行业属性为初始数据集。为确保数据有效性，本文实施如下预处理：首先，剔除样本期间内被标记为 ST、PT 或 ST* 的异常经营企业；其次，针对部分缺失值采用 KNN 算法进行填补，并删除存在严重数据异常的公司。最终构建了一个包含610家公司、共6100个企业年度观测值的平衡面板数据集。

在评级划分方面，由于华证 ESG 评价分为九级，且样本高度集中于中位等级，为缓解类别失衡并提升模型预测精度，本文参照中证指数（CSI）官方标准，将原始等级整合为三类：将 AAA 至 BBB 级归为“领导者”，BB 至 CCC 级归为“平均者”，CC

及 C 级归为“落后者”。为便于模型计算，将三类标签编码为0、1、2。

此外，为进一步处理整合后仍存在的样本分布差异，本文采用 SMOTE 算法对数据集缓解不平衡处理，样本总量由初始的6100个扩充至10429个。样本均衡后的频率百分比见表2。

表2：样本均衡后频率的百分比

级别	频率	百分比（%）
领导者	3212	30.80
平均者	5589	53.59
落后者	1628	15.61
总计	10429	100.00

（三）代价敏感 XGBoost 模型参数设计

XGBoost 模型的性能依赖于超参数的调优^[21]。本研究针对 ESG 三分类任务，对核心超参数进行了系统化配置，通过随机搜

索策略与交叉验证的结合，最终确定的最优超参数配置见表3。

表3：最优超参数配置超参数

超参数	取值	说明
最大树深度	12	控制基学习器复杂度，用于刻画非线性关系
迭代轮数	1000	表示基学习器数量，影响模型拟合能力
学习率	0.015	控制新增树对模型的贡献权重
样本采样比例	0.8	每轮随机抽取80%样本进行训练
特征采样比例	0.8	每棵树随机选取80%特征建模
随机种子	[42,100,2024,888,123]	多随机种子重复实验并取平均值
代价敏感系数	{0.3,1:1,2:2}	提高少数类别的惩罚权重

（四）模型评估指标

混淆矩阵用于评估多分类模型性能，通过统计实际类别与预测结果的对应关系，可以直观看出模型判别偏差。在本研究的ESG评级三分类模型中，混淆矩阵C是一个3×3的方阵。横轴表示预测等级，纵轴表示实际等级，类别划分为“落后者”“平均者”“领导者”。矩阵中的元素 C_{ij} 表示实际等级为i但被分类为j的样本数量。

1. 准确性

准确率表示预测正确的样本在总体中的占比，其计算方式见公式（8）。

$$Acc = \frac{\sum_{i=0}^2 C_{ii}}{\sum_{i=0}^2 \sum_{j=0}^2 C_{ij}} \quad (8)$$

其中： $i, j \in \{0,1,2\}$ 分别代表“落后者”“平均者”与“领导者”。

2. 精度、召回率和F1分数

精度用于衡量预测结果的可靠性，即被判为某一等级的样本中，真实属于该等级的比例。对应计算方式见公式（9）。

$$P_i = \frac{C_{ii}}{\sum_{j=0}^2 C_{ji}} \quad (9)$$

召回率反映识别能力，即在实际属于某一等级的样本中，被正确识别的比例，对于等级*i* 其召回率公式如（10）所示。

$$R_i = \frac{C_{ii}}{\sum_{j=0}^2 C_{ij}} \quad (10)$$

F1分数综合考虑精度与召回率，取二者调和平均值，用以衡量单一等级上的整体判别效果，计算方式见公式（11）。

$$F1_i = 2 \times \frac{P_i \times R_i}{P_i + R_i} \quad (11)$$

精度、召回率和F1分数的数值越高，表示模型在相应分类层面上表现更优越。考虑到“平均者”样本数量较多，容易对整体评价产生主导影响，本文引入宏平均方法。该方法对各类别指标

赋予相同权重后取平均值，从而更合理地反映模型在少数类别上的识别能力，其计算方式见公式（12）。

$$M_{macro} = \frac{1}{3} \sum_{i=0}^2 m_i \quad (12)$$

其中 m_i 表示第*i*类的对应性能指标。本文在未特指某一等级的性能指标情况下（如宏精度、宏召回率、宏F1值及Macro AUC），均指代经宏平均法计算后的结果。

3.ROC 曲线与 AUC 值

ROC 曲线以图形方式展示了真向正类率与假向正类率之间的权衡关系。本研究分别绘制“落后者”“平均者”与“领导者”的ROC 曲线。AUC 作为衡量模型分类性能的标量值，其数值越接近1，代表模型的判别准确度越高。

三、模型预测结果与行业异质性分析

（一）预测性能与策略有效性分析

本文将平衡后的数据集按8:2的比例划分为训练集和测试集，并基于改进后的代价敏感XGBoost模型进行等级判别。为了实现精准的等级分类建模，本文将华证ESG评级整合为三个核心类别：第0、1和2类分别代表“落后者”“平均者”与“领导者”。为兼顾结果的可视化直观性与统计稳健性，本文在图示层面报告单一随机种子下的混淆矩阵结果，而在模型评价层面采用多随机种子重复实验并取均值作为最终结论依据。测试集样本的预测分布结果由图1的混淆矩阵呈现（以随机种子123的单次实验结果为例）。

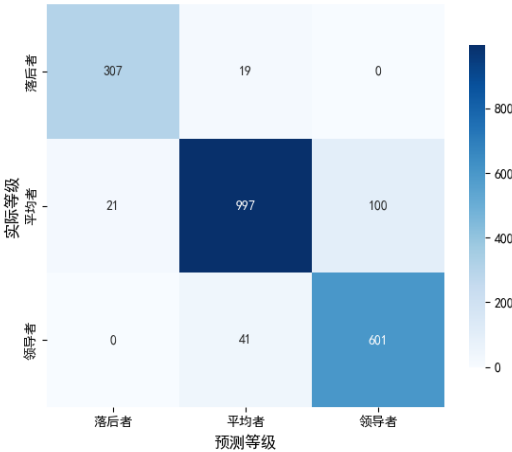


图1：代价敏感XGBoost模型的混淆矩阵

从图1可以看出，代价敏感XGBoost模型展现了显著的分类性能，样本高度集中在主对角线上。并且模型对于“落后者”与“领导者”的判别没有跨级误判，这意味着模型能够精准识别出表现最优与最差的企业。误判样本主要分布在相邻等级之间，如有100个样本所属于“平均者”被归类为“领导者”行列。为消除数据随机划分对模型评价的潜在干扰，本文在相同的超参数配置下，选取了5组不同的随机种子进行重复实验，并以各评估指标的算术平均值作为最终评价依据。实验结果如表4所示：

表4：基于代价敏感优化的模型预测性能对比检验

	准确率	宏精度	宏 F1 值	落后者召回率	领导者召回率
XGboost 模型	89.95%	89.91%	89.70%	87.85%	90.53%
代价敏感 XGboost 模型	90.22%	89.62%	90.18%	91.72%	92.24%

通过表4，基准 XGboost 模型与代价敏感 XGBoost 模型在整体平均准确率和宏平均精度都能达到89%以上，体现了 XGBoost 框架在 ESG 评级预测中的效能。同时代价敏感学习发挥了重要的调节作用，能够在提升整体准确率的同时显著改善极端类别识别能力，使“落后者”召回率从87.85%提升至91.72%，“领导者”召回率从90.53%优化至92.24%。这种针对性优化，增强了模型对少数等级样本的捕捉能力，为投资者锚定尾部风险与优质标的提供了可靠的技术支撑。

本文进一步通过 ROC 曲线考察了代价敏感 XGBoost 模型对不同 ESG 等级的判别区分能力。根据图2，“落后者”“平均者”与“领导者”三类预测的 ROC 曲线均高度趋近于左上角，对应的单类 AUC 均值分别达到了0.9948、0.9685和0.9815。表明该模型在为投资者提供公司在 ESG 领域的相对位置方面具有高度准确性。

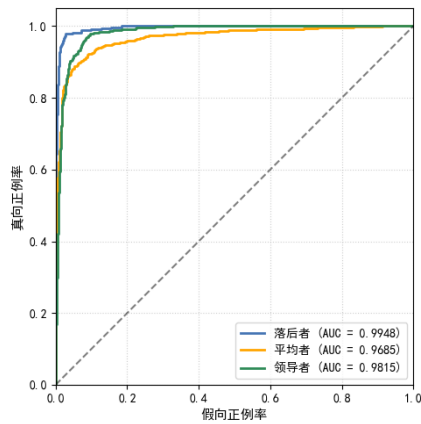


图2：代价敏感 XGboost 模型的 ROC 曲线

（二）特征重要性分析

在完成代价敏感 XGBoost 模型构建后，本文引入 SHAP 方法分析特征贡献。为量化各变量影响，通过计算各特征在三个类别上的 SHAP 平均绝对值并进行堆叠排序。由此可以识别各特征在模型决策中的相对重要程度。

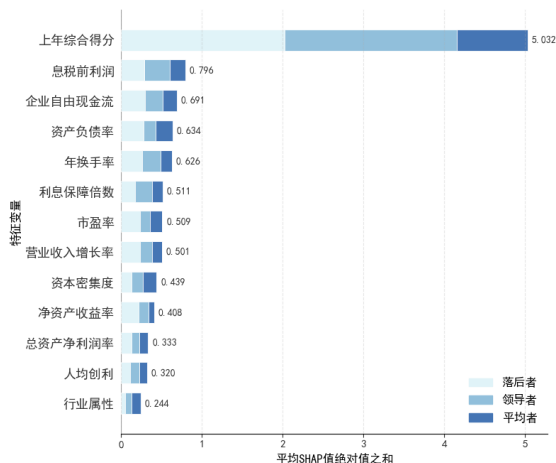


图3：特征全局贡献度量化排名

图3显示，“上年综合得分”对应 SHAP 均值之和达到5.032，

远高于其他变量。该结果表明，企业 ESG 表现存在明显路径依赖，历史评级对未来结果具有关键影响。

虽然“上年综合得分”占主导地位，但其他变量同样发挥作用。息税前利润、企业自由现金流以及资产负债率对应 SHAP 均值之和均超过0.6，对模型输出形成补充影响。其余指标也参与模型判别过程，并带来一定增量信息，相关结果见表5。

表5：财务变量对模型性能的整体增量分析

	准确率	宏精度	宏平均召回率	宏 F1 值	Macro AUC
仅历史得分模型	74.18%	74.07%	68.32%	70.35%	87.48%
历史 + 财务 + 行业模型	90.22%	89.62%	90.86%	90.18%	97.81%

消融实验统一采用代价敏感超参数配置。表5结果表明，引入财务变量与行业属性后，单一维度预测局限得到缓解，模型整体性能明显改善。具体来看，准确率由74.18%提升至90.22%，宏平均精度由74.07%提高至89.62%，Macro AUC 达到97.81%。上述结果说明，历史评级具有较强解释力，但单一依赖会带来路径依赖偏差，财务与行业信息可作为补充进行修正。

（三）行业异质性分析

鉴于各行业在环境规制、资源消耗及社会责任履行方面差异明显，本文将样本划分为重污染行业与非重污染行业两类。基于该划分，进一步检验代价敏感 XGBoost 模型在不同情境下的适用性，并分析其驱动机制差异。

1. 预测性能的稳健性检验

从表6可见，模型在不同行业样本中均表现出稳定的预测能力，并能够有效识别“落后者”与“领导者”。其中，不同行业属性下 Macro AUC 表现均高于全样本模型下的97.81%，表明当样本空间缩小至特定行业内部时，行业内部结构更为同质化，有利于模型学习稳定判别边界，从而展现出更优的判别灵敏度。其中非重污染行业的各项评价指标略高于重污染行业，其 F1 值达到93.29%，Macro AUC 值高达98.97%。这表明尽管重污染行业面临更为复杂的监管环境，其预测准确率与精度依然维持在较高水平，并与非重污染行业的差距极小。

表6：基于行业异质性的模型预测性能对比检验

模型组别	准确率	宏精度	宏平均召回率	宏 F1 值	Macro AUC
重污染行业	92.06%	92.49%	92.05%	91.87%	98.57%
非重污染行业	93.43%	93.70%	93.43%	93.29%	98.97%

2. 驱动机制的差异化分析

本文引入 SHAP 方法，对重污染行业与非重污染行业的驱动差异进行分析。图4展示“领导者”类别预测结果中各特征的边际贡献分布，横轴为 SHAP 值，用于衡量变量对预测结果的影响方向与强度。结合特征取值分布可以进一步判断变量作用方向：红色表示取值较高，蓝色表示取值较低。若高取值样本主要集中在横轴正向区间，则说明该变量与评级结果呈正向关系。

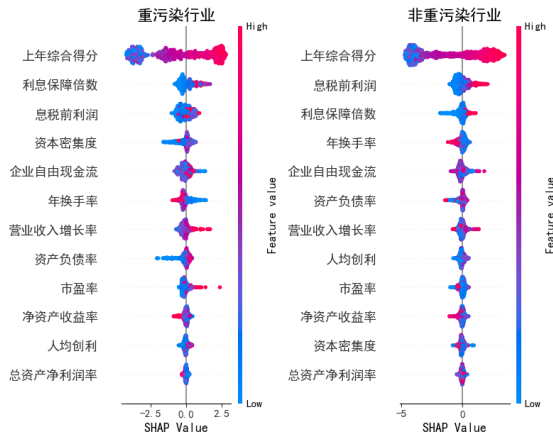


图4：“领导者”行业驱动机制对比分析

结果显示，不论企业是否属于重污染行业，“上年综合得分”始终处于核心位置，是能否进入“领导者”行列的关键因素。SHAP 分布进一步表明，该指标取值较高时，多集中在横轴正向区间，对预测结果产生明显正向作用。与此同时，其他财务变量在不同行业中呈现差异化特征。

就重污染行业而言，资本密集度重要性排名明显靠前，高于非重污染行业。结合 SHAP 分布可见，较高资本密集度通常对应更高评级结果，二者呈正向关系。这一现象反映出，在重资产行业中，固定资产投入规模往往与企业治理能力相关，进而影响 ESG 表现。

就非重污染行业而言，人均创利与年换手率的重要性更为突出，其排名均高于重污染行业。该结果表明，对于轻资产企业，模

型更依赖人力产出效率以及市场交易活跃程度来判断 ESG 等级。

四、结论

（一）企业 ESG 表现呈现明显路径依赖特征

各模型结果均显示，“上年综合得分”始终处于特征重要性首位。对应 SHAP 值集中于正向区间，对预测结果产生持续推动作用。该现象说明，ESG 评级具有延续性，既往建立的 ESG 治理体系与声誉积累会对后续表现产生长期影响。

（二）财务信息对评级预测具有补充作用

消融结果表明，引入财务指标后，模型准确率由 74.18% 提高至 90.22%，Macro AUC 达到 97.81%。这一变化表明，财务数据不仅反映经营状况，同时能够缓解历史评级带来的偏置，对预测结果形成有效补充。

（三）不同行业存在差异化驱动路径

分行业结果显示，重污染行业中资本密集度排名更高，而非重污染行业中，人均创利与年换手率更为关键。上述差异说明，重资产企业更依赖资本投入支撑表现，轻资产企业则更多受效率与市场活跃度影响。

（四）代价敏感方法提升极端类别识别能力

针对样本分布两端较少的情况，引入差异化权重后，模型对“落后者”与“领导者”的识别能力得到增强，对应召回率分别达到 91.72% 与 92.24%。该方法有助于提高极端类别识别效果，为风险识别与优质样本筛选提供参考。

参考文献

[1]JENSEN M C, MECKLING W H. Theory of the firm: managerial behavior, agency costs and ownership structure[J]. Journal of Financial Economics, 1976, 3(4): 305–360.

[2]D’ AMATO V, D’ ECCLESIA R, LEVANTESI S. ESG score prediction through random forest algorithm[J]. Computational Management Science, 2022, 19(2): 347–373.

[3]VELTE P. Does ESG performance have an impact on financial performance? Evidence from Germany[J]. Journal of Global Responsibility, 2017, 8(2): 169–178.

[4]SANDBERG H, ALNOOR A, TIBERIUS V. Environmental, social, and governance ratings and financial performance: evidence from the European food industry[J]. Business Strategy and the Environment, 2022, 32(4): 2471–2489.

[5]史瑜港. 融入 ESG 信息的制造业上市公司信用风险预警研究 [D]. 兰州: 西北师范大学, 2025.

[6]袁蓉丽, 江纳, 刘梦瑶. ESG 研究综述与展望 [J]. 财会月刊, 2022(17):128–134.

[7]梁四安, 李淑萍, 段军山, 等. ESG 表现如何影响企业绿色创新? —— 基于融资约束与分析师关注的中介效应 [J/OL].(2025–12–26)[2026–01–21].<https://link.cnki.net/urlid/61.1294.N.20251226.1133.008>.

[8]尚煊川, 蒋建洪. ESG 评级对企业高质量发展的影响研究 [J]. 商业观察, 2025, 11(34):106–111.

[9]DE LUCIA C, PAZIENZA P, BARTLETT M. Does good ESG lead to better financial performances by firms? Machine learning and logistic regression models of public enterprises in Europe [J]. Sustainability, 2020, 12(13): 5317.

[10]GARCÍA F, GONZÁLEZ-BUENO J, GUIJARRO F, et al. Forecasting the environmental, social, and governance rating of firms by using corporate financial performance variables: a rough set approach[J]. Sustainability, 2020, 12(8): 3324.

[11]D’ AMATO V, D’ ECCLESIA R, LEVANTESI S. Fundamental ratios as predictors of ESG scores: a machine learning approach[J]. Decisions in Economics and Finance, 2021, 44: 1087–1110.

[12]CINI F, FERRARI A. Towards the estimation of ESG ratings: a machine learning approach using balance sheet ratios [J]. Research in International Business and Finance, 2024, 73: 102653.

[13]Kartal M T, Taşkın D, Shahbaz M, Depren S K, Pata U K. Effects of environment, social, and governance (ESG) disclosures on ESG scores: investigating the role of corporate governance for publicly traded Turkish companies[J]. Journal of Environmental Management, 2024, 368: 122205.

[14]TASKIN D, SARIYER G, ACAR E, et al. Do past ESG scores efficiently predict future ESG performance?[J]. Research in International Business and Finance, 2025, 74: 102706.

[15]ZHANG J F, ZHAO Z X. Corporate ESG rating prediction based on XGBoost–SHAP interpretable machine learning model[J]. Expert Systems with Applications, 2026, 295: 128809.

[16]CHEN T, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016: 785–794.

[17]张影. 基于代价敏感的 XGBoost 算法在 T2DM 患者并发糖尿病肾病预测中的应用 [D]. 湘潭: 湘潭大学, 2024.

[18]李爱华, 刘婉婷, 陈思帆, 等. 面向不平衡数据的 SMOTE–BO–XGBoost 集成信用评分模型研究 [J/OL].(2024–10–11)[2025–11–20].<https://doi.org/10.16381/j.cnki.issn1003-207x.2023.0635>.

[19]BEN YAN, ZIQI Y, YAN W. The measurement and early warning of daily financial stability index based on XGBoost and SHAP: Evidence from China[J]. Expert Systems with Applications, 2023, 227: 120245.

[20]HAND D J, TILL R J. A simple generalisation of the area under the ROC curve for multiple class classification problems[J]. Machine Learning, 2001, 45(2): 171–186.

[21]MEHDARY A, CHEHRI A, JAKIMI A, et al. Hyperparameter optimization with genetic algorithms and XGBoost: A step forward in smart grid fraud detection[J]. Sensors, 2024, 24(4): 1121.

基于图神经网络和多目标粒子群优化的智能路径规划方法及其应用

吴荣康¹, 吴伟佳², 黄惠婷^{1*}

1. 广东技术师范大学 数学与系统科学学院, 广东 广州 510080

2. 广东石油化工学院 自动化学院, 广东 茂名 525000

DOI:10.61369/ASDS.2026060011

摘要 : 随着智能交通以及旅游规划的不断发​​展, 对有限时间、预算的约束条件下优化旅行路线、提高赏花效益的问题越来越突出。为了克服与假日赏花旅游有关的城市访问和路线规划难题, 本文提出了将图神经网络 (GNN) 和多目标粒子群优化 (MOPSO) 结合起来的智能路线规划方法 (GNN-MOPSO)。该方法用 GNN 从城市网络中提取结构特征, 用 MOPSO 优化多个目标, 从而达到同时最大化赏花价值、最小化旅行成本的目的。从实验结果可以看出, 在处理大规模城市网络的时候, GNN-MOPSO 比传统的 PSO、NSGA-II 等方法在解决方案质量和多样性上要​​好。另外 GNN-MOPSO 算法可以很好地处理复杂的约束, 具有在智能交通、旅游路线优化等方面的应用前景。最后, 本文对未来的几个研究方向进行了论述, 即强化学习的集成以及算法在更大的城市网络中是否可以推广。

关键词 : 图神经网络; 多目标粒子群优化; 路径规划; 赏花; 智能旅游

Intelligent Path Planning Method Based on Graph Neural Network and Multi-Objective Particle Swarm Optimization and its Application

Wu Rongkang¹, Wu Weijia², Huang Huiting^{1*}

1.College of Mathematics and Systems Science, Guangdong Polytechnic Normal University,
Guangzhou, Guangdong 510080

2.College of Automation, Guangdong University of Petrochemical Technology
Maoming, Guangdong 525000

Abstract : With the development of intelligent transportation and tourism planning, many people have long been eager to find a good travel route and maximise the enjoyment of flowers in a short amount of time and money. To solve the problem of city visits and route planning for holiday flower viewing tourism, this paper puts forward an intelligent route planning method (GNN-MOPSO) based on graph neural networks (GNN) and multi-objective particle swarm optimisation (MOPSO). A GNN is employed to obtain the structure information of the urban network, and then MOPSO is applied for simultaneous multi-objective optimisation of various objectives, that is to say maximisation of flower-viewing value and minimisation of travel cost. According to experiments, GNN-MOPSO is better than the previous optimisation methods PSO and NSGA-II in terms of the quality of solutions and the diversity of solutions for all cases, especially in large-scale urban networks. GNN-MOPSO can also be used to address the problem of multiple constraints in the optimisation of smart transportation and tourism paths. Therefore, based on the reasons mentioned above, the direction of future research in this paper will be to apply reinforcement learning and the scalability of algorithms to large-scale urban networks.

