

Discussion on the application effect of big data technology in Internet financial credit evaluation

Weiha Chen¹ Jing Zhao²

1. Guangzhou Business School, Guangzhou, Guangdong, 511363, China

2. Guangdong Vocational and Technical College of Mechanical and Electrical Engineering, Guangzhou, Guangdong, 510515, China

Abstract

In the digital economy era, internet finance has emerged as a pivotal force in advancing inclusive finance, while traditional credit assessment systems face limitations in data dimensions. This paper examines the application effectiveness of big data technology in internet finance credit evaluation, elaborating on its theoretical foundations and technical architecture. It analyzes specific implementations of big data in expanding data dimensions, optimizing algorithmic models, and restructuring risk control processes. Research findings indicate that big data technology significantly enhances risk identification capabilities, service efficiency, and coverage scope, driving business scale expansion and economic growth. However, challenges persist in data quality standardization and privacy protection compliance. To address these issues, improvements in data governance and innovations in privacy-preserving computation technologies are essential. With technological advancements and regulatory enhancements, big data credit evaluation will bolster financial system stability. Future research should explore new technological integrations and framework development under global data governance.

Keywords

big data; Internet finance; credit evaluation; risk control model; machine learning

大数据技术在互联网金融信用评估中的应用效果探讨

陈薇好¹ 赵静²

1. 广州商学院, 中国·广东 广州 511363

2. 广东机电职业技术学院, 中国·广东 广州 510515

摘要

在数字经济时代, 互联网金融成为推动普惠金融发展的关键力量, 而传统信用评估存在数据维度有限等问题。本文探讨大数据技术在互联网金融信用评估中的应用效果, 阐述其理论基础与技术架构, 分析大数据在数据维度扩展、算法模型优化、风控流程重构等方面的具体应用。研究表明, 大数据技术显著提升风险识别能力、服务效率与覆盖范围, 促进业务规模与经济效益增长, 但也面临数据质量与标准化、隐私保护与合规等挑战。对此, 需完善数据治理、创新隐私计算技术等。随着技术进步与监管完善, 大数据信用评估将为金融体系稳定提供支撑, 未来可探索新技术融合及全球数据治理下的体系构建。

关键词

大数据; 互联网金融; 信用评估; 风控模型; 机器学习

1 引言

数字经济时代, 互联网金融作为传统金融的重要补充, 已成为推动普惠金融发展的关键力量。根据 2024-2025 年中国互联网金融信用评级行业报告显示, 我国互联网金融信用评级市场规模正以年均 20% 以上的速度增长, 预计 2025 年将突破千亿元大关。信用评估作为金融风险控制的环节, 其准确性与效率直接关系到互联网金融行业的健康发展。

传统信用评估主要依赖央行征信数据, 采用逻辑回归等经典统计方法, 存在数据维度有限、评估滞后和覆盖人群不足等问题。数据显示, 央行征信系统收录的 11 亿自然人中, 有信贷数据的不足一半, 大量长尾客户难以获得合理的信用评估。随着大数据、人工智能等技术的发展, 互联网金融机构开始将社交数据、交易记录、行为特征等多维度信息纳入信用评估体系, 通过机器学习算法构建更精准的信用评分模型。基于此, 本文将对大数据技术在互联网金融信用评估中的应用效果展开研究。

【作者简介】陈薇好(2005-), 女, 中国河北定州人, 在读本科生, 从事互联网金融、大数据技术应用研究。

2 大数据信用评估的理论基础与技术架构

2.1 理论基础

信息不对称是导致信贷市场失灵的主要原因，传统金融机构通过严格的尽职调查降低信息不对称，但成本高昂且效率低下。大数据技术通过拓展信息获取渠道和优化信息分析方法，为缓解信息不对称提供了新路径^[1]。根据信息经济学理论，当评估主体能够获取更全面的信息时，逆向选择和道德风险发生的概率将显著降低。数据驱动决策理论认为，基于海量数据的量化分析能够克服人类决策中的主观偏差，提高预测准确性。在信用评估领域，这一理论体现为通过机器学习算法从历史数据中挖掘风险模式，建立违约概率预测模型。中国人民大学的研究表明，基于大数据的机器学习模型在区分好坏客户的 K-S 指标、预测准确性的 AUC 指标上均显著优于传统逻辑回归模型。

2.2 技术架构

大数据信用评估体系由数据层、算法层和应用层构成。数据层负责多源数据的采集与预处理，包括结构化的征信数据、交易记录和非结构化的行为日志、文本信息等。恒丰银行在构建信用评分模型时，不仅使用了央行征信和公积金数据，还引入了银联交易等外部数据，涵盖资产、收入、借贷、消费场景等多维度信息。

算法层是大数据信用评估的核心，主流技术有传统机器学习算法、集成学习算法、隐私计算技术。传统机器学习算法如逻辑回归、决策树等，具有可解释性强的特点，仍被广泛应用于信用评分模型的基础架构^[4]。芝麻信用在其评分体系中就采用了逻辑回归作为基础算法框架。集成学习算法如 GBDT（梯度提升决策树）和 XGBoost（极端梯度提升），通过组合多个弱分类器提高预测性能。研究显示，XGBoost 算法在信用评估中的准确率明显高于随机森林与 GBDT 算法。隐私计算技术包括联邦学习、多方安全计算和同态加密等，解决数据共享与隐私保护的矛盾。京东金融采用联邦学习技术，实现了在不共享原始数据的情况下与合作机构联合训练信用评估模型。

应用层负责将算法输出转化为实际的信贷决策，包括客户准入、额度授信和风险监控等环节。恒丰银行通过构建“数据-模型-策略”闭环，将信用评分模型应用于网贷新产品的客户引入和存量产品的迭代优化中。

3 大数据技术在信用评估中的具体应用

3.1 数据维度的扩展与创新

大数据信用评估突破了传统征信的局限，实现了数据来源的多元化^[2]。芝麻信用的评分体系包含信用历史、行为偏好、履约能力、身份特质和人脉关系五大维度，每个维度下又细分出数十甚至上百个变量，建模初期变量数可轻松破千。这些变量不仅包括传统的还款记录，还涵盖了支付频率、消费层次、社交网络等新型数据。在数据处理方面，大

数据技术解决了传统评估中的数据质量问题。通过缺失值清洗（剔除缺失率超过 50% 的字段）、变量去重（剔除皮尔逊系数 > 0.7 的高度相