Keywords : graph neural network; multi-objective particle swarm optimization; path planning; flower viewing; smart tourism

基金项目: 2023年度广东省普通高校青年创新人才类项目 (2023KQNCX038)。

作者简介:

吴荣康, 广东技术师范大学数学与系统科学学院, 本科生, 专业应用统计学;

吴伟佳, 广东石油化工学院自动化学院, 本科生, 专业自动化。

通讯作者: 黄惠婷, 广东技术师范大学, 讲师, 研究方向为非参数统计。

引言

随着生活水平的提高以及文化旅游需求的变化,假日短途旅行也成了越来越受欢迎的一种交通方式^[1]。赏花旅游,在清明、五一等节假日的时候特别受人欢迎。赏花旅游有时间敏感性、空间异质性,开花季、花卉种类、天气、交通状况等各地区差别很大^[2]。在有限的假日时间和预算内安排跨城市旅行,使游客可以尽量多的欣赏到盛花期的花朵,又对交通费用加以控制,提出了一项具有挑战性的智能路线规划问题^[3]。

目前一些学者用 TSP (旅行商问题)、贪婪算法、遗传算法、蚁群算法等方法来优化旅行路线。另外一些研究把兴趣点推荐和预算约束等条件结合起来使用多目标建模的方法。但是这些方法大多存在以下问题,不能完全建模城市之间复杂的时空关系,如花期的重叠、城市的吸引力等,容易陷入局部最优,缺少全局视角,缺少针对花卉观赏旅游的专业建模,这是一种“多目标+时间+空间异质”的场景^[4-6]。

本文提出的跨城市花卉观赏路线智能规划方法 (GNN-MOPSO) 是将图神经网络 (GNN) 和多目标粒子群优化算法 (MOPSO) 结合起来的一种方法^[7]。GNN 从城市图中提取节点之间结构特征,即可达性、花期重叠、吸引力评分等。MOPSO 用多目标优化的方法,在一定的预算和时间内最大程度上提高花卉的观赏价值,同时使总旅行成本最小化^[8]。利用 GNN 特征加入到初始化、优化阶段,可以得到更加高效、结构感知的搜索过程^[9]。

此项研究既充实了智能交通及智能旅游领域所用算法的方法,又给实际旅行需求如有限的花期、有限的行程、有限的预算等给予了解决途径。GNN-MOPSO 方法具有很强的可扩展性、可解释性,给之后加入强化学习、实时交通反馈提供算法基础,具有重要的实际应用价值以及广泛的使用前景。

一、路径规划建模主要符号与含义

为了方便读者阅读和理解后面建模所用到的主要符号及意义,表 1 给出了本文中路径规划建模所用到的主要符号及其含义。之后的目标函数和约束会根据符号展开。

表 1: 测试集上的模型性能比较

符号	描述
C	候选城市集合,总共有 N 个城市
S_i	城市 i 的花期开始时间 (小时)
e_i	城市 i 的花期结束时间 (小时)
$[s_i, e_i]$	城市 i 的完全开花区间 (小时)
v_i	城市 i 的赏花价值分数 (例如,花的多样性,用户评分)
C_i	住在城市 i 的日常生活成本 (货币单位)
t_{ij}	从城市 i 到城市 j 的旅行时间 (小时)
C_{ij}	从城市 i 到城市 j 的旅行成本 (货币单位)
t_i	在城市 i 的停留时长 (小时)
x_i	x_i 城市 i 是否被访问;若访问则为 1, 否则为 0
R_j	花期重叠率,表示停留期与开花窗口的对齐程度
Z_1	总赏花价值目标
Z_2	总旅行成本目标

本文主要创建出一个公共假期赏花旅行的多目标路径优化模型^[10]。目的在于,在限定的时间与经费之下,促使游客有条理地安排起跨城市之行的行程,尽可能多地观赏到花开美景,并且也力求减少出行所耗费的成本。不同于传统的旅行商问题 (TSP) 或者最短路径问题,该任务具有时间依赖性、空间异质性以及多目标权衡的特点,需要将花开花期、景观价值、时间成本、货币预算等各方面因素一起考虑进去^[11-13]。

假设总旅行时长为 T , 候选城市数量为 N 。对于每个城市 C_i ,

其花期记为 $[s_i, e_i]$, 景观价值为 v_i , 该景观价值反映了花卉多样性和用户满意度得分。从城市 C_i 到 C_j 的交通成本和时长分别记为 d_{ij} 和 t_{ij} , 城市 C_i 中的其他停留相关成本记为 C_{i0} 。给定一条旅行路径 $P = \{C_{i_1}, C_{i_2}, \dots, C_{i_k}\}$ 及其对应的停留时间 $\{t_1, t_2, \dots, t_k\}$, 我们的目标是在若干约束条件下优化多个目标^[14]。

令每个访问城市的赏花贡献为:

$$F_i = f(t_i, s_i, e_i) \cdot v_i, \quad (1)$$

其中 $f(t_i, s_i, e_i) \in [0, 1]$ 衡量停留时间 t_i 与当地花期窗口 $[s_i, e_i]$ 的重叠程度。那么, 总奖励函数为:

$$\max Z_1 = \sum_{i \in P} F_i. \quad (2)$$

时间约束: 总时间不得超过允许的假期时长:

$$\min Z_2 = \sum_{i=1}^{k-1} d_{i,i+1} + \sum_{i=1}^k c_i. \quad (3)$$

时间约束: 总时间不得超过允许的假期时长:

$$\sum_{i=1}^k t_i + \sum_{i=1}^{k-1} t_{i,i+1} \leq T_{\max}. \quad (4)$$

预算约束: 总旅行成本不得超过可用预算 B :

$$Z_2 \leq B. \quad (5)$$

花期有效性, 若一个城市在花期之外被完全访问, 则该城市的贡献被忽略,

$$f(t_i, s_i, e_i) = 0 \Rightarrow F_i = 0. \quad (6)$$

为了对这个问题结构进行建模，我们将多城市旅行规划表示为一个加权图 $G=(V,E)$ ，其中：顶点集 $V=\{C_1, C_2, \dots, C_N\}$ 表示所有候选城市。边集 $E=\{(C_i, C_j)\}$ 表示城市之间的两两连接，由旅行成本和时间加权。每个节点包含诸如花期窗口 $[s_i, e_i]$ 、景观价值 v_i 、当地成本 c_i 和地理位置 (lat_i, lon_i) 等属性。每条边与诸如 d_{ij} 和 t_{ij} 等属性相关联。此图不仅编码了空间连通性，还嵌入了与赏花体验相关的时间和经济背景^[15]。通过利用图神经网络（GNN），我们提取城市的结构感知嵌入，这些嵌入进一步集成到多目标粒子群优化（MOPSO）算法中，以增强初始化并在搜索过程中指导粒子更新。最终，任务是找到一个城市访问序列，在满足时间和预算约束的同时最大化赏花价值并最小化总成本。下一节将详细阐述集成 GNN 和 MOPSO 的算法设计。

二、方法

为了实现假日赏花路线高效规划和智能优化，本文提出了一种集成 GNN 和 MOPSO 的路径规划算法框架整体方法流程如图1所示。



图1：方法框架图

它包含模块有输入模块（城市数据、花期窗口、预算、时间约束）、GNN 图构建与节点嵌入提取、多目标粒子编码与初始化、GNN 引导的 MOPSO 进化过程、帕累托解集输出和路径可视化模块

（一）城市图构建与特征编码

在这项工作中，我们将所有候选城市表示为一个带权无向图 $G=(V,E)$ ，其中 $V=\{v_1, v_2, \dots, v_n\}$ 表示城市集合， $E \subseteq V \times V$ 表示城市间的连通性，边权对应城市间的旅行时间和成本。每个城市节点 v_i 与一个多维属性向量 x_i 相关联，该向量包括城市的赏花价值 v_i 、花期 $[s_i, e_i]$ 、当地成本 c_i 、地理坐标 (lat_i, lon_i) ，以及可能的其他因素，如用户评分或天气预报 r_i 。这些特征表示为：

$$x_i = [v_i, s_i, e_i, c_i, lat_i, lon_i, r_i]. \quad (7)$$

为了得到本地节点属性和城市之间的结构关系，使用图注意力网络（GAT）来计算上下文感知节点嵌入。GAT 框架中每一个节点都会用到自己学到的注意力系数去聚合自己的邻居信息。节点表示的更新过程如下：

$$h_i = \sigma \left(\sum_{j \in N(i)} \alpha_{ij} W h_j \right), \quad (8)$$

其中 h_j 表示相邻节点 v_j 的特征向量， W 是一个可学习的权重矩阵， α_{ij} 是在节点 i 和节点 j 之间计算得到的注意力权重，并且 σ 是一个非线性激活函数，例如 ReLU。得到的输出 h_j 捕捉了节点级语义和拓扑依赖性。

在 GAT 传播之后，每个城市由一个高维嵌入向量 z_i 表示，该向量作为下游优化任务的基础。这些嵌入不仅编码了城市的内在属性（例如，花卉类型和开花期），还编码了其在旅行图中的相对重要性和连通性。在后续模块中，这些节点嵌入用于 MOPSO 框架内的粒子初始化、路线评估和优化指导。

（二）多目标粒子群优化建模（MOPSO）

为解决多约束和多目标下的赏花旅行规划问题，我们采用多目标粒子群优化（MOPSO）算法作为核心搜索引擎。群体中的每个粒子代表一个候选旅行方案，对访问城市的顺序和相应停留时间进行编码。具体来说，一个粒子 P 被定义为一个元组：

$$P = \{(C_{i_1}, t_1), (C_{i_2}, t_2), \dots, (C_{i_k}, t_k)\}. \quad (9)$$

为了便于计算以及用向量来表示，将粒子转换成固定长度的向量形式

$$X = [i_1, i_2, \dots, i_k, t_1, t_2, \dots, t_k], \quad (10)$$

前半部分为城市访问顺序，后半部分为停留时间。群体由 M 个这样的粒子组成，粒子被迭代更新来同时优化两个目标函数，即最大化总赏花效益和最小化总旅行成本。

第一个目标就是考虑计划访问的时间和每一个城市实际的开花期之间是否重叠，使得所有的选定城市的有效赏花奖励最大化。将该目标定义为：

$$Z_1 = \sum_{j=1}^k f(t_j, s_{ij}, e_{ij}) \cdot v_{ij}, \quad (11)$$

其中 $f(t_j, s_{ij}, e_{ij}) \in [0, 1]$ 是一个时间重叠函数，用于评估观赏时间 t_i 与城市花期 $[s_{ij}, e_{ij}]$ 的匹配程度，并且 v_{ij} 表示城市 C_{ij} 的赏花价值。

第二个目标是使市内、市际费用之和最小化，即最小化总出行成本。

$$Z_2 = \sum_{j=1}^k c_{ij} + \sum_{j=1}^{k-1} d_{ij, i_{j+1}}. \quad (12)$$

为了对这些互相矛盾的目标进行管理，用帕累托存档的方式存储在优化过程中得到的非支配解集。每一代中根据支配性和多样性来评定粒子。拥挤距离度量是沿着帕累托前沿保持多样性的一种方式，保证解集包含两个目标之间所有的权衡。

使用以下方程迭代更新每个粒子的位置和速度：

$$V^{(t+1)} = \omega V^{(t)} + c_1 r_1 (P_{\text{best}} - X^{(t)}) + c_2 r_2 (G_{\text{best}} - X^{(t)}), \quad (13)$$

$$X^{(t+1)} = X^{(t)} + V^{(t+1)}, \quad (14)$$

其中 ω 为惯性权重， c_1 和 c_2 为认知和社会学习因子， $r_1, r_2 \in [0, 1]$ 为随机数。 P_{best} 是粒子找到的个人最佳解， G_{best} 是基于拥挤距离或 GNN 引导从外部存档中选择的精英解。

多目标公式可以使得算法去寻找许多可行的旅行路线，从而达到赏花效果与总成本的平衡。用迭代优化粒子群的方法来逼近

帕累托前沿,得到一组优化的、多样化的旅行规划建议。

(三) GNN 与 MOPSO 的集成策略

为了提高粒子初始化的质量和进化搜索的过程的效率,本文把 GNN 学习得到的结构特征加到 MOPSO 优化模型里,如图2所示。

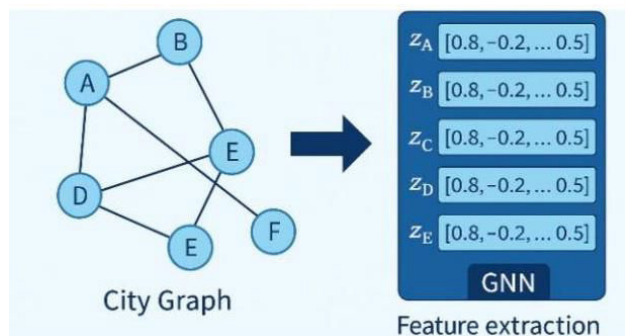


图2: 基于 GNN 从城市图中提取嵌入

GNN 的集成有两方面目的,第一,用结构感知的城市嵌入来指导高质量初始解的产生;第二,在迭代中给群体提供辅助评分信息,促使它向更有价值的赏花路线前进。

在初始化阶段, GNN 生成的节点嵌入 z_i 用于根据每个城市的连通性和属性特征评估其潜在重要性。在初始粒子生成过程中,基于嵌入得分较高的城市被赋予更高的优先级。这有助于生成一个初始群体,该群体不是纯粹随机产生的,而是偏向于结构上有利且语义上有意义的旅行序列。

迭代优化阶段用 GNN 引导的评分函数来评价每一个候选路线,该函数把已经访问过的城市嵌入得分聚合起来,再根据它的开花效果进行调制。评分函数定义为:

$$Score(P) = \sum_{j=1}^k softmax(z_{i_j}) \cdot f(t_j, s_{i_j}, e_{i_j}). \quad (15)$$

该评分函数被加入到外部存档的选择中。当多个非支配解竞争存储或者指导粒子更新的时候,具有高 GNN 引导得分的解会被优先考虑。用这种方法, MOPSO 的搜索轨迹向更好的结构定位、更高的预测效用解趋近,加快收敛速度,使最后的帕累托前沿更加多样化。

通过把基于 GNN 的表示学习同 MOPSO 的进化过程紧密结合起来,提出的方法用拓扑感知和自适应优化的方式,更高效、智能地解决了赏花路径规划的问题。

三、实验设计与结果分析

为了检验本文提出的 GNN-MOPSO 方法在优化假日赏花路径方面是否有效,本文做了若干实验,城市规模、预算约束、花期分布等都不同。用各种主流的优化算法进行实验,从路径质量、目标值性能、解集多样性、计算效率等几个方面对算法性能进行评价。

(一) 实验设置

本文创建了一个包含20个城市、不同的地理位置、交通联系、花卉种类分布以及观赏评分的城市赏花旅行图。

每个城市被赋予一个开花窗口 $[s_i, e_i]$ 、一个赏花价值 v_i 、城市间旅行成本 d_{ij} 、住宿成本 c_i , 以及诸如天气评级和用户满意度得分等辅助变量。

城市之间网络是由高铁连通性所形成的。旅行时间、成本根据实际区域数据做归一化处理。假日时长: 72小时(3天)。预算限制: 1500单位。最多访问城市数量: 6。

在此设置下对所提方法进行评价,并与下面四种算法进行比较,如表2所示。

表2: 测试集上的模型性能比较

算法	描述
GNN-MOPSO	本文提出的方法,将 GNN 嵌入与 MOPSO 相结合。
标准粒子群优化算法	没有 GNN 和多目标建模的基线粒子群优化算法。
非支配排序遗传算法 II	经典多目标遗传算法。
贪心 + 回溯	基于花朵得分 / 开花时间排序和剪枝的启发式算法。

所有实验均在相同的计算平台(英特尔 i7、16GB 内存、Python 3.10、PyTorch Geometric)上进行,并且每种算法运行10次以获得平均结果。

(二) 评估指标

为了全面评估算法的性能,本文使用以下评估指标:

$$R_f = \frac{1}{k} \sum_{j=1}^k f(t_j, s_{i_j}, e_{i_j}). \quad (16)$$

总赏花价值 Z_1 在3.2节中定义,用于衡量审美和体验效益。

总路径成本 Z_2 包括交通 + 住宿,单位越小越好。

解集拥挤指数 (Spacing) 用来度量帕累托解集的多样性以及分布是否均衡,其定义如下

$$S = \frac{1}{|P|-1} \sum_{i=1}^{|P|} |d_i - \bar{d}|. \quad (17)$$

平均运行时间,用于衡量算法的计算效率(秒)。

(三) 实验结果与比较分析

表3展示了四种算法(包括 GNN-MOPSO、PSO、NSGA-II 和贪心 + 回溯算法)在路径质量方面的比较。评估指标包括平均赏花价值 (Z_1)、平均旅行成本 (Z_2)、平均花期重叠率 (R_f) 以及衡量帕累托解多样性的 Spacing 指标。可以观察到, GNN-MOPSO 实现了最高的赏花价值和花期重叠率,同时保持了具有竞争力的旅行成本和最佳的多样性表现。虽然贪心 + 回溯算法实现了最低的成本,但其赏花效益和帕累托多样性明显较差证明了所提方法的整体优越性。

表3: 路径质量比较

算法	平均 Z_1	平均 Z_2	平均 R_f	间距
GNN-MOPSO	53.42	1260.3	0.84	0.013
PSO	47.31	1294.6	0.76	0.021
NSGA-II	51.28	1272.5	0.81	0.018
贪心 + 回溯	42.19	1247.0	0.69	0.045

图3展示了四种算法在总赏花价值 (Z_1) 和总旅行成本 (Z_2) 之间的帕累托前沿比较。GNN - MOPSO (蓝色圆圈) 在赏花价值和旅行成本之间实现了最佳平衡,展现出高赏花价值和低旅行

成本，且解的分布相对均匀，证明了其优越性。相比之下，虽然标准 PSO（红色方块）对某些解提供了较低的旅行成本，但它的赏花价值较低且解的多样性较差。NSGA - II（绿色三角形）在赏花价值和旅行成本之间实现了平衡，但其整体性能略逊于 GNN - MOPSO。贪心回溯（紫色叉号）实现了最低的旅行成本，但其赏花价值和解的多样性明显不足，无法有效提供高质量的帕累托解集。

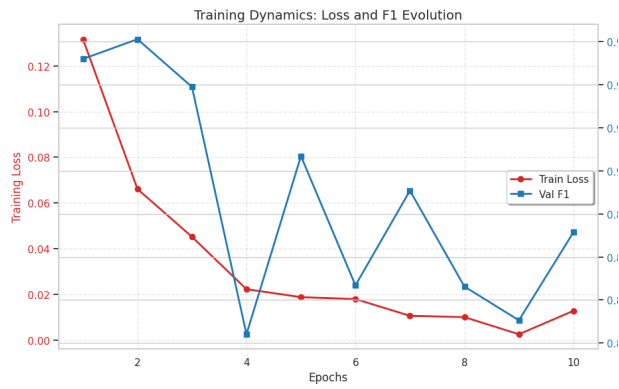


图3：帕累托前沿比较（ Z_1 与 Z_2 ）

表4展示了不同算法（包括 GNN - MOPSO、PSO、NSGA - II 和贪心 + 回溯）之间的运行效率比较。表中显示了每种算法的平均运行时间，GNN - MOPSO 需要18.3秒，PSO 需要9.1秒，NSGA - II 需要23.5秒，而贪心 + 回溯最快，为2.8秒。

表4：运行效率比较

算法	平均运行时间（秒）
GNN-MOPSO	18.3
PSO	9.1
NSGA-II	23.5
贪婪 + 回溯	2.8

虽然和传统的算法相比，GNN-MOPSO 的运行时间要长得多，但是它的解质量、多样性都很好，因此更适合于大规模的城市网络优化。

实验结果清楚地表明，本文所提出的 GNN - MOPSO 方法对于多城市赏花路线优化任务是有效的、有优势的。多目标权衡、复杂结构建模、解集分布控制，它比传统的优化方法好。GNN 模块给出的结构感知嵌入可以提高搜索空间引导的能力，减少无效路径的探索，加快收敛速度，增大帕累托前沿的宽度。

当城市数量增多到50个以上的时候，传统的解法质量会明显降低，但是我们提出的方法仍然可以保持稳定的性能，具有较好的可扩展性。另外 GNN 模块加入了部分模型可解释性，节点注意力权重以及路径得分来源可以被用作决策支持以及路线推荐系统的支撑。

四、结论与未来工作

本文提出了一种基于 GNN 和 MOPSO 相结合的路线优化方法，专门用于解决假日赏花旅游城市游览和路线规划问题。将开花期、交通成本、停留时间等各个约束条件加入到目标函数中，可以很快得到最优路线，也可以较好地兼顾各个目标。GNN 模型可以从中提取出城市网络的结构特征，使路线搜索更加智能。MOPSO 用模拟粒子群的进化过程来完成多目标优化，即在控制旅行成本的同时达到最大的赏花效益。该种将图学习与多目标优化相结合的算法，在计算机算法领域有着较大的应用前景。

虽然 GNN-MOPSO 的计算效率比传统的 PSO、NSGA-II 等要高一些，但是它的解质量、多样性都比其它算法好得多。从实验结果可知，GNN-MOPSO 既能保证较低的旅行成本，又能较好地改善赏花效益，对大城市的网络也具有较好的优化能力。另外，由于城市数量的增多，传统的算法已经不能满足需要。但是 GNN-MOPSO 可以有效地解决大规模的问题，并且有很好的可扩展性以及鲁棒性。因此，本文研究结果表明，粒子群优化算法和图神经网络相结合，给解决复杂的城市路径规划问题提供了一种新的方法，有很强的实际应用价值。

未来研究可以在几个方面进一步优化和扩展现有方法。虽然目前的算法表现很好，但是它的计算时间仍然比传统的算法要长。未来的研究可以进一步研究更加高效并且适合 GPU 的并行计算方法，从而提高计算速度。其次，用强化学习（RL）来实现路径规划的自适应优化，可以使得算法对于实时环境条件（天气、交通等）作出更加灵活的改变。最后随着智慧城市、物联网的发展，可以将实时交通流量、用户偏好等更多的多模态数据加入到路径优化算法当中来，从而提高它在实际应用中的准确性以及鲁棒性。经过上述改进之后，GNN-MOPSO 算法在更加复杂的、动态的应用环境中会发挥出更大的作用。

参考文献

- [1]McNulty A ,Berman O B ,Engelbrecht A .A comparative study of evolutionary algorithms and particle swarm optimization approaches for constrained multi-objective optimization problems[J].Swarm and Evolutionary Computation,2024,91101742-101742.DOI:10.1016/J.SWEVO.2024.101742.
- [2]Dezvarei M ,Tomovic K ,Sun S J , et al.A graph neural network framework for security assessment using topological measures[J].Electric Power Systems Research,2025,249111972-111972.DOI:10.1016/J.EPSR.2025.111972.
- [3]Kajla I N ,Missen S M M ,Coustaty M , et al.A histogram-based approach to calculate graph similarity using graph neural networks[J].Pattern Recognition Letters,2024,186286-291.DOI:10.1016/J.PATREC.2024.10.015.
- [4]Zhao Y ,Dong J ,Wang W , et al.A multi-typed multi-relational heterogeneous graph neural network model for complex networks[J].Knowledge-Based Systems,2025,329(PA):114291-114291.DOI:10.1016/J.KNOSYS.2025.114291.

- [5]Quintero H P , Triviño G P , Gonz á lez C D , et al.Adaptive multi-objective real-time hierarchical control for isolated microgrid clusters utilizing an enhanced particle swarm optimization strategy to optimize costs and emissions[J].Electric Power Systems Research, 2026, 250112169–112169.DOI: 10.1016/J.EPSR.2025.112169.
- [6]Meng X , Li H .An adaptive co-evolutionary competitive particle swarm optimizer for constrained multi-objective optimization problems[J].Swarm and Evolutionary Computation, 2024, 91101746–101746.DOI: 10.1016/J.SWEVO.2024.101746.
- [7]Cunha B C , Massarotto F D , Fornazza L S , et al.An ALNS metaheuristic for the family multiple traveling salesman problem[J].Computers and Operations Research, 2024, 169106750–106750.DOI: 10.1016/J.COR.2024.106750.
- [8]Luo X , Yi Z .Efficiency management of engineering projects based on particle swarm multi objective optimization algorithm[J].Systems and Soft Computing, 2025, 7200320–200320.DOI: 10.1016/J.SASC.2025.200320.
- [9]Yang X , Wang R , Li K , et al.Exploratory landscape analysis on black-box optimization problems via Graph Neural Network[J].Swarm and Evolutionary Computation, 2025, 99102136–102136.DOI: 10.1016/J.SWEVO.2025.102136.
- [10]Wu J , Xu Z , Qiao S .Fast Heterogeneous Graph Neural Network Generation via Meta Contrastive Learning[J].Neural networks : the official journal of the International Neural Network Society, 2025, 192107727.DOI: 10.1016/J.NEUNET.2025.107727.
- [11]Wang X , Lv Y , Sun H , et al.Multi-modal travel route planning considering environmental preference under uncertainties: A distributionally robust optimization approach[J].Transportation Research Part E, 2025, 198104097–104097.DOI: 10.1016/J.TRE.2025.104097.
- [12]Zhao Y , Wang W , Wang S , et al.Over-smoothing problem of heterogeneous graph neural networks: A heterogeneous graph neural network with enhanced node differentiability[J].Information Processing and Management, 2026, 63(2PA): 104395–104395.DOI: 10.1016/J.IPM.2025.104395.
- [13]Dinçer H , Y ü ksel S , Eti S , et al.Q-learning algorithm and molecular fuzzy multi-objective particle swarm optimization-based decision-making approach to circular economy-oriented investment alternatives for renewable energy technologies[J].Information Sciences, 2025, 718122378–122378.DOI: 10.1016/J.INS.2025.122378.
- [14]Lin C , Xie Y , Wang H C .Significant wave height prediction at multiple sites using sequence decomposition and dynamic spatiotemporal graph neural networks[J].Ocean Engineering, 2025, 341(P2): 122548–122548.DOI: 10.1016/J.OCEANENG.2025.122548.
- [15]Feng B , Zhou P X .The novel physics-enhanced graph neural network for phase-field fracture modelling[J].Computer Methods in Applied Mechanics and Engineering, 2025, 446(PB): 118284–118284.DOI: 10.1016/J.CMA.2025.118284.

广东省绿色低碳创新效率的统计测度与时空演化分析

张倩生*, 朱曹了了

广东外语外贸大学 数学与统计学院, 广东 广州 510006

DOI:10.61369/ASDS.2026060012

摘 要 : 本文以2014–2024年广东省21个地级市的面板数据为样本, 采用考虑非期望产出的超效率SBM–DEA模型测度广东省各地市的绿色低碳创新效率, 并结合Malmquist–Luenberger (ML) 指数分析其动态演变趋势。进而, 还检验考察了广东省绿色低碳创新效率的区域收敛性, 最后通过Tobit回归模型探究识别影响广东绿色低碳创新效率的关键因素。研究表明, 广东省绿色低碳创新效率整体呈波动上升趋势, 但区域差异明显, 珠三角绿色创新效率最高, 粤西地区最低; ML指数分解显示, 技术进步是推动广东低碳创新效率提升的主要动力, 技术效率变化的作用相对有限; 收敛性检验发现, 广东省绿色低碳创新效率存在显著的 σ 收敛和 β 收敛, 区域差距有缩小趋势。建议通过扩大对外开放程度, 并提升产业结构和各地城镇化水平, 有助于广东省绿色经济高质量发展和低碳目标的实现。

关 键 词 : 绿色低碳创新; 超效率DEA模型; ML指数; 收敛性; Tobit回归

Statistical Measurement and Spatiotemporal Evolution Mechanisms of Green and Low-Carbon Innovation Efficiency in Guangdong Province

Zhang Qiansheng*, Zhu Caoliaoliao

School of Mathematics and Statistics, Guangdong University of Foreign Studies, Guangzhou, Guangdong 510006

Abstract : This study takes panel data from 21 cities in Guangdong Province for the period 2014–2024 as its sample. We employ the super-efficiency SBM–DEA model with undesirable output to measure the green and low-carbon innovation efficiency of these cities, and further analyzes its dynamic evolution using the Malmquist–Luenberger (ML) index. In addition, the regional convergence of low-carbon innovation efficiency in Guangdong Province is examined. Finally, a Tobit regression model is applied to identify the key factors influencing this efficiency. The results indicate that while the overall green and low-carbon innovation efficiency in Guangdong Province shows a fluctuating upward trend, significant regional disparities exist: the Pearl River Delta region exhibits the highest efficiency, whereas western Guangdong ranks the lowest. Decomposition of the ML index reveals that technological progress is the primary driver of efficiency improvements, while changes in technical efficiency play a relatively limited role. Convergence tests suggest significant σ -convergence and β -convergence in green and low-carbon innovation efficiency, indicating a narrowing trend in regional disparities. Policy recommendations include expanding Guangdong Province's openness to the outside world, upgrading its industrial structure, and enhancing the urbanization level across regions. These measures facilitate the development of green economy and the achievement of its low-carbon goal.

Keywords : green and low-carbon innovation; Super-efficiency DEA model; Malmquist–Luenberger index (ML); convergence; Tobit regression

引言

在全球气候变暖加上资源环境约束日益加剧的背景下, 绿色低碳创新已成为推动经济高质量发展、实现“碳达峰、碳中和”目标的核心动力。广东省作为中国经济第一大省, 既是制造业与科技创新的高地, 也面临较大的能源消耗和污染物排放所带来的压力。如何测度广东绿色低碳效率并在保证经济增长质量的同时, 用更少的资源与环境代价去获得更多的创新与经济产出, 是广东省乃至全国可持续发展面临的重要问题。目前关于绿色创新效率的测度方法主要分为随机前沿法(郭树华^[1])与数据包络分析(DEA)两大类。特别地

基金项目: 2025年度广东省自然科学基金项目“特征驱动下基于机器学习的投资决策研究”(2025A1515010937); 2024年度广东省社科规划办项目“不确定需求下动态设施布局的多目标滚动优化研究”(GD24CGL54)。

作者简介: 朱曹了了, 广东外语外贸大学数学与统计学院, 硕士研究生, 研究方向为数据驱动与统计优化。

通讯作者: 张倩生, 广东外语外贸大学数学与统计学院, 教授, 研究方向为智能决策。

Tone^[2]提出超效率SBM-DEA模型,能有效处理非期望产出并考虑投入冗余和产出不足问题。Wang等^[3]和钱丽等^[4]分别基于共同前沿DEA方法,测算了群组前沿与共同前沿下的碳减排效率和绿色创新效率。为处理CO₂排放等数据估计误差和不确定性, Ignatius等^[5]提出了基于模糊DEA的碳效率评价框架; Zha等^[6]采用机会约束DEA模型对区域能源效率评估。张辽^[7]采用三阶段SBM-DEA模型测度了工业企业真实绿色技术创新效率。张盼^[8]运用四阶段DEA-超效率SBM模型测度了中国30个省份的真实绿色创新效率,并通过泰尔指数分析了区域差异。赵雅婷^[9]运用超效率网络SBM-DEA模型对中国绿色技术创新进行效率测度和时空演化分析。Zhang等^[10]提出了基于公平视角的交叉效率网络DEA测度方法。赵志芳^[11]运用超效率SBM模型测度探究全国30个省份绿色技术创新效率的区域差异。滕琳娜^[12]也采用超效率SBM模型测度省际工业绿色创新效率的区域差异。刘子豪等^[13]运用Dagum基尼系数分析了省域绿色创新效率的差异。张平凤^[14]和王金河^[15]探究了中国绿色创新效率在空间上存在的收敛特征。

近年来,学者们还从政府、市场、社会与开放等多个维度探讨了绿色创新效率的影响因素。钱丽等^[4]运用随机效应的Tobit模型,分别从技术差距无效率和管理无效率两个层面分析影响因素。在区域层面,赵志芳^[11]运用Tobit模型检验影响因素,发现地区经济发展水平、政府支持力度、环境规制和产业结构均与效率呈显著正相关。刘子豪等^[13]同样采用Tobit模型发现经济发展水平、市场竞争强度、对外开放程度和政府扶持力度可提高绿色创新效率,而环境规制因挤占企业资源呈现抑制作用。

现有研究大多集中在全国或重点城市群层面,针对广东省地级市层面的系统性研究相对较少,尤其是在效率的收敛性检验和影响因素的综合分析方面仍存在不足。因此,本文的贡献是将考虑非期望产出的超效率SBM-DEA模型,利用2014-2024年广东省21个地级市的面板数据,从城市和地域板块两个维度系统测度绿色低碳创新效率,并引入ML指数刻画其动态演变过程,同时结合 σ 收敛与 β 收敛与Tobit回归,全面揭示广东绿色低碳创新效率的区域差异和时空演化机理,旨在为促进广东区域绿色协同发展提供科学理论支撑与实践建议,为广东省绿色转型和区域创新资源优化配置、制定差异化绿色政策提供实证依据。

一、广东省绿色低碳创新效率测度模型设计

(一) 考虑非期望产出的超效率SBM-DEA模型构建

在绿色创新效率的测度研究中,传统DEA模型要求所有投入或产出都按同一比例进行调整,易使效率评价产生偏差。针对这一不足,本文采用如下超效率SBM-DEA模型(1)来评估广东省各地市的绿色低碳创新效率,对于处于效率前沿的有效决策单元先把它们从参照集中剔除,再基于剩余单元构建新的前沿面进行重新测算,这样有利于各地市效率区分和排序;它既考虑了非期望产出,又能获得投入和产出的松弛变量,以期提升效率落后地区的技术管理。

$$\rho_k^* = \min \frac{1 - \frac{1}{m} \sum_{i=1}^m \frac{s_i^-}{x_{ik}}}{1 + \frac{1}{s_1 + s_2} \left(\sum_{r=1}^{s_1} \frac{s_r^+}{y_{rk}} + \sum_{l=1}^{s_2} \frac{s_l^-}{z_{lk}} \right)} \quad (1)$$

$$\text{s.t.} \begin{cases} \sum_{j=1, j \neq k}^n \lambda_j x_{ij} + s_i^- = x_{ik}, & i=1, \dots, m \\ \sum_{j=1, j \neq k}^n \lambda_j y_{rj} - s_r^+ = y_{rk}, & r=1, \dots, s_1 \\ \sum_{j=1, j \neq k}^n \lambda_j z_{lj} + s_l^- = z_{lk}, & l=1, \dots, s_2 \\ \lambda_j \geq 0, & s_i^- \geq 0, s_r^+ \geq 0, s_l^- \geq 0 \end{cases}$$

其中, ρ_k^* 为广东省被评价的第 k 个地市(决策单元)的综合技术创新效率, x_{ij} 表示第 j 个决策单元(地市)的第 i 种投入量, y_{rj} 表示第 j 个决策单元的第 r 种期望产出量, z_{lj} 表示第 j 个决策

单元的第 l 种非期望产出量, s_i^- 、 s_r^+ 和 s_l^- 分别为投入、期望产出与非期望产出的松弛变量, λ 为广东省各地市的资源权重向量。

(二) 构建 Malmquist-Luenberger 指数模型

超效率SBM-DEA模型主要测度的是各年份的静态效率,无法反映效率随时间的变化趋势。同时,广东省绿色低碳创新效率包含非期望产出,传统Malmquist指数无法处理这类问题。因此,本文采用基于方向性距离函数的Malmquist-Luenberger(ML)指数,考察广东省绿色低碳创新效率的动态演变。

为了定义ML指数,首先需要构建方向性距离函数。设定生产可能性集为 $P^1(x^1) = \{(y^1, b^1): x^1 \text{ 可以生产 } (y^1, b^1)\}$, 其中 y 表示期望产出, b 表示非期望产出。方向性距离函数为 $\bar{D}(x^1, y^1, b^1; g_y, -g_b) = \sup \{\beta: (y^1 + \beta g_y, b^1 - \beta g_b) \in P^1(x^1)\}$ 。式中, g_y 和 g_b 分别为好产出的扩张方向和坏产出的缩减方向; β 表示在给定方向上可以同时扩张好产出、缩减坏产出的最大比例。该函数能够恰当地描述“增加期望产出的同时减少非期望产出”这一绿色生产目标。

基于方向性距离函数,以第 t 期和第 $t+1$ 期的技术为参照, ML指数定义为:

$$ML_t^{t+1}(x^{t+1}, y^{t+1}, b^{t+1}; x^t, y^t, b^t) = \left[\frac{\bar{D}^t(x^{t+1}, y^{t+1}, b^{t+1})}{\bar{D}^t(x^t, y^t, b^t)} \times \frac{\bar{D}^{t+1}(x^{t+1}, y^{t+1}, b^{t+1})}{\bar{D}^{t+1}(x^t, y^t, b^t)} \right]^{\frac{1}{2}}$$

其中ML指数还可以进一步分解为技术效率变化(EC)和技术进步变化(TC):

$$ML_t^{t+1} = \frac{\bar{D}^t(x^{t+1}, y^{t+1}, b^{t+1})}{\bar{D}^t(x^t, y^t, b^t)} \times \left[\frac{\bar{D}^t(x^{t+1}, y^{t+1}, b^{t+1})}{\bar{D}^{t+1}(x^{t+1}, y^{t+1}, b^{t+1})} \times \frac{\bar{D}^{t+1}(x^t, y^t, b^t)}{\bar{D}^{t+1}(x^t, y^t, b^t)} \right] = EC \times TC \quad (2)$$

当 $ML > 1$ 时,表明广东绿色全要素生产率提高;当 $ML = 1$ 时,表明绿色全要素生产率不变;当 $ML < 1$ 时,表明绿色全要素

生产率下降。当技术效率变化 $EC>1$ 表明技术效率改善, 存在追赶效应; 当技术进步变化 $TC>1$ 表明技术进步, 存在生产前沿面外移。

(三) 广东省绿色低碳创新效率的区域收敛性检验

绿色创新效率收敛性检验旨在考察落后地区是否具有追赶先进地区的趋势, 是分析区域差异动态演变的重要工具。本文从 σ 收敛和 β 收敛两个维度展开检验, 其中 β 收敛又可以进一步分为绝对 β 收敛和条件 β 收敛。

1. 广东省绿色低碳创新效率的 σ 收敛检验

σ 收敛是用来判断广东省各地区绿色低碳创新效率的差距是否随着时间推移而缩小。本文采用变异系数 (CV) 作为衡量指标, 其计算公式为:

$$CV_t = \frac{\sigma_t}{\mu_t} \tag{3}$$

其中, σ_t 为第 t 年广东省各地区效率值的标准差, μ_t 为第 t 年广东省绿色创新效率的均值, 若 CV_t 随时间呈下降趋势, 则表明存在 σ 收敛, 即区域间的效率差距在缩小; 反之, 区域间的差距在扩大。

2. 广东省绿色低碳创新效率的绝对 β 收敛检验

绝对 β 收敛检验假定广东所有地区具有相同的稳态水平, 旨在探究在不考虑其他外部条件的情况下, 初期效率较低的地区会不会有更高的增长率, 从而慢慢追赶先进地区。本文设定如下面板回归模型:

$$\frac{1}{T} \ln \left(\frac{EFF_{i, t+T}}{EFF_{i, t}} \right) = \alpha + \beta \ln (EFF_{i, t}) + \varepsilon_{i, t} \tag{4}$$

其中, 被解释变量为地区 i 在 t 到 $t+T$ 期间绿色低碳创新效率的年均增长率, 核心解释变量为初期效率水平的对数。若回归系数 β 显著为负, 则表明初期效率越低的地区, 其后期的增长越快, 存在绝对 β 收敛; 反之, 若 β 显著为正, 则表明初期效率越高的地区增长越快, 区域差距呈现扩大趋势, 存在绝对 β 发散。考虑到样本时间跨度为 11 年, 本文将 T 设为 1, 采用年度增长率进行面板固定效应回归。

3. 广东省绿色低碳创新效率的条件 β 收敛检验

在现实中, 广东省各地区在产业结构、经济发展水平等结构性因素存在的显著差别会影响到各地区效率长期的稳态水平, 进而制约了收敛效应的发生。条件 β 收敛检验正是为了解决这一问题而设计出来的措施, 它在回归模型中引入一些控制变量, 以此来考察在给定那些结构性差异的条件下, 落后地区是否具有更高的增长率。本文借鉴已有研究的做法, 设定如下条件 β 收敛面板回归模型:

$$\frac{1}{T} \ln \left(\frac{EFF_{i, t+T}}{EFF_{i, t}} \right) = \alpha + \beta \ln (EFF_{i, t}) + \gamma X_{i, t} + \mu_i + \lambda_t + \varepsilon_{i, t} \tag{5}$$

其中, 被解释变量为广东省地区 i 在 t 到 $t+T$ 期间绿色低碳创新效率的年均增长率, 核心解释变量为初期效率对数 $\ln (EFF_{i, t})$, $X_{i, t}$ 为控制变量向量, 包括影响绿色创新效率的结构性因素, μ_i 和 λ_t 分别为个体固定效应和时间固定效应。若回归系数

β 显著为负, 则表明存在条件 β 收敛, 即结构性因素只能部分解释增长差异, 收敛效应具有独立存在的稳健性; 反之, 则表明初始差异完全可由结构性因素解释, 不存在独立追赶趋势。

(四) 基于 Tobit 回归模型探究广东省绿色低碳创新因素

为探究广东省绿色低碳创新效率的影响因素, 需要建立一个回归模型。由于超效率 SBM 模型测算出来的广东各地市的绿色低碳创新效率值会有一个下限截断的情况, 也就是效率值会大于或等于零, 这就属于一种受限因变量, 如果采用普通最小二乘法 (OLS) 进行回归估计, 会导致得到的参数估计变得有偏且不一致。所以本文采用如下 Tobit 回归模型 (6) 对广东省绿色低碳创新效率进行影响因素分析。

$$Y_{it}^* = \alpha + \beta X_{it} + \varepsilon_{it}, \tag{6}$$
$$Y_{it} = \begin{cases} Y_{it}^*, Y_{it}^* > 0; \\ 0, Y_{it}^* \leq 0. \end{cases};$$

其中, Y_{it} 为可观测的被解释变量 (广东省城市 i 在 t 年的绿色低碳创新效率), Y_{it}^* 为潜变量, α 为截距项, X_{it} 为解释变量向量, β 为待估系数, ε_{it} 为随机扰动项。

二、广东省绿色低碳创新效率相关评价指标

(一) 广东省绿色低碳创新的投入产出指标

在广东省绿色低碳创新投入产出指标方面, 投入指标选择 R&D 经费内部支出、R&D 人员全时当量、能源消耗总量, 用来说明资金投入、劳动力投入和地区的资源消耗。期望产出指标选择发明专利授权量作为知识产出, 其能够有效反映创新成果的技术含量与实质性突破, 规避低质量专利对创新效率测算的干扰; 新产品销售收入作为经济产出指标, 直接度量了绿色创新成果的市场化转化能力, 能精准反映创新活动对区域经济增长的实际贡献。非期望产出指标选取工业 SO_2 排放量、工业废水排放量、一般工业固体废物产生量, 能够反应地区的环境污染状况; 详细投入产出指标如表 1 所示。

表 1: 广东省绿色低碳创新效率评价的投入和产出指标

指标类型	指标构成	变量	单位
投入	资金	R&D 经费内部支出	万元
	劳动力	R&D 人员全时当量	人年
	能源	能源消耗总量	万吨标准煤
	知识产出	发明专利授权量	件
产出	经济产出	新产品销售收入	万元
		工业 SO_2 排放量	吨
	非期望产出	工业废水排放量	万吨
		一般工业固体废物产生量	万吨

(二) 广东省绿色低碳创新效率条件 β 收敛的控制变量指标

在针对广东省绿色低碳创新效率的条件 β 收敛检验中, 我们选取广东各地市经济发展水平 ($\ln pgdp$)、产业结构 (is)、城镇化水平 ($urban$)、对外开放程度 ($open$)、市场竞争强度 ($\ln imc$) 作为控制变量 (如表 2), 旨在剔除各地区结构性差异对

绿色低碳创新效率收敛趋势的干扰，从而检验在给定这些结构性特征后，落后地区是否仍然具有更快的增长速率。其中，经济发展水平用人均 GDP 来表示，能够反映地区经济发展阶段和要素禀赋。产业结构以第三产业增加值占 GDP 比重来衡量，第三产业占比的提升，通常意味着更低的能耗与更高的附加值，能通过产业协同与技术外溢加速绿色创新成果转化。城镇化率反映了人口与经济活动的空间集聚程度，因此选取城镇人口占常住人口比例衡量城镇化水平，考察其对绿色创新效率收敛的影响。对外开放程度以地区货物进出口总额与 GDP 之比衡量，反映了地区参与国际技术溢出和市场竞争的深度。市场竞争强度以规模以上工业企业个数对数近似表征的市场竞争程度，会影响企业进行绿色技术创新的激励和资源获取能力。

表 2：控制变量指标

变量名称	变量说明
经济发展水平（lnpgdp）	地区人均 GDP 的对数
产业结构（is）	第三产业增加值占 GDP 比重
城镇化水平（urban）	地区城镇人口占常住人口的比例
对外开放程度（open）	地区货物进出口总额与地区 GDP 之比
市场竞争强度（lnimc）	地区规模以上工业企业个数的对数

（三）广东省绿色低碳创新效率的影响因素指标

为探究广东省绿色创新效率的影响因素，我们选取经济发展水平（lnpgdp）、产业结构（is）、城镇化水平（urban）、对外开放程度（open）、市场竞争强度（lnimc）、政府支持（gov）作为研究变量（如表 3 所示）。然后通过 Tobit 回归，识别出哪些因素会显著地促进或抑制广东省各地级市的绿色创新效率，这为政策制定提供实证依据。其中，政府支持（gov）以一般预算内财政收入与财政支出之比来衡量，比值越高表明政府财政状况越好、对绿色创新的直接扶持能力可能越强。但该指标同时也反映了政府对经济的干

预程度和支出效率，若财政支出主要投向传统基建而非绿色技术，或存在低效补贴，则可能对绿色创新产生“挤出效应”。

表 3：广东省绿色低碳创新效率的影响因素指标

变量类型	变量名称	变量说明
被解释变量	绿色创新效率（gtie）	各城市的绿色低碳创新效率值
	经济发展水平（lnpgdp）	地区人均 GDP 的对数
	产业结构（is）	第三产业增加值占 GDP 比重
	城镇化水平（urban）	地区城镇人口占常住人口的比例
解释变量	对外开放程度（open）	地区货物进出口总额与地区 GDP 之比
	市场竞争强度（lnimc）	地区规模以上工业企业个数的对数
	政府支持（gov）	一般预算内财政收入与一般预算内财政支出之比

（四）指标数据来源

本文选取 2014–2024 年广东省 21 个地级市的面板数据作为研究样本。数据取自历年的《中国统计年鉴》《中国城市统计年鉴》《广东省统计年鉴》及各市统计年鉴、国民经济和社会发展公报、环境公报等，对于个别年份的缺失数据，采用线性插值法进行补充。

三、实证测度分析

（一）广东省绿色低碳创新效率的整体演变分析

利用考虑非期望产出的超效率 SBM–DEA 模型（1），本研究运用 MATLAB 软件测算 2014–2024 年广东省 21 个地级市的绿色低碳创新效率，结果如下表 4 所示。

表 4：2014–2024 年广东省各城市的绿色低碳创新效率

城市	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	平均值
深圳市	1.623	1.631	1.562	1.625	1.574	1.531	1.505	1.450	1.438	1.453	1.462	1.532
东莞市	1.008	1.019	1.090	1.130	1.238	1.177	1.160	1.098	1.082	1.087	1.074	1.106
汕尾市	1.039	1.008	1.060	1.031	1.072	1.111	1.187	1.077	1.014	1.111	1.085	1.072
珠海市	1.016	1.063	1.171	1.104	1.087	0.704	1.084	1.100	1.118	1.039	1.065	1.050
惠州市	1.122	1.121	1.076	1.036	1.020	1.027	1.005	1.013	1.017	0.475	0.432	0.940
河源市	1.045	1.019	1.033	1.020	1.050	0.630	0.658	0.374	0.468	1.029	0.322	0.786
佛山市	0.617	0.555	0.590	0.655	0.642	0.498	0.601	1.038	1.033	1.053	1.054	0.758
清远市	0.305	1.022	0.413	1.018	0.330	0.415	0.442	1.077	1.086	0.280	1.014	0.673
广州市	0.429	0.364	0.401	0.435	0.580	0.486	0.551	1.023	1.004	1.028	1.014	0.665
中山市	0.221	0.291	0.258	0.413	0.563	1.129	0.325	0.515	0.461	0.452	1.026	0.514
汕头市	1.022	1.026	0.314	0.287	0.238	0.225	0.214	0.313	0.308	0.299	0.334	0.416
梅州市	0.177	0.203	0.245	0.358	0.293	1.152	0.397	0.407	0.430	0.362	0.352	0.398
江门市	0.287	0.319	0.357	0.314	0.242	0.257	0.279	0.383	0.449	0.421	0.399	0.337
肇庆市	0.227	0.262	0.383	0.191	0.238	0.233	0.339	0.464	0.440	0.334	0.372	0.317
潮州市	0.328	0.282	0.198	0.232	0.205	0.265	0.243	0.349	0.341	0.300	0.403	0.286
湛江市	0.275	0.259	0.193	0.255	0.302	0.304	0.240	0.286	0.257	0.265	0.291	0.266
阳江市	0.083	0.273	0.060	0.114	0.329	0.137	0.221	0.311	0.353	0.361	0.491	0.249
揭阳市	0.270	0.293	0.271	0.223	0.222	0.167	0.099	0.260	0.236	0.273	0.364	0.243
韶关市	0.201	0.147	0.126	0.156	0.171	0.132	0.168	0.307	0.400	0.361	0.355	0.229
云浮市	0.100	0.159	0.089	0.098	0.238	0.252	0.283	0.246	0.235	0.213	0.258	0.197
茂名市	0.257	0.179	0.107	0.095	0.132	0.103	0.108	0.148	0.189	0.205	0.113	0.149
平均值	0.555	0.595	0.524	0.561	0.560	0.568	0.529	0.630	0.636	0.591	0.632	0.580

1. 广东省绿色低碳创新效率整体变化趋势分析

从时间上看，广东省绿色低碳创新效率呈波动上升态势，但不同阶段差异明显。2014–2020年期间，效率均值在0.52至0.60之间反复波动，未形成持续增长趋势，2021–2024年，效率均值出现明显跃升，从2020年的0.529升至2021年的0.630，2022年达到峰值0.636，2023年回落至0.591，2024年又回升至0.632，这表明在“双碳”目标政策驱动下，广东省绿色创新效率有明显的提升。

从整体来看，2014–2024年广东省绿色低碳创新效率的总均值为0.580，这说明广东省在绿色低碳创新方面虽然取得了一定的进展，但还没有达到生产前沿面的有效状态，存在投入冗余或产出不足的问题，绿色创新效率还有巨大提升空间。

2. 广东省各城市绿色低碳创新效率演变分析

表4将广东省21个地级市2014–2024年绿色创新效率由高到低进行排序，可看出广东省各城市的绿色创新效率发展极不平衡。深圳、东莞、汕尾、珠海这四个城市效率较高，其中深圳遥遥领先，历年效率均大于1.45，表现出绝对优势。惠州、河源、佛山、清远、广州、中山处于中等水平，效率均值在0.5–1.0之间。而汕头、梅州、江门等城市效率较低，特别是茂名、云浮、韶关等城市，多数年份效率不足0.3，绿色创新效率存在巨大提升空间。从时间纵向来看，绝大部分城市的效率在这11年内表现出波动中略有上升的趋势。深圳、东莞等城市效率值始终大于1，并且年度波动也比较小，这显示出它们有着持续的创新能力；广州、佛山等城市则在2021年后效率从低于0.6跃升至1.0以上，说明它们的追赶效应较明显；而河源、清远、中山等城市的效率起伏较大，缺乏长期的稳定性。

（二）基于地域板块的广东省绿色低碳创新效率分析

按照广东省“十四五”规划的区域划分标准，本文将21个地级市划分为珠三角核心区、粤东地区、粤西地区和粤北山区四大板块，计算各区域研究期内的平均绿色低碳创新效率，结果如表5所示。

表5：广东省各区域绿色低碳创新效率		
区域	包含城市	平均效率
珠三角	广州、深圳、珠海、佛山、江门、肇庆、惠州、东莞、中山	0.838
粤东	汕头、汕尾、潮州、揭阳	0.504
粤北	韶关、河源、梅州、清远、云浮	0.457
粤西	湛江、茂名、阳江	0.221

由上表5可知，2014–2024年珠三角核心区的绿色低碳创新效率均值远高于其他板块，且效率呈现由珠三角向粤东、粤北、粤西依次递减的态势。结合上述城市排序可知，效率较高的城市，如深圳、东莞、珠海、广州等全部位于珠三角核心区，而排名靠后的城市，如茂名、云浮、阳江等主要位于粤西和粤北地区。这说明了珠三角地区凭借雄厚的经济实力以及密集的创新资源，在推动绿色低碳创新方面取得了显著成效；而粤西、粤北地区受到经济发展水平较低、创新人才短缺、产业转型缓慢等因素的制约，使得绿色创新效率长期处于低位。因此，未来应加大对粤东西北地区的绿色技术转移与政策扶持，以此促进区域绿色创新的协调发展。

（三）广东省绿色低碳创新效率的动态演变——基于 ML 指数的测度与分解

为进一步揭示广东省绿色低碳创新效率的动态演变趋势，本节运用 Malmquist–Luenberger(ML) 指数模型（2）对2020–2024年四个时期的变化情况进行测度，结果如表6所示。

表6：广东省各城市绿色低碳创新效率 ML 指数及因素分解

城市	2020–2021			2021–2022			2022–2023			2023–2024		
	ML	EC	TC	ML	EC	TC	ML	EC	TC	ML	EC	TC
广州市	1.523	0.974	1.563	0.986	1.000	0.986	1.065	1.000	1.065	1.055	1.000	1.055
深圳市	1.024	1.000	1.024	0.989	1.000	0.989	1.020	1.000	1.020	1.096	1.000	1.096
珠海市	1.292	1.000	1.292	1.014	1.000	1.014	0.903	1.000	0.903	1.250	1.000	1.250
佛山市	1.246	0.878	1.418	0.959	1.000	0.959	1.004	1.000	1.004	1.102	1.000	1.102
江门市	1.064	0.849	1.254	1.128	0.868	1.298	1.036	0.986	1.051	0.968	1.039	0.932
肇庆市	1.095	0.777	1.409	0.930	1.017	0.915	0.853	1.242	0.687	1.112	0.921	1.207
惠州市	0.923	1.000	0.923	0.965	1.000	0.965	1.000	1.004	0.997	0.988	1.024	0.965
东莞市	0.938	1.000	0.938	0.979	1.000	0.979	1.031	1.000	1.031	1.084	1.000	1.084
中山市	1.133	0.789	1.437	0.881	1.075	0.819	1.027	1.002	1.026	1.182	0.893	1.325
珠三角	1.138	0.919	1.251	0.981	0.996	0.992	0.993	1.026	0.976	1.093	0.986	1.113
汕头市	1.024	0.902	1.135	0.921	1.033	0.892	0.981	1.065	0.921	1.323	0.804	1.646
汕尾市	0.722	1.000	0.722	0.733	1.000	0.733	1.306	1.000	1.306	1.137	1.000	1.137
潮州市	1.059	0.861	1.229	0.955	0.985	0.970	1.050	1.042	1.008	1.324	0.842	1.572
揭阳市	1.362	0.726	1.875	0.868	1.107	0.784	0.970	1.072	0.905	1.463	0.772	1.894
粤东	1.042	0.872	1.240	0.869	1.031	0.845	1.077	1.045	1.035	1.312	0.855	1.562
韶关市	1.186	0.894	1.327	1.547	0.630	2.454	0.819	1.052	0.779	1.107	1.012	1.094
河源市	0.734	1.041	0.705	0.998	0.930	1.074	1.159	0.994	1.166	0.808	1.309	0.618
梅州市	0.657	1.351	0.486	1.106	0.902	1.227	0.953	0.956	0.996	1.039	1.089	0.954
清远市	1.034	0.989	1.045	0.936	1.000	0.936	0.615	1.309	0.469	1.584	0.764	2.074
云浮市	0.646	1.347	0.480	0.939	1.018	0.923	1.033	1.042	0.991	1.235	0.842	1.468
粤北	0.851	1.124	0.809	1.105	0.896	1.323	0.916	1.071	0.880	1.155	1.003	1.242

城市	2020-2021			2021-2022			2022-2023			2023-2024		
	ML	EC	TC	ML	EC	TC	ML	EC	TC	ML	EC	TC
湛江市	1.245	0.902	1.380	1.029	0.912	1.129	0.851	0.902	0.943	1.307	1.035	1.263
茂名市	0.993	1.014	0.979	1.037	0.956	1.085	1.017	0.981	1.036	0.932	1.082	0.861
阳江市	0.870	0.960	0.907	1.046	0.876	1.194	1.103	1.011	1.092	1.304	0.824	1.583
粤西	1.036	0.959	1.089	1.037	0.915	1.136	0.990	0.965	1.024	1.181	0.980	1.236
平均值	1.037	0.964	1.120	0.997	0.967	1.063	0.990	1.031	0.971	1.162	0.964	1.247

从总体上看，2020-2024年广东省绿色低碳创新效率的 ML 指数平均值为 1.047，说明广东省绿色低碳创新效率总体呈上升趋势，各要素投入得到了较好利用。从各时期变化来看，ML 指数在 2020-2021 年相对较高，随后在 2021-2022 年和 2022-2023 年分别下降至 0.997 和 0.990，这表明中期增长动力有所减弱；而 2023-2024 年又大幅跃升至 1.162，这反映出在“双碳”政策持续推动下，末期的绿色全要素生产率实现了显著的提速。进一步将 ML 指数分解为技术效率变化（EC）和技术进步变化（TC），其均值分别为 0.982 和 1.100；这表明广东省绿色低碳创新效率的提升主要源于技术进步，也就是说近年来广东省在绿色低碳转型、绿色技术研发投入以及政策体系建设方面取得了积极的成效，但是技术效率改善不足的问题还需引起重视。

从区域对比来看，粤东、粤西、珠三角、粤北四大区域的 ML 指数分别为 1.075、1.061、1.051、1.007，呈现粤东最高、粤北最低的格局。其中，粤东地区由于原有经济基础和技术水平相对落后，在区域发展条件改善和政策扶持加大的背景下，绿色创新水平的增长率最高；珠三角地区绿色低碳创新已处于较高水平，增长趋于平缓；粤西地区稳步追赶，增速较为均衡；粤北地区虽然略有增长，但波动较大，总体增速最低，需提升增长的稳定性。

从 ML 指数分解项进一步分析，珠三角、粤东和粤西地区的技术效率变化（EC）均小于 1，这表明这些区域在技术效率方面存在退步；而粤北的 EC 达到了 1.024，是唯一实现技术效率改善的区域，也就显示出一定的追赶效应。与此同时，所有区域的技术进步变化（TC）均大于 1，其中粤东最高。这些情况进一步印证了技术进步是驱动各区域绿色全要素生产率增长的核心力量。

（四）广东省绿色低碳创新效率的收敛性分析

1. 广东省绿色低碳创新效率的 σ 收敛分析

运用前文提出的变异系数公式（3），可计算广东省每年的 σ 收敛变异系数结果如图 1 所示。从中可看出，广东省绿色低碳创新效率的变异系数整体呈波动下降态势，从 2014 年的 0.80 下降至 2024 年的约 0.62，表明各地区效率差异在逐步缩小。尽管这段时期反复出现上升和下降的情况，但变异系数的长期趋势是明显的向下走向，峰值也在逐年降低。这一结果表明，广东省绿色低碳创新效率的区域差距正在稳步收窄；粤东、珠三角、粤西、粤北四大板块间的“分化格局”有所改善， σ 收敛特征较显著。

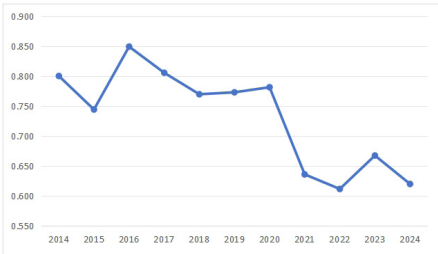


图 1：广东省绿色低碳创新效率的变异系数

2. 广东省绿色低碳创新效率的绝对 β 收敛分析

为考察广东省各地市绿色低碳创新效率是否存在无条件追赶趋势，采用面板回归模型（4）进行绝对 β 收敛检验，结果如表 7 所示。

表 7：广东省绿色低碳创新效率的绝对 β 收敛结果

变量	系数	标准差	t 值	P 值
初期效率对数	-0.737***	0.162	-4.56	0.000
常数项	-0.501**	0.135	-3.71	0.001
R ²	0.289			
收敛速度 λ	1.337			
半生命周期 τ	0.518			

从回归结果来看，初期效率对数的系数为 -0.737，且在 1% 的显著性水平上通过检验。这一结果表明，在 2014-2024 年这个时间段里，初始绿色低碳创新效率较低的城市，它们后期的年均增长率显著高于初始效率较高的城市。换句话说，落后地区会对先进地区表现出一种追赶的态势，绿色创新效率存在显著的绝对 β 收敛。

进一步地，根据收敛速度公式 $\lambda = \frac{\ln(1+\beta)}{T}$ ，可计算出收敛速度 λ 为 1.337，半生命周期 $\tau = \frac{\ln(2)}{\lambda}$ 约为 0.518 年。这意味着，广东省各地级市之间的绿色低碳创新效率差距缩小一半所需时间仅约 0.518 年。这一较快的收敛速度反映出，在样本期内，落后城市的绿色创新效率增长势头强劲，能够迅速对先进城市形成有效追赶。

3. 广东绿色低碳创新效率的条件 β 收敛分析

为剔除各地区结构性差异对收敛效应的影响，本文还进一步引入经济发展水平、产业结构、城镇化水平、对外开放程度及市场竞争强度作为控制变量，进行条件收敛 β 检验。采用面板回归模型（5）得到结果如表 8 所示。

表 8：广东省绿色低碳创新效率的条件 β 收敛结果

变量	系数	标准差	t 值	P 值
初期效率对数	-0.769***	0.175	-4.40	0.000
经济发展水平（lnpgdp）	0.525	0.529	0.99	0.333
产业结构（is）	0.002	0.015	0.15	0.881

变量	系数	标准差	t 值	P 值
城镇化水平 (urban)	-0.005	0.022	-0.24	0.809
对外开放程度 (open)	0.004	0.003	1.35	0.193
市场竞争强度 (lnimc)	-0.100	0.238	-0.42	0.679
常数项	-5.509	3.858	-1.43	0.169
R ²	0.306			
收敛速度 λ	1.464			
半生命周期 τ	0.474			

从核心解释变量来看，初期效率对数的系数为-0.769，且在1%的水平上显著。这一结果与绝对 β 收敛检验中的结论一致，表明即使在控制了各地区的结构性特征之后，初期效率较低的城市仍然表现出更高的后续增长率，故绿色低碳创新效率的条件 β 收敛成立。与绝对 β 收敛相比，条件 β 收敛的系数绝对值从0.737略微增大至0.769，收敛速度 λ 从1.337提升至1.464，半生命周期 τ 从0.518年缩短至0.474年。这一变化说明，在剔除结构性差异的干扰后，收敛效应反而有所增强，即如果进一步缩小产业结构、城镇化水平等方面的差距，落后地区的绿色创新效率可以有更快的提升。

从各控制变量来看，经济发展水平、产业结构、城镇化水平、对外开放程度和市场竞争强度的系数均未通过10%水平的显著性检验，这可能是因为样本量比较有限、变量间存在多重共线性或者各城市的发展阶段差异比较大所导致。尽管如此，各系数的方向基本上符合理论预期，经济发展水平和对外开放程度系数为正，这表明较高的经济基础与开放水平有助于促进绿色创新效率的增长；产业结构系数为正，说明第三产业占比的提升有利于绿色转型；城镇化水平和市场竞争强度的系数为负，反映出快速城镇化过程会带有一定的环境压力，再加上过度竞争可能会把创新资源给挤占掉。

综上所述，条件 β 收敛检验显示广东省绿色低碳创新效率存在稳健的追赶机制，即使把各地区的结构性差异考虑进去，落后城市仍表现出显著的增长优势，且收敛速度比绝对 β 收敛还要更快一些。这表明，推动区域绿色创新协调发展的政策具有坚实的基础，未来应加大对粤东、粤西、粤北等落后地区的绿色技术转移与制度支持，同时还需要着力优化各地区的产业结构与开放环境，以便能够更好地进行绿色转型。

（五）广东省绿色低碳创新效率的影响因素分析——基于Tobit 回归模型

为进一步探究影响广东省绿色低碳创新效率的关键因素，本研究采用 Stata 软件面板 Tobit 回归模型（6）对绿色创新效率影响因素实证分析，结果如表9所示。

表9：广东省绿色低碳创新效率 Tobit 回归结果

变量	系数	标准差	Z 值	P 值
经济发展水 (lnpgdp)	0.045	0.080	0.56	0.573
产业结构 (is)	0.009**	0.003	2.71	0.007
城镇化水平 (urban)	0.007**	0.002	2.95	0.003
对外开放程度 (open)	0.008***	0.001	9.10	0.000
市场竞争强度 (lnimc)	-0.063	0.036	-1.74	0.084
政府支持 (gov)	-0.005*	0.002	-2.48	0.014
常数项	-0.414	0.715	-0.58	0.563
方差参数	0.074***	0.007	10.75	0.000
Log likelihood	-27.66			

结果表明，经济发展水平的系数为0.045，但标准误较大，并且没有通过显著性检验，这可能是因为经济发展水平对绿色创新效率的影响并不是简单的线性关系，而且广东省内各城市经济发展阶段的差异较大，部分城市仍然处于工业化中期，人均 GDP 的提高还没有能够有效转化为绿色创新动力，因此导致整体回归不显著。

产业结构的系数为0.009，在5%水平上显著。这表明，服务业的发展，能够为制造业绿色转型提供技术支撑和资金保障，同时第三产业本身具有较低的能耗和污染排放强度，有助于整体绿色创新绩效的改善。

城镇化水平的系数为0.007，在5%水平上显著。城镇化进程通过人口集聚、产业集聚和基础设施共享，降低了绿色技术研发与推广的平均成本，此外，高城镇化地区通常拥有更完善的教育、医疗等公共服务，有利于吸引高素质人才，为绿色创新提供人力资本支持。因此，城镇化水平的提升对绿色创新效率产生了显著的正向拉动。

对外开放程度的系数为0.008，在1%水平上显著。这一结果表明，货物进出口总额占 GDP 比重越高，地区融入全球市场的程度越深，越有利于通过技术引进、知识溢出和竞争效应促进绿色创新。

市场竞争强度的系数为-0.063，且结果不显著。从理论上来看，一是市场竞争有可能通过“倒逼效应”促进绿色创新，二是它可能因为利润微薄而抑制长期研发投入。在广东省各地级市中，深圳、东莞等城市的企业数量多而且竞争激烈，但龙头企业的绿色创新能力会比较强；而像粤北山区的企业数量少、市场发育也不足，竞争强度低的同时创新动力也会比较弱。这种非线性或异质性的影响，就有可能导致市场竞争强度在整个样本回归中没能呈现出显著的统计关系。

政府支持的系数为-0.005，并且在10%的水平上显著。其系数为负的可能原因，一是财政支出结构中可能存在一种“重基建、轻绿色技术”的倾向，这会导致绿色创新领域的实际支持变得不足；二是政府补贴若缺乏精准性和绩效评估，就可能引发企业的补贴依赖，反而抑制自主创新的动力；三是财政收入占比过高可能意味着企业税负较重，挤占了企业本可用于绿色研发的那部分资金。

四、结束语

本文通过运用超效率 SBM 数据包络分析模型发现，广东省效率水平整体呈上升趋势，但区域之间的发展极不平衡，珠三角的效率均值达到了0.838，而粤西地区却只有0.221，形成了典型的“核心-边缘”格局。从城市层面看，深圳、东莞、汕尾、珠海属于高效领先城市，而茂名、云浮、韶关则长期处在低效率的状态。通过动态曼奎斯特指数分析发现，技术进步是推动广东绿色创新效率提升的主导力量，技术效率的改善就相对滞后一些。另外，广东绿色低碳创新效率收敛的趋势比较明显，区域差距有望持续地缩小， σ 收敛变异系数从2014年的0.80下降到了2024年

的0.62，绝对 β 收敛系数为 -0.737 且在1% 的水平上显著，条件 β 收敛系数为 -0.769 同样也在1% 的水平上显著；这结果表明落后地区存在显著的追赶效应，区域绿色创新效率的不平衡问题正逐步得到缓解。最后通过 Tobit 回归还发现，政府对外开放程度的扩大、产业结构的升级、城镇化水平的提升都会对广东绿色创

新效率产生显著的正向影响，而政府支持表现出负向的影响，这可能与财政支出结构不合理或补贴效率低下有关。广东省政府在推动绿色低碳转型的过程中，应该优化财政支出结构并提高资金使用效率。

参考文献

[1] 郭树华, 方鹏飞, 毕福芳. 基于 SFA 方法的企业绿色创新效率分析 [J]. 工业技术经济, 2024, 43(07):112-119.

[2]Tone K. A slacks-based measure of super-efficiency in data envelopment analysis[J]. European Journal of Operational Research, 2002, 143(1):32-41.

[3]Nannan Wang, Ji Chen, Shengnan Yao, Yen-Chiang Chang.A meta-frontier DEA approach to efficiency comparison of carbon reduction technologies on project level[J]. Renewable and Sustainable Energy Reviews.2018, 82(03):2606-2612.

[4] 钱丽, 王文平, 肖仁桥. 高质量发展视域下中国企业绿色创新效率及其技术差距 [J]. 管理工程学报, 2021, 35(06):97-114.

[5]Joshua Ignatius, M. R. Ghasemi, Feng Zhang, Ali Emrouznejad, Adel Hatami-Marbini.Carbon efficiency evaluation: An analytical framework using fuzzy DEA[J]. European Journal of Operational Research.2016, 253(02):428-440.

[6] Yong Zha, Linlin Zhao, Yiwen Bian. Measuring regional efficiency of energy and carbon dioxide emissions in China: A chance constrained DEA approach[J].Computers & Operations Research.2016, 66:351-361.

[7] 张辽, 黄蕾琼. 中国工业企业绿色技术创新效率的测度及其时空分异特征——基于改进的三阶段 SBM-DEA 模型分析 [J]. 统计与信息论坛, 2020, 35(12):50-61.

[8] 张盼. 中国省域绿色创新效率测度及聚类分析 [D]. 东北林业大学, 2023.

[9] 赵雅婷. 中国绿色技术创新效率测度、时空分异及影响因素研究 [J]. 现代工业经济和信息化, 2023, 13(06):184-188+191.

[10]Weipan Zhang, Xianhua Wu, Jihong Chen. ow-carbon efficiency analysis of rail-water multimodal transport based on cross efficiency network DEA approach[J]. Energy.2024, 305:132348.

[11] 赵志芳. 中国区域绿色技术创新效率的测度及驱动因素研究——基于超效率 SBM 和 Tobit 模型 [J]. 特区经济, 2022, (08):54-57.

[12] 滕琳娜. 中国省际工业绿色创新效率测度及影响因素研究 [D]. 安徽理工大学, 2022.

[13] 刘子豪, 王子涵, 师路路. 中国省域绿色创新效率测度、差异分析及影响因素 [J]. 科技和产业, 2024, 24(17):63-68.

[14] 张平凤. 中国绿色创新效率的区域差异及收敛性研究 [D]. 中南财经政法大学, 2022.

[15] 王星河, 张玲梅. 中国绿色创新效率的地区差异及空间收敛性研究 [J]. 生态经济, 2023, 39(01):81-86+100.

公司创业投资对创业企业 ESG 表现的影响

——基于深圳创业板的实证分析

乔明哲*, 余俊, 肖文慧

上海对外经贸大学 金融管理学院, 上海 201620

DOI:10.61369/ASDS.2026060013

摘 要 : 随着公众对可持续发展和企业社会责任的日益关注, ESG 表现已成为评价企业的重要指标。基于深圳创业板上市公司 2009–2023 年数据的实证分析显示: (1) 公司创业投资 (CVC) 支持的创业企业 ESG 表现显著高于没有 CVC 支持的创业企业; (2) CVC 支持的创业企业 ESG 表现高于独立创业投资 (IVC) 支持的创业企业; (3) CVC 双方业务垂直相关性对创业企业 ESG 表现存在正向影响; (4) CVC 持股比例对业务垂直相关性和创业企业 ESG 表现之间的关系存在负向调节作用。异质性分析显示, 国有背景、制造业和高 ESG 关注度年份的样本中, CVC 对创业企业 ESG 表现的正向影响更强。

关 键 词 : 公司创业投资; 创业企业; ESG 表现; 风险投资

The Impact of Corporate Venture Capital on the ESG Performance of Startups — An Empirical Analysis Based on Shenzhen Growth Enterprise Market

Qiao Mingzhe*, Yu Jun, Xiao Wenhui

School of Finance and Management, Shanghai University of International Business and Economics, Shanghai 201620

Abstract : As public pay more and more attention to sustainable development and corporate social responsibility, ESG performance has become an important indicator for evaluating enterprises. Based on an empirical analysis of data from listed companies of A-share Growth Enterprise from year 2009 to 2023, this study reveals the following findings: (1) Startups supported by Corporate Venture Capital (CVC) exhibit significantly higher ESG performance than those without CVC support; (2) Startups with CVC support enjoy higher ESG performance than their counterparts supported by Independent Venture Capital (IVC); (3) The business vertical relationship between CVC parties has a positive impact on startups ESG performance; (4) CVC shareholding ratio exerts a negative moderating effect on the relationship between business vertical relationship and startups ESG performance. Heterogeneity analysis indicates that the positive effects of CVC on startups ESG performance are stronger in samples with state-owned backgrounds, manufacturing industries, and years with high ESG attention.

Keywords : corporate venture capital; start-up enterprises; ESG performance; venture capital

引言

在党的十八大以后, 生态文明建设被纳入中国特色社会主义事业“五位一体”战略布局, 建设“美丽中国”和绿色、可持续发展经济成为中国社会的远景目标。然而, 一些企业在追求利益最大化时忽视了环境责任、社会责任, 形成了对社会可持续发展能力的挑战。因此, 企业的 ESG 表现日益成为中国实践界和理论研究人员关注的重要问题。

研究显示, 外部宏观因素和企业内部因素都可能影响企业的 ESG 表现^[1, 2]。风险投资 (VC) 主要投资于科创企业, 是支撑创新发展的重要金融主体。虽然已经有研究开始关注 VC 对创业企业 ESG 的影响^[3], 但是仍然不足。和独立风险投资 (IVC) 相比, 公司创业投资 (CVC) 的投资主体主要是拥有产业资源的大型企业, 它们通常把 CVC 作为战略手段。然而, CVC 究竟如何影响创业企业的 ESG 表现? 迄今, 仍然缺乏对该问题的深入探索。

基于此, 本文运用深圳创业板上市公司的数据进行分析, 关注 CVC 投资对创业企业 ESG 表现的影响机制, 以期进一步回答上述问题。

基金项目: 2024 年度国家社会科学基金一般项目“公司创业投资推动企业创新生态系统演化研究”(24BJY140)。

作者简介:

余俊, 上海对外经贸大学金融管理学院, 硕士研究生, 研究方向为创业投资、企业社会责任;

肖文慧, 上海对外经贸大学金融管理学院, 硕士研究生, 研究方向为创业投资。

通讯作者: 乔明哲, 上海对外经贸大学金融管理学院, 副教授, 硕士生导师, 研究方向为创业投资、企业社会责任。

一、文献综述与研究假设

（一）CVC 与创业企业 ESG 表现

风险投资家凭借专业能力筛选出优质的创业企业，并积极参与创业企业的管理^[4]，对创业企业进行监督^[5]。VC 可以改善其董事会治理^[6]以及缓解管理层代理问题^[7]。可见，VC 的投后管理和增值服务有助于改善创业企业的治理水平。直接证据也显示，VC 可以提升被投资企业的 ESG 表现^[8]。

不过也有证据揭示，VC 的介入可能不利于创业企业的企业社会责任（CSR）表现。比如，“印刻”效应会导致 VC 支持的创业企业 CSR 表现较差^[8]。来自国内的证据显示，上市公司的 CSR 评价在 VC 股东退出后显著提升，但主要是缺乏经验、低声誉的 VC 促使企业把资源集中在股东利益最大化上所致^[9]。

虽然有研究显示，开展 CVC 项目可以促进投资企业的 CSR 表现^[10]，但是关于 CVC 对创业企业 ESG 影响的研究非常少。从相关研究来看：一方面，许多证据表明 CVC 有助于创业企业的发展，比如提供互补性资源^[11]，助力其提升创新绩效^[12]。这得益于 CVC 投资者丰富的经营资源和参与治理。另一方面，CVC 投资可以促进双方组织间学习，创业企业得以从投资企业的知识流获益^[13]。投资企业通常具备更加成熟的 ESG 实践，可以在和创业企业的互动中传递 ESG 理念和知识，进而推动后者提高 ESG 表现。因此，提出假设：

H1：CVC 支持的创业企业 ESG 表现好于没有 CVC 支持的创业企业。

CVC 通常注重战略回报，而不是像 IVC 那样纯粹追求财务收益^[14、15]，这使得 CVC 通常比 IVC 投资周期更长、金额更大^[16]。相对于 IVC 而言，CVC 具有丰富的行业知识，更能够包容创业企业的失败，其支持的创业企业更具有创新性^[17、18]。IVC 主要由独立投资人或投资机构管理，更加关注投资的财务回报最大化，这使得它们在投资决策中难以做到像 CVC 那样长期导向，对创业企业的 ESG 表现和长期可持续发展能力关注不足。

当然，也有研究认为 CVC 存在负面影响。比如，基于美国数据的研究显示，IVC 支持的创业企业 ESG 表现好于 CVC 支持的同类^[19]。不过国内证据显示，VC 对创业企业 ESG 表现的负面效应主要由于缺乏经验和低声誉的投资者所致^[9]。CVC 通常来自具有市场声誉的大企业，产业经验丰富，且注重长远的战略利益。中国的 CVC 主要基于市场扩张逻辑开展投资^[20]，关注在产业链上下游布局创业企业。因此，国内的 CVC 投资者有理由敦促投资企业提升其 ESG 表现，为自身营造“负责任”的良好形象。基于此，提出假设：

H2：CVC 支持的创业企业 ESG 表现好于 IVC 支持的创业企业。

（二）业务垂直相关性与创业企业 ESG 表现

当 CVC 投资双方处于相关行业时，知识储备重合度高，有助于组织间学习^[21]，更容易实现协同效应^[22]。此时，投资企业与创业企业的组织间学习会随之增强。研究表明，CVC 双方在战略技术重叠度较高情况下收益更大^[23]，创业企业能够从投资企业获取

更多收益^[24]。

然而，创业企业也有可能陷入“与鲨共泳”的困境^[25]，证据表明，与投资企业战略重合度很高时，创业企业接受 CVC 的可能性降低^[26]。这意味着在双方业务相关时，创业企业既可能从投资企业获取提升 ESG 表现的知识，也面临利益受到侵害导致 ESG 表现下降的风险。当然，业务水平相关和垂直相关受到负面影响的程度可能存在差异。业务水平相关时，双方处于竞争关系，投资企业通过抑制创业企业减弱竞争的动机增强。双方业务垂直相关时，双方更可能通过互补性资源实现互惠，有利于创业企业的发展^[11]。此外，敦促供应链上下游的创业企业提升 ESG 水平也有助于维护投资企业的市场形象。综上，提出以下假设：

H3：双方业务垂直相关时，CVC 支持的创业企业 ESG 表现更好。

（三）CVC 持股比例、业务垂直相关性与创业企业 ESG 表现

CVC 持股比例既决定了投资企业与创业企业之间的利益关联程度，也反映了前者对后者的支配能力。研究显示，CVC 投资者与创业企业的互动可以促进后者的创新活动^[18]。虽然缺少 CVC 持股比例与创业企业 ESG 表现之间关系的证据，但已有研究显示，CVC 投资强度与创业企业的创新绩效正相关^[27]。当 CVC 持股比例较高时，投资企业能够从创业企业发展中获取更大的收益，会加强与创业企业的互动，进而增强业务垂直相关双方的学习效果。此外，较高股权比例也会使得投资企业出于自身利益督促创业企业更好地发展。

然而，相关研究显示，CVC 持股比例与创业企业创新绩效之间的关系具有复杂性^[28]。CVC 持股比例高意味着投资企业对创业企业的控制能力强，增加了其侵占后者利益的风险^[25、29]。此时，投资企业支配业务垂直相关的创业企业将会攫取到更多经济利益，从而可能去侵占创业企业的利益，妨碍其发展。鉴于上述两种可能情况，提出竞争性假设：

H4a：CVC 持股比例正向影响业务垂直性与创业企业 ESG 表现之间的关系。

H4b：CVC 持股比例负向影响业务垂直性与创业企业 ESG 表现之间的关系。

二、研究设计和数据来源

（一）样本选择和数据来源

本文选取 2009–2021 年完成 IPO 的深圳创业板上市公司作为样本，运用 2012–2023 年 ESG 数据进行分析。对样本进行如下处理：剔除金融行业与 ST 企业；剔除数据缺失的观测值；对连续变量进行了 1% 缩尾处理。最终得到 858 家样本企业，ESG 数据来自 WIND 数据库的华证 ESG 评级数据，其他数据取自 CSMAR 和 iFIND 数据库。

（二）变量定义

1. 因变量

创业企业 ESG 表现（ESG）。华证 ESG 评价体系参照了国际主流的评价方法，本文以样本企业 IPO 当年及以后两年（ $t-t+2$ ）

华证 ESG 总得分的均值衡量创业企业 ESG 表现。

2. 解释变量

CVC 投资（CVC）。参照相关研究^[23, 30]，以非金融类企业直接或间接向创业企业进行的少数股权投资来界定 CVC。通过对创业板公司 IPO 时的前十大股东数据根据定义进行人工筛选取得。

VC 类型（VCtype）。本文定义了 VCtype 分类变量，用于比较 CVC 与 IVC 对创业企业 ESG 表现的影响，其中包括 3 个类型变量：PureIVC、CVCIVC 和 nonVC，分别表示仅有 IVC 持股、兼有 CVC 和 IVC 持股和没有 VC 持股。

业务垂直相关性（Vrelated）。参考相关研究^[31]，本文将业务领域处于上下游行业或存在供应商及顾客关系的 CVC 投资企业和创业企业定义为业务垂直相关，虚拟变量 Vrelated 记为 1，否则记为 0。

3. 调节变量

CVC 持股比例（CVCratio）。以 CVC 投资企业或者 CVC 基金持有的创业企业股权比例衡量，数据通过人工查询样本企业上市时前十大股东中 CVC 投资方的持股比例获取。

4. 控制变量

参考相关 ESG 和 VC 领域的研究^[9, 10, 20, 32, 33]，引入企业年龄（Age）、企业规模（Size）、资产负债率（Lev）、总资产报酬率（ROA）、营业收入增长率（Growth）、第一大股东持股比例（Top1share）、董事会独立性（Indepe）、研发投入（RD）、国有背景（State）、制造业（Manuf）、东部区域（East）、上市年份（IPOyear）作为控制变量。

变量定义详见表 1。

表 1：变量定义

变量类型	名称	符号	变量计量
因变量	创业企业 ESG 水平	<i>HZesg</i>	企业 IPO 的 <i>t-t+2</i> 年华证 ESG 评分均值
解释变量	有无 CVC 持股	<i>CVC</i>	无 CVC：0；有 CVC：1
	VC 类型	<i>VCtype</i>	PureIVC、CVCIVC、NonVC
	业务垂直相关性	<i>Vrelated</i>	垂直相关：1；其他：0
调节变量	CVC 持股比例	<i>CVCratio</i>	CVC 持有被投资企业股权的比例
控制变量	企业年龄	<i>Age</i>	观测年份减去企业成立年份
	企业规模	<i>SIZE</i>	企业当年总资产取对数
	资产负债率	<i>LEV</i>	企业当年年末负债 / 年末总资产
	总资产报酬率	<i>ROA</i>	企业当年净利润 / 当年平均资产总额
	营业收入增长率	<i>Growth</i>	当年营业收入 / 上年营业收入 - 1
	第 1 大股东持股比例	<i>Top1share</i>	第一大股东持股数 / 总股数
	董事会独立性	<i>Indepe</i>	独立董事人数 / 董事会总人数
	研发投入	<i>RD</i>	研发投入加 1 取对数
	国有背景	<i>State</i>	是否为国有背景
	制造业	<i>Manuf</i>	公司所属行业为制造业
	东部区域	<i>East</i>	是否位于中国东部区域
	上市年份	<i>IPOyear</i>	企业上市年份

（三）模型设定

式（1）用于检验 H1：

$$HZesg = \alpha_0 + \alpha_1 CVC + \alpha_i \sum Control_i + \varepsilon \quad (1)$$

其中 HZesg 为因变量，CVC 为解释变量，Control_i 为控制变量。

式（2）用于检验 H2：

$$HZesg = \alpha_0 + \alpha_1 PureCVC + \alpha_2 CVCIVC + \alpha_3 NonVC + \alpha_i \sum Control_i + \varepsilon \quad (2)$$

其中 PureCVC、CVCIVC、NonVC 为分类变量 VCtype 转换的虚拟变量。

式（3）用于检验 H3：

$$HZesg = \alpha_0 + \alpha_1 Vrelated + \alpha_i \sum Control_i + \varepsilon \quad (3)$$

式（4）用于检验 H4，其中 VRelated×CVCratio 为 Vrelated 和调节变量 CVCratio 的交乘项。

$$HZesg = \alpha_0 + \alpha_1 Vrelated + \alpha_2 CVCratio + \alpha_3 Vrelated \times CVCratio + \alpha_i \sum Control_i + \varepsilon \quad (4)$$

式（1）和（2）使用全样本数据进行回归，式（3）和（4）使用股东中存在 CVC 的样本进行回归。

三、实证分析结果

（一）描述性统计

描述性统计分析见 2。44% 的创业企业有 CVC 介入，表明其已经成为重要的投资者。Top1share 均值为 34.31%，表明样本企业的股权集中度较高。RD 均值为 0.167，说明样本企业研发投入较高。State 均值为 0.12，表明国有背景企业占比不高。表格下方是 CVC 样本主要研究变量的描述性统计。可以看出：有 CVC 支持的创业企业 ESG 得分均值高于总体样本；CVC 的持股比例最小值、最大值分别为 0.06% 和 24.03%，差距较大；40.9% 的投资企业与被投资企业的业务垂直相关。

表2：描述性统计

样本	变量	观测值	均值	标准差	中值	最小值	最大值
全样本	<i>HZesg</i>	858	74.547	3.116	74.715	59.73	84.25
	<i>CVC</i>	858	0.444	0.497	0	0	1
	<i>PureCVC</i>	858	0.05	0.218	0	0	1
	<i>PureIVC</i>	858	0.395	0.489	0	0	1
	<i>CVCIVC</i>	858	0.394	0.489	0	0	1
	<i>nonVC</i>	858	0.161	0.368	0	0	1
	<i>Age</i>	858	2.531	0.393	2.565	1.386	3.466
	<i>Size</i>	858	20.846	0.677	20.77	19.49	25.91
	<i>Lev</i>	858	0.22	0.145	0.189	0.011	0.868
	<i>ROA</i>	858	0.117	0.058	0.108	0.001	0.511
	<i>Growth</i>	858	0.211	0.303	0.174	-0.702	2.618
	<i>Top1share</i>	858	0.344	0.132	0.332	0.067	0.9
	<i>Indepe</i>	858	0.375	0.050	0.333	0.25	0.6
	<i>RD</i>	858	0.167	0.025	0.169	0	0.214
	<i>State</i>	858	0.121	0.327	0	0	1
	<i>Manuf</i>	858	0.707	0.455	1	0	1
	<i>East</i>	858	0.544	0.498	1	0	1
CVC 样本	<i>HZesg</i>	381	75.007	3.096	75.1	65.14	84.25
	<i>Vrelated</i>	381	0.409	0.492	0	0	1
	<i>CVCratio</i>	381	0.059	0.062	0.035	0.001	0.311

（二）回归分析结果

表3是多元回归结果，其中模型（1）仅包含控制变量。模型（2）引入解释变量 CVC，回归系数显著为正，H1 得到支持。模型（3）中虚拟变量 PureIVC、CVCIVC、NonVC 的回归系数均显著为负，表明 IVC 支持的企业 ESG 表现显著低于 CVC 支持的企业，H2 得到支持。模型（4）到（6）是 CVC 持股样本的回

归结果，其中模型（4）仅包含控制变量。模型（5）中解释变量 Vrelated 的系数显著为正，H3 得到支持。模型（7）引入调节变量 CVCratio 和交互项 Vrelated×CVCratio，交互项系数为负值且显著，表明 CVCratio 对 Vrelated 和 HZesg 之间的关系存在负向调节作用，因此拒绝 H4a 而接受 H4b。

表3：多元回归结果

变量	HZesg					
	(1)	(2)	(3)	(4)	(5)	(6)
CVC		1.090*** (0.208)				
PureIVC			-1.767*** (0.513)			
CVCIVC			-1.013** (0.509)			
nonVC			-2.509*** (0.533)			
Vrelated					0.720** (0.316)	0.731** (0.321)
CVCratio						4.841 (3.118)
Vrelated×CVCratio						-9.044** (4.583)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
_cons	62.814*** (4.473)	61.940*** (4.427)	63.859*** (4.386)	60.757*** (7.447)	60.050*** (7.435)	59.029*** (7.469)
<i>N</i>	858	858	858	381	381	381
<i>R</i> ²	0.099	0.127	0.138	0.107	0.119	0.126
<i>F</i>	4.309***	5.384***	5.436***	2.135***	2.461***	2.329***

注：***、**、* 分别表示在 1%、5% 和 10% 的水平上显著，括号内为稳健标准误。

（三）稳健性检验

1. Heckman 二阶段模型

为应对样本选择性偏误问题，进行了 Heckman 二阶段分析。第一阶段，以 CVC 作为因变量，参照相关研究^[31]，选用企业年龄（Age）、总资产报酬率（ROA）、营业收入增长率（Growth）、高科技行业（虚拟变量 Hightech，生物医药行业和 IT 行业的企业赋值 1）、企业位于 4 大 VC 聚集城市（虚拟变量 Top4city，企业注册地址在北京、上海、广州、深圳等城市则赋值 1）、独立创业投资持股（IVC）等变量，运用 Probit 模型进行回归，结果如下：Hightech，0.048；Top4city，0.001；IVC，0.731***；Age，-0.209*；ROA，-2.022*；Growth，0.316*（*代表显著度）。第

二阶段，生成逆米尔斯比率（IMR）并将其引入回归方程，结果见表 4 中模型（1），IMR 系数不显著而 CVC 的系数仍然显著为正，表明样本不存在明显的选择性偏误。

2. 替换因变量

构建虚拟变量 HZESG，当 HZesg 大于等于样本的中位数时取 1，否则取 0。Logit 回归结果见表 4 中模型（2）-（5）。其中：CVC 的回归系数显著为正；PureIVC、CVCIVC、NonVC 回归系数显著为负，Vrelated 的回归系数显著为正；交互项 Vrelated×CVCratio 的回归系数虽未达到 10% 显著度但方向仍为负。总体而言，回归结果相对稳健。

表 4：稳健性检验

变量	HZesg	HZESG	HZESG	HZESG	HZESG
	(1)	(2)	(3)	(4)	(5)
CVC	1.050*** (0.217)	0.746*** (0.172)			
IMR	-0.427 (0.525)				
PureIVC			-1.502*** (0.351)		
CVCIVC			-1.004*** (0.343)		
nonVC			-2.036*** (0.415)		
Vrelated				0.452* (0.250)	0.480* (0.251)
CVCratio					2.454 (2.842)
Vrelated×CVCratio					-6.405 (4.082)
	(0.456)	(0.348)	(0.349)	(0.520)	(0.524)
Controls	Yes	Yes	Yes	Yes	Yes
_cons	62.209*** (4.454)	-11.417*** (3.138)	-10.249*** (3.096)	-10.492** (4.464)	-11.301** (4.483)
N	858	858	858	381	381
Log Likelihood/R ²	0.128	-446.590	-441.051	-212.441	-211.057
χ ² /F	5.192***	59.947***	65.479***	38.092**	40.020**

注：***、**、* 分别表示在 1%、5% 和 10% 的水平上显著，括号内为稳健标准误。

（四）异质性分析

1. 产权性质

表 5 中模型（1）（2）分别给出了国有背景和非国有背景样本下的回归结果。在国有背景样本下，CVC 持股对创业企业 ESG 表现的影响更大（回归系数 1.692）。

2. 制造业

模型（3）（4）分别给出了制造业和非制造业样本下的回归结果。制造业样本中 CVC 的回归系数显著为正，而非制造业样本中

不显著，表明 CVC 对创业企业 ESG 表现的促进作用主要体现在制造业企业中。

3. ESG 高关注年份

鉴于国内对于 ESG 的关注开始于 2015 年左右，按 IPO 年份把数据划分为 2015 年前、后两个样本。两个样本中 CVC 的回归系数均显著为正，但 2015 年之后样本中系数更大（1.274），表明社会普遍关注 ESG 后，CVC 对创业企业 ESG 表现的促进作用有所增强。

表5：异质性分析

变量	HZesg					
	(1)	(2)	(4)	(3)	(6)	(5)
	国有	非国有	制造业	非制造业	15年之后	15年之前
CVC	1.692** (0.757)	0.962*** (0.225)	1.381*** (0.238)	0.518 (0.424)	1.274*** (0.275)	0.780** (0.322)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
_cons	51.456*** (11.448)	62.858*** (4.761)	56.696*** (5.062)	69.335*** (8.616)	53.253*** (5.543)	77.723*** (7.083)
N	104	754	607	251	542	316
R ²	0.405	0.108	0.187	0.086	0.140	0.142
F	4.243***	4.299***	6.626***	0.967***	5.283***	2.837***

注：***、**、* 分别表示在 1%、5% 和 10% 的水平上显著，括号内为稳健标准误。

四、结论和政策建议

本文基于深圳创业板市场 2009–2023 年上市公司数据分析了 CVC 对创业企业 ESG 表现的影响，主要结论如下：（1）CVC 支持的创业企业 ESG 表现显著高于没有 CVC 支持的创业企业；（2）CVC 支持的创业企业 ESG 表现显著高于 IVC 支持的创业企业；（3）业务与 CVC 投资企业垂直相关的创业企业 ESG 表现更好；（4）CVC 持股比例对业务垂直相关性和创业企业 ESG 表现之间的关系具有显著调节作用；（5）国有背景、制造业、ESG 高关注年份的样本中，CVC 对创业企业 ESG 表现的促进作用更强。

上述结论也给相关实践带来启示：（1）引入 CVC 投资者尤其是业务垂直相关的投资者，有助于提升创业企业的 ESG 水平；（2）接受 CVC 投资的创业企业应考虑通过谈判将其持股比例限定在合理水平，过高的 CVC 持股比例不利于投资企业治理效应的发

挥，反而由于投资企业控制力的增强而产生侵害创业企业利益的风险。（3）对于 ESG 理念投资者而言，在 IPO 后三年内的创业板上市公司中，有 CVC 投资支持的公司更值得作为投资标的加以关注。

文章揭示了 CVC 支持的创业企业 ESG 表现不仅高于没有 CVC 支持的同类，还显著高于 IVC 支持的同类。这为关于 VC 对创业企业 ESG 或者 CSR 影响的研究结论^[3、8、34]存在较大分歧提供了一种理解途径，即不同 VC 投资者在目标和资源上的差异，可能导致其对创业企业 ESG 绩效的影响不同。此外，本研究还表明，双方业务垂直相关性对接受 CVC 的创业企业的 ESG 绩效存在正向影响，为理解 CVC 投资双方互补性和竞争性的复杂影响提供了证据。当然，本文也存在不足之处：首先，样本数据来自 A 股创业板市场，结论未必适用于其他市场；其次，未能充分关注 CVC 投资者的其他特征对创业企业 ESG 表现的影响。

参考文献

[1]Khaw T Y and Amran A and Teoh A P. Factors influencing ESG performance: A bibliometric analysis, systematic literature review, and future research directions [J]. Journal of Cleaner Production, 2024, 448.

[2]Martiny A, et al. Determinants of environmental social and governance (ESG) performance: A systematic literature review [J]. JOURNAL OF CLEANER PRODUCTION, 2024, 456: 142213.

[3]Drobetz W, et al. Entrepreneurial finance and sustainability: Do institutional investors impact the ESG performance of SMEs? [J]. Journal of Business Venturing Insights, 2024, 22: e00498.

[4]Sahlman W A. The structure and governance of venture capital organizations [J]. Journal of Financial Economics, 1990, 27(2): 473–521.

[5]Lerner J. Venture Capitalists and the Oversight of Private Firms [J]. The Journal of Finance, 1995, 50(1): 301–18.

[6]Fried V H and Bruton G D and Hisrich R D. Strategy and the board of directors in venture capital–backed firms [J]. Journal of Business Venturing, 1998, 13(6): 493–503.

[7]Hochberg Y V. Venture Capital and Corporate Governance in the Newly Public Firm [J]. Review of Finance, 2012, 16(2): 429–80.

[8]Alakent E and Goktan M S and Khoury T A. Is venture capital socially responsible? Exploring the imprinting effect of VC funding on CSR practices [J]. Journal of Business Venturing, 2020, 35(3): N.PAG–N.PAG.

[9]Cheng C, et al. Venture Capital and Corporate Social Responsibility [J]. Journal of Corporate Finance, 2022, 75.

[10]Battisti E, et al. Corporate venture capital and CSR performance: An extended resource based view’s perspective [J]. Journal of Business Research, 2022, 139: 1058–66.

[11]Park H D and Steensma H K. When Does Corporate Venture Capital Add Value For New Ventures? [J]. Strategic Management Journal, 2012, 33(1): 1–22.

[12]Uzuegbunam I and Ofem B and Nambisan S. Do Corporate Investors Affect Entrepreneurs’ IP Portfolio? Entrepreneurial Finance and Intellectual Property in New Firms [J]. Entrepreneurship: Theory & Practice, 2019, 43(4): 673–96.

[13]Yang Y. Bilateral inter–organizational learning in corporate venture capital activity: Governance characteristics, knowledge transfer, and performance [J]. Management Research Review, 2012, 35(5): 352–78.

- [14]Dushnitsky G and Lenox M J. When do firms undertake R&D by investing in new ventures? [J]. Strategic Management Journal, 2005, 26(10): 947–65.
- [15]Wadhwa A and Kotha S. Knowledge creation through external venturing: evidence from the telecommunications equipment manufacturing industry [J]. Academy of Management Journal, 2006, 49(4): 819–35.
- [16]Mazza P and Shuwaikh F. Industry-relatedness, geographic proximity and strategic decisions of corporate and independent venture capital-backed companies [J]. Journal of Small Business Management, 2024, 62(2): 966–1003.
- [17]Chemmanur T J and Loutskina E and Tian X. Corporate Venture Capital, Value Creation, and Innovation [J]. Review of Financial Studies, 2014, 27(8): 2434–73.
- [18]Paik Y and Woo H. The Effects of Corporate Venture Capital, Founder Incumbency, and Their Interaction on Entrepreneurial Firms’ R&D Investment Strategies [J]. Organization Science, 2017, 28(4): 670–89.
- [19]Shuwaikh F, et al. Insights for sustainable business practices: Comparative impact of independent and corporate venture capital funding on financial and environmental performance [J]. Research in International Business and Finance, 2025, 73: 102632.
- [20]Dushnitsky G and Yu L. Why Do Incumbents Fund Startups? A Study of the Antecedents of Corporate Venture Capital in China [J]. Research Policy, 2022, 51(3): 104463.
- [21]Cohen W and Levinthal D. Absorptive capacity: a new perspective on learning and innovation [J]. Administrative Science Quarterly, 1990, 35(1): 128–52.
- [22]Montgomery C A and Wernerfelt B. Diversification, Ricardian Rents, and Tobin’ s q [J]. The RAND Journal of Economics, 1988, 19(4): 623–32.
- [23]Gompers P A and Lerner J. The determinants of corporate venture capital success: organizational structure, incentives, and complementarities [Z]//MORCK R K. Concentrated Corporate Ownership USA; University of Chicago Press. 2000: 17–54
- [24]Balachandran S. The inside track: Entrepreneurs’ corporate experience and startups’ access to incumbent partners’ resources [J]. Strategic Management Journal, 2024, 45(6): 1117–50.
- [25]Katila R and Rosenberger J D and Eisenhardt K M. Swimming with sharks: technology ventures, defense mechanisms and corporate relationships [J]. Administrative Science Quarterly, 2008, 53(2): 295–332.
- [26]Bae J and Lee J M. How Technological Overlap between Spinouts and Parent Firms Affects Corporate Venture Capital Investments in Spinouts: The Role of Competitive Tension [J]. Academy of Management Journal, 2021, 64(2): 643–78.
- [27]Lin J–Y. What affects new venture firm’ s innovation more in corporate venture capital? [J]. European Management Journal, 2020, 38(4): 646–60.
- [28] 乔明哲, et al. 公司创业投资、防护机制对创业企业技术创新绩效的影响研究 [J]. 管理学报, 2024: 1–11.
- [29] 董静 and 孙嘉来. 公司风险投资对创业企业价值的影响——基于投资阶段的悖论视角 [J]. 研究与发展管理, 2025, 37(01): 168–83.
- [30]Allen S A and Hevert K T. Venture capital investing by information technology companies: Did it pay? [J]. Journal of Business Venturing, 2007, 22(2): 262–82.
- [31] 乔明哲, et al. 公司创业投资究竟怎样影响创业企业的 IPO 抑价——来自深圳创业板市场的证据 [J]. Lkgp, 2017, 20(1): 167–80.
- [32]Duque–Grisales E and Aguilera–Caracuel J. Environmental, Social and Governance (ESG) Scores and Financial Performance of Multilatinas: Moderating Effects of Geographic International Diversification and Financial Slack [J]. Journal of Business Ethics, 2021, 168(2): 315–34.
- [33]Wang L and Huang H and An Y. Technological Fit, Control Rights Allocation, and Innovation Performance of Corporate Venture Capital–Backed Enterprises [J]. Venture Capital, 2021, 23(3): 229–55.
- [34]Cheng M and Liu J and Chen Z. The Effect of Venture Capital on Invested Firms’ Social Responsibility: Evidence from Venture Capital’ s Exits [J]. Emerging Markets Finance & Trade, 2023, 59(6): 1668–89.

科技金融助推中国西部地区全要素生产率提升的作用路径及对策研究

石光周¹, 钱成^{2*}

1. 陕西神木农村商业银行股份有限公司, 陕西 神木 719000

2. 中国建设银行股份有限公司温州分行, 浙江 温州 325000

DOI:10.61369/ASDS.2026060014

摘 要 : 在创新驱动发展战略与区域协调发展战略深入推进的背景下, 如何借助科技金融提升中国西部地区全要素生产率已成为推动经济高质量发展的重要议题。本文以科技金融试点政策为准自然实验, 基于2000—2023年中国西部地区地级市面板数据, 构建多时点双重差分模型, 系统考察科技金融对中国西部地区全要素生产率的影响效应、作用机制及异质性影响。研究发现: 科技金融试点显著提升了中国西部地区全要素生产率, 且该结论经过替换被解释变量和更换回归模型等多种稳健性检验后依然成立; 机制分析表明, 科技金融通过提升人力资本水平和推动产业结构高级化促进中国西部地区全要素生产率增长; 异质性分析显示, 该政策效应在中国西北地区以及高创新水平城市中更为显著。基于此, 应完善中国西部地区科技金融支持体系, 夯实创新与人才基础, 并推动产业结构优化升级, 以进一步释放科技金融对区域高质量发展的赋能效应。

关 键 词 : 科技金融; 全要素生产率; 西部地区; 人力资本水平

The Action Path and Countermeasure Research of Sci-Tech Finance Boosting Total Factor Productivity Growth in Western China

Shi Guangzhou¹, Qian Cheng^{2*}

1. Shaanxi Shenmu Rural Commercial Bank Co., Ltd., Shenmu, Shanxi 719000

2. China Construction Bank Corporation Wenzhou Branch, Wenzhou, Zhejiang 325000

Abstract : Against the background of the in-depth implementation of the innovation-driven development strategy and the regional coordinated development strategy, how to improve total factor productivity in western China through sci-tech finance has become an important issue for promoting high-quality development. Taking the sci-tech finance pilot policy as a quasi-natural experiment, this research develops a time-varying DID model using panel data from prefecture-level cities located in Western China during 2000–2023, and systematically investigates the impact, mechanism and heterogeneous effects of sci-tech finance on total factor productivity. According to the empirical findings, the science and technology finance pilot policy has exerted a significantly positive effect on enhancing total factor productivity in western China, and the finding remains robust even when employing different dependent variables and utilizing alternative regression specifications for robustness testing. Mechanism analysis indicates that sci-tech finance promotes total factor productivity growth by improving human capital level and advancing industrial structure upgrading. Heterogeneity analysis reveals that the policy effect is more significant in the northwest region and cities with high innovation capacity. Accordingly, this paper proposes to improve the sci-tech finance support system in western China, consolidate the foundation of innovation and talents, and promote the optimization and upgrading of industrial structure, so as to further release the empowering effect of sci-tech finance on regional high-quality development.

Keywords : sci-tech finance; total factor productivity; western China; human capital level

作者简介: 石光周 (1984.03–), 男, 陕西人, 硕士, 中级经济师, 邮箱: sgz0303@163.com。

通讯作者: 钱成 (2001.09–), 男, 温州人, 硕士, 邮箱: qc15355981393@126.com。

引言

在中国经济由高速增长阶段转向高质量发展阶段的背景下，全要素生产率逐渐成为衡量区域经济发展质量与效率的重要指标。《中共中央关于制定国民经济和社会发展第十五个五年规划的建议》指出，应大力发展科技金融、绿色金融、普惠金融、养老金融、数字金融¹。其中，科技金融作为连接科技创新与金融资源的重要方式，正在成为提升资源配置效率、推动全要素生产率增长的重要抓手。为推动科技金融发展，中国早于2011年启动“促进科技和金融结合试点”，通过政策引导金融资源向科技创新领域集聚，推动科技成果转化与产业升级。这一政策在为科技型企业提供更加多元化融资渠道的同时，也为从政策视角识别科技金融对经济效率影响提供了重要现实基础。从区域发展角度来看，西部地区在中国区域协调发展战略中具有重要地位。国家层面先后提出西部大开发战略、“一带一路”倡议以及区域协调发展战略，旨在缩小东西部差距、提升西部地区发展能力。然而，受制于产业基础相对薄弱、金融资源配置不足以及创新能力有限等因素，中国西部地区在全要素生产率提升方面仍存在明显短板。与此同时，随着新发展格局的构建，西部地区正逐步成为承接产业转移和推动国内大循环的重要区域，对科技金融的依赖程度不断增强。因此，在区域协调发展战略背景下，进一步探讨科技金融如何赋能西部地区全要素生产率提升具有重要的理论和现实意义。

已有部分研究从企业或区域视角出发证实科技金融能够显著提升全要素生产率，但仍存在进一步研究空间：其一，多数文献聚焦中国或中国东中部地区，对西部地区这一发展相对滞后的区域关注不足；其二，现有研究对科技金融影响全要素生产率的作用路径缺乏系统识别，尤其在人力资本积累与产业结构高级化方面的机制分析仍有待深化。基于此，本文以2011年启动的科技金融试点政策为准自然实验，构建多时点双重差分模型，使用2000—2023年中国西部地区地级市面板数据，系统考察科技金融对西部地区全要素生产率的影响效应、作用机制及异质性影响。

一、文献综述

以往文献围绕科技金融和全要素生产率相关内容已展开了广泛探究，本文从科技金融、全要素生产率以及两者关联三方面入手对以往文献进行梳理，进而明晰该领域研究共识与进一步研究空间，明确本文研究切入点。

（一）科技金融相关研究

已有研究普遍认为科技金融本质上是科技资源与金融资源在制度、工具和服务层面的深度融合，其核心功能在于缓解科技创新过程中的融资约束、分散高风险创新活动的不确定性，并通过资本配置引导技术进步和产业演进^[1、2]。在此基础上，部分文献开始从资源配置效率等角度考察中国科技金融发展状况，发现中国科技金融投入规模持续上升，但科技金融资源配置效率整体尚未达到最优状态，区域差异较为明显^[3]。

随着研究不断深入，学者们逐步将关注点转向“科技金融能

否有效促进经济发展”问题上。一类研究发现，科技金融能够在特定条件下显著促进地区创新水平提升，并且这一作用在政府效率较高、创新基础较好的地区表现得更为明显^[4]；另一类研究则进一步指出，科技金融不仅影响创新活动本身，还会通过促进金融发展、优化资源流动、改善产业资本配置等途径推动产业结构升级与经济高质量发展，且这种作用在不同类型城市中具有明显的异质性^[5、6]。总体而言，既有研究已较充分地揭示了科技金融对创新和结构转型的促进作用。

（二）全要素生产率相关研究

全要素生产率长期以来一直是经济学研究中的重要议题。起初，相关文献多围绕其测度方法展开研究：相关研究多使用索洛残差、隐性变量与潜在产出法测算全国全要素生产率，并指出该测算方式受假设约束与效率缺口影响^[7]。在此基础上，部分学者以非农就业密度表征集聚，从空间维度补充生产率差异测度思路，测算地级市层面的全要素生产率^[8]。还有学者从绿色角度引入方向距离函数与ML指数，将能源、污染纳入非期望产出，测度绿色全要素生产率，为资源环境约束下的生产率核算提供规范框架^[9]。

其次，在驱动因素方面，部分研究强调技术创新是推动全要素生产率提升的核心动力。创新活动能够通过提高技术效率、改善生产工艺和增强知识外溢效应来提升企业或地区生产效率，但这一作用往往受到人力资本、制度环境和产业基础的共同制约，因而表现出显著的异质性特征^[10、11]。另一部分研究则从结构视角出发，认为产业结构调整 and 升级并不只是部门比重的简单变化，而是要素由低效率部门向高效率部门重新配置的过程，产业结构合理化有助于释放“结构红利”，进而促进全要素生产率提升^[12、13]。由此可见，全要素生产率提升来源于技术进步、结构优化和资源配置改善的共同作用。

（三）科技金融与全要素生产率关系

相较于科技金融和全要素生产率较为成熟的研究脉络，直接考察二者关系的研究出现较晚。在宏观层面，部分研究发现科技金融发展能够显著提升地区或行业全要素生产率，其作用体现为通过融资便利化和资源配置优化提升效率；同时，这种促进作用往往伴随明显的空间溢出效应和区域差异^[14、15]。在微观层面，一些研究从企业层面展开分析，认为科技金融及相关政策可以通过缓解融资约束、增强创新激励、改善资本获取条件和降低融资成本，进而显著提升企业全要素生产率^[16-18]。同时，部分文献关注科技金融对全要素生产率影响的中介传导机制。一方面，科技金融可以通过促进技术创新提升技术效率与知识产出，进而推动生产率增长；另一方面，科技金融还能通过引导资本流向更高效率部门、促进产业结构调整和改善金融资源配置效率，对全要素生产率产生影响^[19]。

综上，现有研究针对科技金融和全要素生产率已进行了较为系统的探讨，但仍存在进一步拓展的空间。第一，从研究对象看，已有文献多以中国全国样本、东中部地区或企业整体样本为主，对西部地区这一金融发展基础相对薄弱、创新资源相对不足但政策扶持力度较强的区域关注不够，难以充分揭示科技金融在该类地区的真实效应。第二，从研究内容看，关于科技金融与全要素生产率的研究虽然逐渐增多，但对作用机制的识别仍不够系

1 https://www.gov.cn/zhengce/202510/content_7046050.htm

统,尤其是在人力资本积累与产业结构高级化两条路径上的考察仍较少。本文将进一步基于上述研究缺口开展拓展分析。

二、理论分析与研究假说

相较于传统金融,科技金融更强调对创新活动的支持以及对高技术产业的资源倾斜,其作用不仅体现在直接改善融资条件,还可以通过多种路径影响区域经济运行方式和生产效率。因此,有必要从直接效应与间接机制两个层面系统分析科技金融对全要素生产率的影响逻辑。

(一) 直接效应分析

科技金融能够通过优化资源配置与激励技术进步直接推动地区全要素生产率提升。一方面,科技金融通过多元化融资渠道、风险分担机制以及信息技术手段,降低科技型企业融资门槛,能够有效缓解创新活动融资约束,从而促进区域扩大研发投入和技术创新,最终推动地区全要素生产率提升^[16, 17]。另一方面,科技金融通过引导资金流向高效率部门与新兴产业,优化资源在不同产业和主体之间的配置结构,从而提升区域整体经济效率。

对于中国西部地区而言,由于长期存在金融供给不足、资本市场发展滞后以及资源配置效率偏低等问题,科技金融的引入能够在更大程度上发挥“边际改善效应”。随着金融资源向创新领域和高生产率部门集中,低效率部门的资源占比下降,区域整体全要素生产率将随之提升。同时,科技金融还通过降低信息不对称与交易成本,提高市场运行效率,进一步强化其对西部地区全要素生产率的促进作用^[14]。据此,本文提出如下假说:

H₁: 科技金融能够有效推动中国西部地区全要素生产率实现增长。

(二) 机制分析

1. 人力资本水平

从科技金融对人力资本的影响来看,科技金融通过支持高新技术产业与创新型企业发展,增加对高技能劳动力的需求,从而促进教育投入和人才集聚,推动人力资本积累。同时,金融支持创新活动能够改善人才激励机制,提高劳动力技能水平与知识积累,从而提升人力资本质量。此外,科技金融发展还可以通过完善教育融资、促进教育资源配置优化等方式,进一步强化人力资本形成。

从人力资本对全要素生产率的影响来看,人力资本是技术进步与创新扩散的重要载体。较高的人力资本水平不仅能够提升劳动者的生产效率,还能够增强对新技术的吸收与应用能力,从而推动技术效率提升和知识外溢效应的形成。相关研究指出,人力资本能够通过促进技术创新和提升资源配置效率,对全要素生产率产生显著正向影响^[10, 11]。对于中国西部地区而言,人力资本水平的提升能够有效破解其创新人才匮乏、技术转化能力不足的发展瓶颈,依托人才红利释放创新动能,显著带动区域全要素生产率稳步提高。基于此,本文提出如下假说:

H₂: 科技金融通过提高中国西部地区人力资本水平推动全要素生产率增长。

2. 产业结构高级化

从科技金融对产业结构的影响来看,科技金融通过引导资本流向技术密集型和高附加值产业,促进产业由低端向高端转型。

同时,科技金融支持创新活动和新兴产业发展,有助于推动产业结构从以传统制造业为主向以服务业和高技术产业为主转变,从而实现产业结构高级化^[6]。此外,科技金融通过改善融资环境与资源配置效率,促进企业进入新兴行业,加快产业结构调整进程。

从产业结构高级化对全要素生产率的影响来看,产业结构调整是要素在不同部门之间重新配置的过程。产业结构高级化通过发展高附加值产业与知识密集型产业,增强技术溢出与规模经济效益,从而推动全要素生产率提升^[12, 13]。对于中国西部地区而言,产业结构高级化能够有效破解其产业层次偏低、资源配置效率不高的发展困境,通过结构优化释放资源配置红利,显著提升区域全要素生产率水平。据此,本文提出如下假说:

H₃: 科技金融通过促进中国西部地区产业结构高级化推动全要素生产率提升。

科技金融影响中国西部地区全要素生产率的影响机制图如下图1所示:

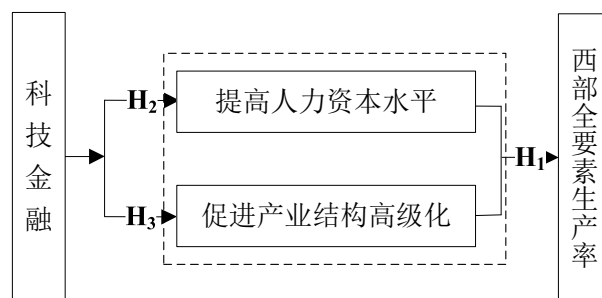


图1: 影响机制图

三、研究设计

2010年,中华人民共和国科学技术部发布《关于印发促进科技和金融结合试点实施方案的通知》,为科技金融试点提供了纲领。随后,中国于2011年、2016年先后推出科技与金融融合试点政策,共计确定了16个试点地区和9个试点城市。在此基础上,本文为识别科技金融试点政策对中国西部地区全要素生产率的影响,将中国“促进科技和金融结合试点”视为外生政策冲击,构建准自然实验,利用中国西部地区2000—2023年地级市面板数据开展多时点双重差分检验,进一步从人力资本和产业结构高级化两个维度考察科技金融影响全要素生产率的传导路径。

(一) 模型设计

科技金融试点分两期设立,因此在不同科技金融试点城市的实施时间并不完全一致,具有典型的多时点政策冲击特征,因此本文采用多时点双重差分模型对政策效应进行识别。具体设定如下:

$$TFP_{i,t} = \alpha_0 + \alpha_1 FT_{i,t} + X_{i,t} + \mu_i + \lambda_t + \varepsilon_{i,t} \quad (1)$$

其中, i 表示地级市, t 表示年份; $TFP_{i,t}$ 为被解释变量,表示中国西部地区地级市的全要素生产率; $FT_{i,t}$ 为核心解释变量,表示科技金融试点政策; $X_{i,t}$ 为控制变量组; μ_i 为城市固定效应; λ_t 为年份固定效应; $\varepsilon_{i,t}$ 为随机扰动项。回归系数 α_1 反映科技金融对中国西部地区全要素生产率的影响,若 α_1 显著为正,则表明科技

金融能够显著促进中国西部地区全要素生产率提升，从而支持本文的研究假说。

（二）变量选取

1. 被解释变量

本文以中国地级市层面的全要素生产率作为被解释变量。参考已有研究思路，将 GDP 作为产出指标，将资本存量以及全社会从业人员作为投入指标，使用 OLS 方法测算地级市全要素生产率指标（TFP），并将其作为被解释变量^[8]。该变量值越大，说明样本城市全要素生产率越高。

2. 核心解释变量

本文构建科技金融试点政策变量 FT_{it} 作为核心解释变量。 FT_{it} 为个体虚拟变量 $treat_i$ 和时间虚拟变量 $post_{it}$ 的交乘项。若某城市属于科技金融试点地区，则 $treat_i=1$ ，否则为 0；在政策实施年份及之后， $post_{it}=1$ ，否则为 0。

3. 控制变量

参考以往文献，本研究采用的控制变量如下所示：一是经济发展程度（Rgdp），计算人均 GDP 取自然对数后的值予以衡量；

二是人口集聚程度（Pop），通过对地区户籍人口除以行政区划面积的比值取自然对数来表征；三是产业结构（Str），利用第三产业增加值占地区生产总值的比例加以测度；四是金融发展程度（Fin），利用金融机构存贷款余额与地区 GDP 的比值作为代理变量^[15]。

（三）数据来源与描述性统计分析

本文在考虑数据可得性和有效性的前提下，选取 2000–2023 年中国西部地区 84 个地级市数据构建面板数据集，其中处理组城市 12 个，其余城市为对照组城市，本文构建变量所使用的数据主要来源为《中国城市统计年鉴》。

表 1 报告了描述性统计结果。总体来看，样本期内中国西部地区地级市全要素生产率均值为 0.968，标准差为 0.445，最小值为 0.208，最大值为 1.746，表明不同城市之间生产效率水平存在显著差异。其余变量在样本城市间也表现出较大离散性，说明中国西部地区在经济基础、人口集聚、金融发展和产业结构方面存在明显异质性，这为后续识别科技金融政策对全要素生产率的影响提供了较好的样本基础。

表 1：描述性统计

变量	(1)	(2)	(3)	(4)	(5)
	样本量	平均值	标准差	最小值	最大值
TFP	2,016	0.968	0.445	0.208	1.746
FT	2,016	0.070	0.255	0	1
Rgdp	2,016	9.916	1.068	7.145	12.880
Pop	2,016	5.119	1.021	1.547	7.020
Str	2,016	0.386	0.108	0.086	0.872
Fin	2,016	2.376	1.313	0.560	14.420

四、实证结果

为验证本文研究假说，本研究首先利用 DID 模型进行基准回归分析，并进一步通过平行趋势检验与多种稳健性检验验证结果的可靠性，从而为后续机制分析提供基础。

（一）基准回归

表 2 为控制城市固定效应和年份固定效应条件下科技金融试点政策对中国西部地区全要素生产率的基准回归结果。其中，列（1）为未经聚类标准误调整的回归结果，列（2）经聚类标准误调整后的结果。核心解释变量 FT 的估计系数均显著为正，意味着科技金融试点政策对我国西部地区全要素生产率增长具有显著的推动作用，假说 H_1 得证。该结果表明，科技金融试点政策已经成为推动中国西部地区全要素生产率提升的重要制度工具，这为后续影响机制检验提供了经验基础。

表 2：基准回归结果

变量	(1)	(2)
	TFP	TFP
FT	0.043*** (0.012)	0.043* (0.025)
Rgdp	0.007 (0.015)	0.007 (0.029)

Pop	-0.014 (0.028)	-0.014 (0.052)
Str	0.103** (0.049)	0.103 (0.091)
Fin	0.007 (0.005)	0.007 (0.010)
常数项	0.912*** (0.201)	0.912** (0.385)
聚类标准误调整	否	是
样本量	2,016	2,016
R ²	0.963	0.963

注：括号内为稳健标准误；***，** 和 * 分别代表在 1%，5% 和 10% 的显著性水平上显著；以上回归均控制城市固定效应和年份固定效应；下表同。

（二）平行趋势检验

处理组与对照组在政策冲击前的时间变动趋势保持一致，是双重差分模型估计结果可靠的重要前提。为验证这一前提，本文以样本第一期为基期进行平行趋势检验。平行趋势结果如图 2 所示，由于前后窗口期过长，本文将样本归并为前后九期。根据平行趋势检验结果，在政策实施之前，各期系数整体不显著，且波动较小，表明处理组与对照组在政策实施前的全要素生产率变化趋势基本一致，满足平行趋势假设。

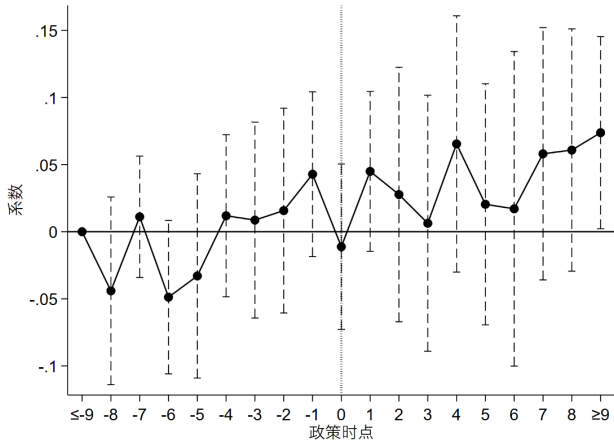


图2：平行趋势检验结果

（三）稳健性检验

本文将进一步从多个角度进行稳健性检验，结果如表3列（1）~（4）所示。

第一，替换被解释变量。这里采用基于系统广义矩估计（SGMM）方法测算的全要素生产率替换基准回归模型中的TFP变量，回归结果如表3列（1）所示。结果显示，虚拟变量FT的估计系数依然显著为正，说明本文结论不依赖于特定的生产率测量方式，具有较强稳健性。

第二，增加控制变量。参照现有研究，本研究采用地方一般公共预算支出占地区GDP的比重来度量政府干预水平（Gov），将其纳入控制变量组来控制政府干预对区域发展的潜在影响^[20]，回归结果如表3列（2）所示。结果表明，在引入该控制变量后，核心解释变量的系数仍保持显著为正，说明基准模型并未因遗漏关键变量而产生明显偏误。

第三，剔除特殊样本。参照已有研究思路，剔除2020年样本，以排除2020年公共卫生事件对估计结果的干扰^[21]，回归结果如表3列（3）所示。回归结果显示，科技金融试点变量的估计系数依然显著为正，具有较好的时间稳健性。

第四，更换回归模型。借鉴已有文献思想，采用随机森林因果推断模型替换DID模型对政策效应进行重新估计^[22]，回归结果如表3列（4）所示。结果显示，科技金融试点变量仍然显著为正，进一步说明本文基准回归结果在不同估计方法下具有一致性。这一结果也表明，本文结论不受模型设定形式的限制，具有较强的稳健性。

表3：稳健性检验

变量	替换变量	增加控制变量	剔除特殊样本	更换回归模型
	(1)	(2)	(3)	(4)
FT	0.067* (0.038)	0.043* (0.025)	0.043* (0.024)	0.134*** (0.031)
Rgdp	0.013 (0.039)	0.006 (0.031)	0.006 (0.029)	
Pop	-0.028 (0.064)	-0.014 (0.052)	-0.022 (0.051)	
Str	0.068 (0.106)	0.103 (0.090)	0.090 (0.086)	
Fin	0.006 (0.010)	0.007 (0.010)	0.007 (0.010)	

Gov	-0.010 (0.084)			
常数项	0.881* (0.528)	0.917** (0.397)	0.964** (0.374)	0.002 (0.005)
聚类标准误调整	是	是	是	是
样本量	2,016	2,016	1,932	2,016
R ²	0.855	0.963	0.963	

（四）机制分析

为进一步揭示科技金融试点政策影响中国西部地区全要素生产率的内在路径，本文使用两步法从人力资本水平与产业结构高级化两个维度进行机制分析，构建的模型如下所示：

$$M_{i,t} = \alpha_0 + \alpha_1 DFI_{i,t} + X_{i,t} + \mu_i + \lambda_t + \varepsilon_{i,t} \quad (2)$$

其中， $M_{i,t}$ 为机制变量，具体为人力资本水平和产业结构高级化水平。借鉴现有研究，通过高校在校生规模占区域总人口的比例来表征人力资本积累水平（Edu）^[23]；以第三产业增值与第二产业增值的比值来衡量产业结构高级化水平（Ais）^[12]。式（2）其他变量含义与式（1）中一致。

表4列（1）展示了以人力资本水平作为被解释变量的回归结果。可以看出，科技金融虚拟变量FT的回归系数显著为正，表明科技金融政策能够有效提升中国西部地区人力资本水平。人力资本作为技术进步与创新扩散的重要载体，能够显著提高劳动生产效率，并通过增强技术吸收能力与知识溢出效应推动全要素生产率提升^[10、11]。综上，科技金融通过促进人力资本积累，对中国西部地区全要素生产率产生间接推动作用，假说H₂得证。

表4列（2）展示了以产业结构高级化水平作为被解释变量的回归结果。结果显示，科技金融虚拟变量FT的回归系数显著为正，说明科技金融试点政策推动城市实现产业结构高级化，假说H₃得证。产业结构高级化通过促进要素在不同部门之间重新配置，提高资源配置效率，并强化技术外溢与规模经济效应，从而显著提升全要素生产率^[12、13]。

表4：影响机制检验结果

变量	人力资本水平	产业结构高级化
	(1)	(2)
FT	0.004* (0.003)	0.123*** (0.040)
Rgdp	0.001 (0.003)	0.024 (0.070)
Pop	0.009* (0.005)	-0.065 (0.101)
Str	0.018 (0.012)	5.477*** (0.466)
Fin	0.003** (0.001)	-0.000 (0.026)
常数项	-0.060 (0.046)	-1.010 (0.870)
聚类标准误调整	是	是
样本量	2,016	2,016
R ²	0.870	0.896

（五）异质性分析

鉴于我国西部地区各城市在经济发展水平与科技创新能力等维度存在显著异质性，这一区域差异可能使得科技金融对全要素生产率的作用效果呈现出明显区别。为此，本文进一步从地理区位与创新水平开展分组回归分析。

首先，本文将全部城市样本划分为西北地区与西南地区进行分组回归，回归结果如表5列（1）~（2）所示。其中，西北地区回归系数显著为正，而西南地区回归系数不显著。这意味着科技金融对西北地区全要素生产率的促进作用更为显著，而在西南地区的影响相对较弱。这一结果可能与区域经济结构及发展阶段差异有关。西北地区整体经济发展水平相对较低，传统金融供给不足问题更为突出，因此科技金融在缓解融资约束、改善资源配置方面的边际作用更大，从而对全要素生产率的促进效应更加明显。相比之下，西南地区市场化程度较高、产业结构相对完善，科技金融的边际提升空间相对有限。

从科技创新维度出发，本文使用地级市专利申请数量和专利授权数量之和的自然对数衡量地级市创新水平。同时，以城市创新水平中位数为界将样本划分为高创新水平城市和低创新水平城市分别进行分组回归，回归结果如表5列（3）~（4）所示。可以看出，列（3）回归系数显著为正，列（4）回归系数则不显著。这说明科技金融试点政策在高创新水平城市中的促进作用更为显著，而在低创新水平城市中的影响相对有限。上述结果表明，科技金融政策效应的发挥在很大程度上依赖于既有创新基础。对于高创新水平城市而言，当地通常拥有更完善的创新生态、更活跃的研发活动以及更强的技术吸收和成果转化能力，因此科技金融资源更容易转化为创新产出和生产率提升。相比之下，低创新水平城市尽管也能从政策中获得一定金融支持，但由于创新主体较弱、技术积累不足和创新环境不完善，科技金融的资源投入难以在短期内充分转化为效率增量，从而使政策效应相对不显著。

表5：异质性检验结果

变量	西北地区	西南地区	创新水平高	创新水平低
	(1)	(2)	(3)	(4)
FT	0.060*	0.020	0.052***	-0.036
	(0.035)	(0.031)	(0.018)	(0.064)
Rgdp	0.066*	-0.100**	0.088***	-0.078**
	(0.038)	(0.042)	(0.031)	(0.030)
Pop	0.040	-0.003	-0.064	0.055
	(0.095)	(0.050)	(0.064)	(0.084)
Str	0.054	0.113	0.096	0.185*
	(0.124)	(0.117)	(0.099)	(0.105)
Fin	0.018**	-0.006	-0.025	0.012
	(0.009)	(0.012)	(0.015)	(0.010)
常数项	0.018	1.953***	0.560	1.317**
	(0.610)	(0.497)	(0.386)	(0.514)
聚类标准误调整	是	是	是	是
样本量	912	1,104	864	1,152
R ²	0.968	0.958	0.947	0.970

五、结论与政策建议

（一）研究结论

基于中国科技金融试点政策构建的准自然实验，本文利用2000—2023年中国西部地区地级市面板数据，采用多时点双重差分模型，系统考察了科技金融对我国西部地区全要素生产率的作用效果及内在传导机制。研究结果表明：

第一，科技金融试点政策显著提升了中国西部地区全要素生产率，表明以科技金融为代表的金融创新工具在促进区域经济增长由要素驱动向效率驱动转型方面具有重要作用。

第二，机制分析结果显示，科技金融能够借助提升区域人力资本积累水平、加快产业结构高级化两大渠道，对全要素生产率产生间接驱动效应。一方面，科技金融的发展有助于提高劳动生产效率，通过增强技术吸收能力与知识溢出效应推动全要素生产率提升；另一方面，科技金融通过引导资本流向高技术、高附加值产业，促进产业结构由低端向高端转型，从而释放结构性效率红利。

第三，异质性分析表明，科技金融试点政策对全要素生产率的提升效应在不同地区和不同创新基础条件下存在明显差异。从地理区位维度来看，政策在西北地区的促进作用更为显著，说明在金融发展相对滞后的地区，科技金融具有更强的边际改善效应。从创新水平维度来看，科技金融在高创新水平城市中的作用更加突出，而在低创新水平城市中的效果相对有限。这表明，科技金融政策的实施效果在很大程度上依赖于区域创新基础，其作用发挥具有明显的条件依赖性。

（二）政策建议

基于上述研究结论，结合中国西部地区经济发展实际，本文从完善科技金融体系、强化人力资本积累赋能、推动产业结构高级化以及促进差异化政策实施等方面提出如下政策建议：

1. 加快完善科技金融服务体系，提升金融资源配置效率。本文基准回归结论表明，金融科技能够显著提升全要素生产率。因此，应进一步完善面向科技创新的多层次金融服务体系，推动银行、资本市场与数字金融平台协同发展，构建覆盖企业全生命周期的科技金融支持机制。同时，针对中国西部地区金融供给不足的问题，应鼓励金融机构加大对科技型中小企业的信贷投放力度，创新抵押担保方式，提升信用评估能力，降低融资门槛。此外，应发展股权融资、风险投资等直接融资渠道，引导长期资本进入科技创新领域，从而提高资源配置效率并增强生产率增长动力。
2. 重视人力资本积累赋能，夯实生产率增长人才根基。本文机制检验结果表明，科技金融能够通过提升区域人力资本水平，进而推动中国西部地区全要素生产率提升。因此，应依托科技金融政策引导，加快构建与创新发展的相适应的人力资本培育体系。一方面，鼓励金融机构开发适配人才培养、科研创新、技能提升的金融产品，支持高校、科研院所与企业联合开展应用型人才培

养,提高劳动力整体素质与技术应用能力;另一方面,通过金融贴息、风险补偿等方式,加大对高端人才引进、科研团队建设的资金支持,优化人才发展环境。同时,健全人力资本与科技金融资源的匹配机制,推动知识要素、技术要素与资本要素高效融合,让人力资本红利持续转化为西部地区全要素生产率增长的重要动力。

3. 推动产业结构高级化,释放结构性效率红利。机制分析表明,产业结构高级化是科技金融影响中国西部地区全要素生产率的有效作用机制。因此,应加快推动西部地区产业结构转型。具体而言,应通过科技金融政策引导资本向战略性新兴产业、高技术制造业和现代服务业集聚,推动传统产业技术改造和绿色转型,实现产业结构合理化与高级化协同推进,进一步提升中国西部地区全要素生产率水平。

4. 实施差异化科技金融政策,提升政策精准性。考虑到科技金融政策在不同区域和创新水平城市中的效果存在显著差异,应更加注重政策的分类施策与精准支持。对于中国西北地

区,应进一步加大政策倾斜力度,重点解决金融供给不足问题,释放科技金融的边际效应;对于中国西南地区,则应侧重优化金融资源配置效率,提高政策实施质量。对于高创新水平城市,应强化科技金融与创新体系的深度融合,推动高质量发展;而对于低创新水平城市,则应优先提升创新基础条件,通过基础设施建设、人才引进和创新平台培育等措施,为科技金融发挥作用创造条件。

5. 强化科技创新基础,提升科技金融政策的转化效率。鉴于科技金融在高创新水平城市中的作用更为显著,应着力夯实西部地区创新基础,提升科技金融资源的吸收与转化能力。一方面,应加大对科研基础设施、创新平台和高水平人才的投入,完善区域创新体系;另一方面,应鼓励企业加大研发投入,提升自主创新能力,促进科技成果转化。通过构建“科技—金融—产业”良性循环机制,使科技金融资源能够更高效地转化为生产率提升的现实动力。

参考文献

- [1] 曹颖, 尤建新, 卢锐等. 我国科技金融发展指数实证研究 [J]. 中国管理科学, 2011, 19: 134-40.
- [2] 薛莹, 胡坚. 金融科技助推经济高质量发展: 理论逻辑、实践基础与路径选择 [J]. 改革, 2020: 53-62.
- [3] 李俊霞, 温小霓. 中国科技金融资源配置效率与影响因素关系研究 [J]. 中国软科学, 2019: 164-74.
- [4] 芦锋, 韩尚容. 我国科技金融对科技创新的影响研究——基于面板模型的分析 [J]. 中国软科学, 2015: 139-47.
- [5] 胡欢欢, 刘传明. 科技金融政策能否促进产业结构转型升级? [J]. 国际金融研究, 2021: 24-33.
- [6] 汪淑娟, 谷慎. 科技金融对中国经济高质量发展的影响研究——理论分析与实证检验 [J]. 经济学家, 2021: 81-91.
- [7] 郭庆旺, 贾俊雪. 中国全要素生产率的估算: 1979—2004 [J]. 经济研究, 2005: 51-60.
- [8] 范剑勇. 产业集聚与地区间劳动生产率差异 [J]. 经济研究, 2006: 72-81.
- [9] 陈超凡. 中国工业绿色全要素生产率及其影响因素——基于 ML 生产率指数及动态面板模型的实证研究 [J]. 统计研究, 2016, 33: 53-62.
- [10] 逯进, 李婷婷. 产业结构升级、技术创新与绿色全要素生产率——基于异质性视角的研究 [J]. 中国人口科学, 2021: 86-97+128.
- [11] 赵宸宇, 王文春, 李雪松. 数字化转型如何影响企业全要素生产率 [J]. 财贸经济, 2021, 42: 114-29.
- [12] 干春晖, 郑若谷, 余典范. 中国产业结构变迁对经济增长和波动的影响 [J]. 经济研究, 2011, 46: 4-16+31.
- [13] 于斌斌. 产业结构调整与生产率提升的经济增长效应——基于中国城市动态空间面板模型的分析 [J]. 中国工业经济, 2015: 83-98.
- [14] 张腾, 刘阳. 科技金融发展是否促进了全要素生产率的提高?——基于空间计量模型的研究 [J]. 金融与经济, 2019: 29-35.
- [15] 冯锐, 马青山, 刘传明. 科技与金融结合对全要素生产率的影响——基于“促进科技和金融结合试点”准自然实验的经验证据 [J]. 科技进步与对策, 2021, 38: 27-35.
- [16] 刘少波, 吴玥. 科技金融政策是否提升了企业全要素生产率? [J]. 产经评论, 2022, 13: 117-32.
- [17] 阳灿, 刘明显, 刘忍妹. 科技金融政策提高了企业全要素生产率吗?——来自中国工业企业的证据 [J]. 武汉金融, 2022: 33-41.
- [18] 张晓莉, 张露文, 孙琪琪. “双循环”下科技金融对企业全要素生产率的影响——基于国家科技金融试点政策的准自然实验 [J]. 金融理论与实践, 2022: 1-12.
- [19] 巴曙松, 白海峰, 胡文韬. 金融科技、企业全要素生产率与经济增长——基于新结构经济学视角 [J]. 财经问题研究, 2020: 46-53.
- [20] 唐松, 赖晓冰, 黄锐. 金融科技如何影响全要素生产率: 促进还是抑制?——理论分析框架与区域实践 [J]. 中国软科学, 2019: 134-44.
- [21] Romer C D, Romer D H. The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks [J]. American Economic Review, 2010, 100(3): 763-801.
- [22] Chernozhukov V, Chetverikov D, Demirer M, et al. Double/debiased machine learning for treatment and structural parameters [J]. The Econometrics Journal, 2018, 21(1): C1-C68.
- [23] 詹新宇, 刘文彬. 中国式财政分权与地方经济增长目标管理——来自省、市政府工作报告的经验证据 [J]. 管理世界, 2020, 36: 23-39+77.

广州市智慧场馆平台的使用现状及消费意愿调查

王婷, 李迪, 陈颖熙, 刘诺柠

广州大学 经济与统计学院, 广东 广州 510006

DOI:10.61369/ASDS.2026060015

摘 要 : 在“健康中国 2030”战略背景下,完善全民健身公共服务体系成为刚需。本文以广州市为调研地,聚焦智慧场馆平台应用功能单一、程序稳定性差等问题,综合运用文本挖掘、深度访谈、问卷调查及多种统计分析方法展开系统研究。通过 Python 爬取群体通等五个平台的用户评论,运用 NLPPIR 进行词频统计,获取现有用户反馈;对 5 位公共场馆管理者、11 位社会场馆老板及 13 位团体运动者进行深度访谈,多维度挖掘各方预期;在 35 个球馆发放问卷,回收有效问卷 605 份。研究采用主成分分析与灰色关联分析,发现平台性能对效果满意度影响最大;通过聚类分析将潜在用户划分为重要潜在用户等五类群体。最终从公共场馆管理者与社会场馆老板双视角,为平台优化升级与潜在市场拓展提出针对性建议。

关 键 词 : 智慧场馆平台; 认知与使用情况; 效果满意度; 潜在市场

A Survey on the Usage Status and Consumption Intention of Guangzhou's Smart Venue Platform

Wang Ting, Li Di, Chen Yingxi, Liu Nuoning

School of Economics and Statistics, Guangzhou University, Guangzhou, Guangdong 510006

Abstract : Against the backdrop of the "Healthy China 2030" strategy, improving the public fitness service system has become an essential requirement. This study takes Guangzhou as the research site and focuses on issues such as the limited functionality and poor program stability of smart venue platforms. A systematic investigation is conducted using text mining, in-depth interviews, questionnaire surveys, and multiple statistical analysis methods. User reviews from five platforms including Quntitong were crawled using Python, and word frequency statistics were performed with NLPPIR to obtain existing user feedback. In-depth interviews were conducted with 5 public venue managers, 11 private venue owners, and 13 group ball game players to explore expectations from multiple dimensions. Questionnaires were distributed across 35 ball courts, yielding 605 valid responses. Principal component analysis and grey relational analysis revealed that platform performance has the greatest impact on effectiveness satisfaction. Cluster analysis categorized potential users into five groups, including key potential users and key development users. Finally, from the dual perspectives of public venue managers and private venue owners, targeted recommendations are proposed for platform optimization and potential market expansion.

Keywords : smart venue platform; awareness and usage; effectiveness satisfaction; potential market

引言

伴随着全民健身战略的深入实施,我国体育公共服务体系加速向智慧化转型。国务院 2021 年发布的《全民健身计划(2021-2025 年)》明确提出构建更高水平的全民健身公共服务体系^[1],强调推动线上和智能体育赛事活动开展,建设国家全民健身信息服务平台,提供健身设施查询预订、体育培训报名、健身指导等智慧化服务。广州市体育局在《2023 年广州市体育工作计划》中也提出,要完善全民健身服务体系,丰富“群体通”应用场景,推动更多体育场馆上线提供优质体育服务^[2]。政策驱动下,“智慧场馆平台”作为连接体育场馆与运动者的线上交互载体应运而生^[3]。

基金项目: 2024 年度“大学生创新创业训练计划”。

作者简介:

王婷,女,广州大学经济与统计学院,本科生,研究方向:金融学;

李迪,女,广州大学经济与统计学院,本科生,研究方向:金融学;

陈颖熙,女,广州大学经济与统计学院,本科生,研究方向:统计学;

刘诺柠,女,广州大学经济与统计学院,本科生,研究方向:统计学。

近年来,随着人们健康意识的增强,参与团体球类运动的人数不断增加,不少体育场馆为扩宽宣传渠道、提高知名度,在运营线下业务的同时引入线上智慧化服务^[4],通过入驻智慧场馆平台满足市民对运动场地及基础设施的需求^[5]。目前广州市主要使用的平台包括群体通、趣运动、掌馆通、VAL排球圈、GolfLive等。然而,这些平台大多只实现了场馆预订、健身地图、赛事活动、加入社群、场馆评价等基础功能,虽然实现了服务的智慧化,但并未体现个性化、亲民化的特点^[6]。团体球类运动重点考验队员的协作能力,参与者在默契与实力较为匹配的基础上才能获得更好的体验感,但许多运动者仍面临因时间、距离、实力、性格等因素差异而找不到契合球友的困境,现有智慧场馆平台尚无法满足这类个性化需求^[7]。

本文以广州市为调研地,聚焦智慧场馆平台功能单一、个性化服务缺失等问题。按照“是否使用过智慧场馆平台”,将调查对象分为使用过平台的用户与未使用过的潜在用户两类人群。针对已使用用户,重点探究其对平台功能完备程度、资源信息质量、匹配队友专业程度等方面的效果满意度及其影响因素;针对未使用用户,深入挖掘其未使用平台的内外部障碍及未来使用意愿。同时,从公共场馆管理者与社会场馆经营者双视角,了解平台运营现状、盈利模式及创新预期,多维度把握智慧场馆平台的发展瓶颈与优化方向。

为解决上述问题,本文以广州市团体球类运动参与者、公共场馆管理者及社会场馆经营者为调研对象^[8],采用多种研究方法展开系统研究:利用Python爬取群体通等五个平台的用户评论,运用NLPIR进行词频统计与情感分析,获取现有用户反馈;对5位公共场馆管理者、11位社会场馆老板及13位团体运动者进行深度访谈,多维度挖掘各方预期与改进方向;在广州市天河区、越秀区、海珠区、黄埔区、白云区5个主城区的35个球馆发放问卷,回收有效问卷605份,分析运动者对平台的认知、使用现状及满意度。研究采用主成分分析与灰色关联分析探究影响效果满意度的关键因子,通过聚类分析将潜在用户划分为重要潜在用户、重要发展用户、次要潜在用户、一般潜在用户和低价值潜在用户五类群体。最终从公共场馆管理者与社会场馆老板双视角,为智慧场馆平台的功能优化、服务升级与潜在市场拓展提出针对性建议。

一、调查方案设计与实施

(一) 调查内容

本文借助问卷调查与深度访谈,探究广州市团体球类运动参与者对智慧场馆平台的认知与使用现状、效果满意度及其影响因素,挖掘未使用平台者的使用意愿与潜在市场需求,并从公共场馆管理者和社会场馆经营者双视角了解平台运营现状与改进预期,为智慧场馆平台的功能优化、服务升级与市场拓展提供建议。

(二) 调查对象

本文通过问卷调查法,采用多阶段抽样,先后以主城区、场馆、场馆进出者为划分依据,对广州市35个球馆的团体球类运动参与者进行线下调研,共发放问卷650份。经数据预处理后,回收有效问卷605份,有效率达93.08%。

同时选取5位公共场馆管理者、11位社会场馆老板及13位团体球类运动参与者进行深度访谈,研究智慧场馆平台的运营现状、用户体验及改进方向。

(三) 问卷设计

1. 问卷框架

本文基于智慧场馆平台用户消费行为、功能使用偏好及市场运行现状,结合网络文本挖掘结果、访谈信息与调查目标,将问卷分为三个部分。第一部分为参与者基本信息与运动习惯,包含性别、年龄、学历、月薪、职业、运动频率、运动目的及运动花费金额;第二部分为智慧场馆平台认知与使用情况。针对使用用户,包含对平台的了解程度、使用过的具体平台、使用频率、花费金额、了解渠道、功能类型、运动效果及平台整体运行情况;针对未使用用户,包含使用意愿、意愿花费、期望功能、传播渠道及不使用原因;第三部分为市场建议,包含对平台发展的改善意见。

2. 问卷优化

正式调查前,团队开展预调查,以调查员所在球类运动者为目标总体,通过滚雪球抽样拦截场馆进出者并请求其引荐球友,共发放51份问卷,回收45份,有效率达88.23%。结合预调查中出现的问题,对问卷语言表达、题目设置进行了修改。

3. 变量定义与信效度检验

本文从平台功能评价、市场运行、未来期望及运动目标重要程度四方面构建量表。信度分析结果显示,各变量Cronbach's Alpha系数均>0.86,表明量表具有良好的内部一致性,问卷可信度较高。检验结果如下表所示:

表1: 信度检验结果

变量名称	ronbach's Alpha	项数	信度评价
团体球类运动目标重要程度	0.860	6	相当好
智慧场馆平台功能评价	0.985	5	非常好
智慧场馆平台市场运行打分	0.996	6	非常好
智慧场馆平台未来期望	0.980	5	非常好

效度分析结果显示,KMO值为0.970,Bartlett球形检验结果 p 值<0.001,量表效度良好,适用于模型构建与数据分析。检验结果如下表所示:

表2: KMO与Bartlett检验系数

项目	数值
KMO值	0.970
Bartlett球形度检验	16973.516
自由度df	231
P值	0.000***

(四) 访谈设计

本文针对不同对象设计访谈大纲。通过访谈公共场馆管理者,从供给侧了解智慧场馆平台的运营现状、政策支持及创新挑战;通过访谈社会场馆老板,从运营侧了解平台的盈利模式、入驻

困境及发展预期；通过访谈团体球类运动参与者，从需求侧了解用户的使用体验、功能偏好及改进建议。

二、统计建模与分析

（一）智慧场馆平台用户特征与满意度调查

1. 平台使用者的运动特征与人群画像

本文结合实地调查数据，将团体球类运动参与者按运动类型、年龄层、收入水平等维度进行划分。如图1所示，羽毛球占比29.1%，篮球17.8%，乒乓球16.1%，排球12.1%，足球12.5%，网球9.4%；其中羽毛球女性占比更高，乒乓球、篮球、足球男性占比更高，排球和网球男女比例较为均衡。如图2所示，排球、网球、篮球群体呈现年轻化特征，35岁以下占比分别为83%、68%、67%；乒乓球46岁以上占比达32%，为各球类最高；羽毛球和足球26-35岁占比分别达42%和45%。



图1：不同球类运动参与人数、性别占比

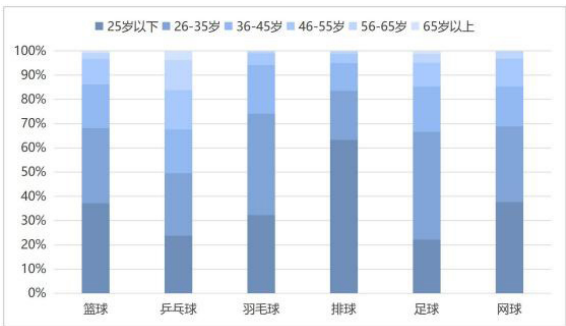


图2：不同球类运动中年龄层占比图

如图3所示，GolfLive平台用户平均月收入最高，反映高尔夫运动的高消费属性；趣运动和群体通用户收入分布较广，印证其价格分层策略扩大了用户覆盖范围。

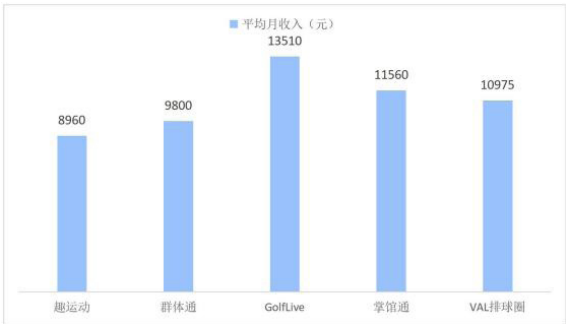


图3：不同平台适用人群收入占比图

基于上述特征，本文通过描述性统计与交叉分析，将平台使用者按运动偏好、年龄结构、收入水平进行细分。各类人群在运动类型、消费能力和平台选择上存在差异，却均对平台的便捷性、信息完备性和社交匹配功能表现出强烈需求。平台可聚焦于羽毛球、篮球等大众球类，针对不同年龄层推出差异化服务，满足用户多元化运动需求。

2. 平台认知程度与渠道偏好

受访者对智慧场馆平台比较了解及非常了解的占比达44%，听说过平台的占26%，七成以上受访群体对平台有所认知。受访者获取平台信息的渠道多元，主要来自亲友介绍、互联网媒介、展会及公益活动等。这表明平台宣传推广策略相对全面，口碑传播与线上推广效果显著，但仍有较大提升空间。建议平台加强宣传力度，举办赛事合作活动，推出实用有趣的功能，提升知名度与用户粘性。

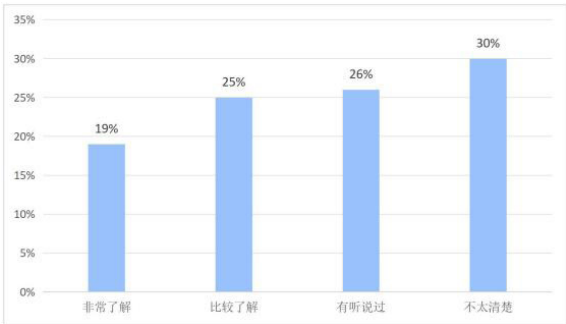


图4：对智慧场馆平台的认知程度图

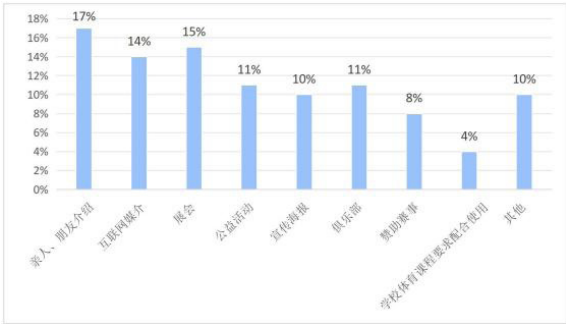


图5：对智慧场馆平台的认知渠道图

3. 平台满意度分析

本文综合平台功能特性与用户实际体验，从页面设计美观度与易用度、用户活跃度、安全性、功能多样性四个维度构建智慧场馆平台效果满意度指标体系，并引入用户资源投入和兴趣动机作为影响因素。

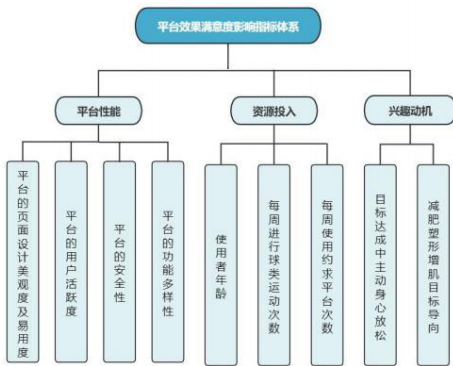


图6：平台效果满意度影响指标体系

本文采用主成分分析法对9个影响因子进行降维处理，提取方差累计贡献率达85%的前三个主成分：平台性能、资源投入和兴趣动机。进一步采用灰色关联分析计算各主成分与效果满意度的关联度。

表3：主成分载荷矩阵表

变量	主成分				
	1	2	3	4	5
年龄	-0.124	0.682	0.432	-0.365	-0.239
平均每周进行团体球类运动的次数	-0.083	0.717	-0.106	0.066	0.476
一周运用智慧场馆平台的平均次数	-0.068	0.757	0.351	0.444	-0.381
主打身心放松功能	0.561	-0.303	0.636	0.163	0.256
减肥增肌塑形功能	0.598	-0.153	0.688	-0.056	0.205
平台页面设计美观度易用度	0.826	0.157	-0.248	0.016	0.008
平台的用户活跃度	0.828	0.101	-0.190	0.020	-0.142
平台的安全性（如支付安全、隐私政策等）	0.804	0.092	-0.122	-0.092	-0.097
平台的功能多样性	0.824	0.166	-0.150	-0.057	-0.093

表4：灰色关联分析结果

评价项	关联度	排名
平台性能	0.743	1
资源投入	0.733	2
兴趣动机	0.710	3
补充因素1	0.672	4
补充因素2	0.667	5

平台性能与满意度的灰色关联度最大，说明页面设计美观度与易用度、用户活跃度、安全性及功能多样性对效果满意度影响最大。资源投入次之，表明用户运动积极性直接影响平台满意度。兴趣动机关联度为0.710，亦属高度相关。综上，若需进一步提升用户满意度，可重点关注平台性能指标，其中页面设计美观度与易用度权重较高且仍有提升空间，优化界面交互体验可最大程度提升整体满意度。

（二）智慧场馆平台潜在市场调查

1. 未使用平台者的使用意愿

在未使用过平台的受访者中，非常愿意和愿意使用的分别占42%和23%，态度中立占21%，不太愿意和非常抗拒仅占11%和3%。

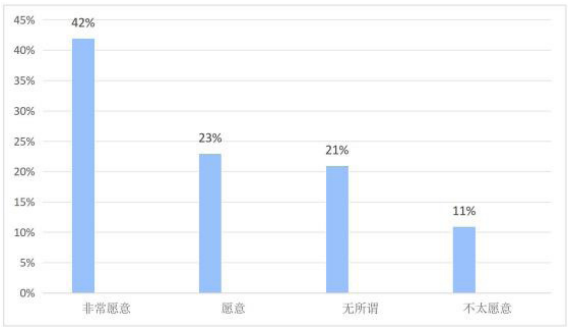


图7：未使用过平台的运动者的使用意愿图

潜在用户对运动数据记录功能预期最高，信息反馈功能次之，社区交流功能预期相对较低。



图8：对平台功能的评分图

2. 潜在用户聚类分析

本文采用K-means聚类法，以月薪、职业、每周运动次数、平均每次运动花费、使用意愿程度为变量，对241位潜在用户进行聚类。通过轮廓系数和DBI指标对比，确定最佳聚类数为4类，结果如下表：

表5：聚类分析结果

聚类种类	月薪	职业	平均每周进行团体球类运动的次数	平均每次参与团体球类运动的花费	意愿程度
1	3000以下	学生	2-3次	35元以下	非常愿意
2	6000—10000	政府机关干部 / 公务员 / 事业单位	2-3次	36-50元	愿意
3	6001—10000	体育相关从业人员	4-5次	51-70元	非常愿意
4	10001—16500	企业工作人员	0-1次	71元以上	无所谓

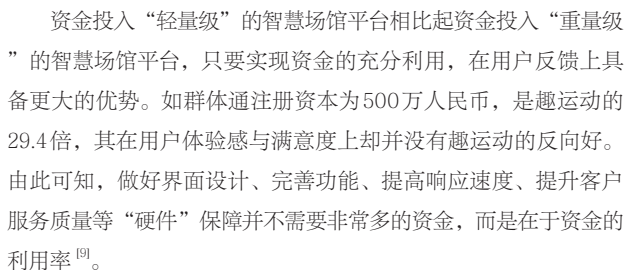
根据意愿程度特征，将潜在用户命名为三类。重要潜在用户：政府机关干部 / 公务员，月收入6000-10000元，运动频率和花费较高，意愿程度较高，偏好身心放松与社交目标。次要潜在用户：学生和体育行业从业人员，运动频率最高，意愿程度最高，偏好工作社交目标，占潜在用户38.18%。一般潜在用户：企业工作人员，月收入10001-16500元，运动次数少，意愿不强烈。平台可针对学生和体育从业人员推出高性价比服务，针对公

以个性化服务为特色，实现用户满意度与市场拓展的双重目标。

三、结论与建议

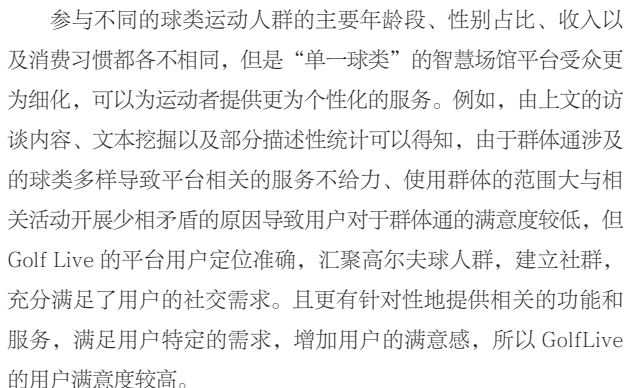
（一）调查结论

1. 资金投入量不决定平台绩效，其关键在于资本的有效利用率。

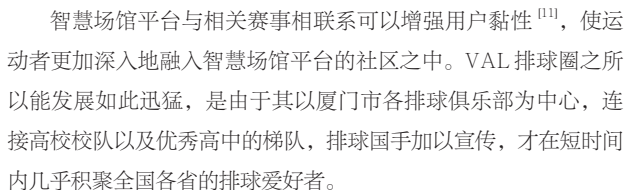


2. 平台用户规模大并不等同于用户使用意愿强。

3. 服务“单一球类”的平台比面向“多种球类”的平台满意度高。



4. 推动平台发展的关键因素是组织赛事, 连接俱乐部和高校圈层。



5. 影响用户使用平台满意度的关键因素是平台性能、资源投入和运动者的兴趣动机。

群体通存在运行速度慢、界面设计不佳、功能不稳定等问题,给用户带来不便和困扰,使其降低对平台的满意度。相反,如GolfLive平台性能良好,用户可以顺利完成操作,并获得良好的使用体验,使其对平台具有较高的满意度。于运动者自身而言,如果其更加重视对于运动项目所带来的目标成就感,则会更愿意使用平台,并对平台的服务和体验有更高的期待和满意度。

ed under A Commons Attribution-Non Commercial 4.0 International License. | 087

6. 政府机关干部、公务员和事业单位是智慧场馆平台的重要潜在用户。

固定的空闲时间、相对稳定的工作和收入是用户频繁使用智慧场馆平台的基础，对社交的需求以及较强的健康意识是用户使用智慧场馆平台的重要原因。

（二）对策和建议

1. 对政府的建议

对平台投入适度资金，将款项重点放在赛事组织、场地租赁、设备采购、宣传推广等方面。这不仅可以很大程度上提高全民对运动的积极性，还可以鼓励大家参加运动赛事从而激发其他更多的运动用户。除此之外，政府需要重新制定对平台的资助标准和资助程序，确保资金使用的透明度和合规性。

2. 对平台的建议

组建专项球类运动的相关社群，以供用户之间结交朋友、建立

团队，组织联赛等；平台要根据客户运动的目的和性质；建立“个性化”的隐私保护机制，根据用户的选择，平台可以调整隐私设置，提供更符合用户期望的个性化体验；平台要不断进行全面的性能测试与优化，确保平台的安全性以及功能多样性在能够满足需求，同时还要不断优化用户体验，如提高页面设计的美观度以及易用度，确保用户在使用平台时感到舒适；给用户平台培训与支持，使更多的用户能够充分利用平台功能；建立有效的用户活跃系统，及时回应用户的问题和反馈；可以在智慧场馆平台上引入奖励和激励机制，给予参与团体球类运动的用户一些实质性的回馈，如积分、优惠券等，这可以有助于激发用户的积极性以及提高对平台的满意度。

参考文献

- [1] 国务院. 全民健身计划（2021-2025年）[EB/OL]. (2021-08-03)[2025-2-14].https://www.gov.cn/zhengce/zhengceku/2021-08/03/content_5629218.htm.
- [2] 广州市体育局. 2023年广州市体育工作计划[EB/OL]. (2023-12-14)[2025-2-14].https://www.gz.gov.cn/zwgk/ghjh/zxgh/content/mpost_9382592.html.
- [3] 郑芳, 徐伟康. 我国智能体育：兴起、发展与对策研究[J]. 体育科学, 2019, 39(12): 14-24.
- [4] 张强, 王家宏. 数字赋能体育场馆智慧转型发展：突破动因、国外镜鉴与立体化路径[J]. 武汉体育学院学报, 2023, 57(07): 54-61.
- [5] 付紫硕, 陈元欣. 国外智慧体育场馆建设经验及启示[J]. 体育文化导刊, 2020(10): 40-46.
- [6] 杨小银, 曾建明, 朱俊鹏, 等. 美国体育场馆智慧化建设经验与启示[J]. 体育文化导刊, 2022(07): 39-44.
- [7] QUINTAL B. U.S. Bank Stadium: A Game-Changing, Multi-Purpose NFL Stadium[EB/OL]. (2019-01-22)[2025-2-14].<https://www.archdaily.com/784289/us-bank-stadium-a-game-changing-multi-purpose-nfl-stadium>.
- [8] 张强, 陈元欣. 国外智慧体育场馆：转型诉求、实践模式及本土启示[J]. 西安体育学院学报, 2023, 40(01): 20-27.
- [9] 国家体育总局. 体育总局办公厅印发《全民健身场地设施提升行动方案（2023-2025）》工作方案[EB/OL]. (2023-08-15)[2025-2-14].<https://www.sport.gov.cn/n315/n20001395/c25666308/content.html>.
- [10] 方琳, 邹月辉. 数字技术助力全民健身场馆智慧化：价值、现实审视与策略[J]. 湖北体育科技, 2024, 43(01): 103-107.
- [11] 国务院. 体育强国建设纲要[EB/OL]. (2019-09-02)[2025-2-14].https://www.gov.cn/zhengce/zhengceku/2019-09/02/content_5426485.htm.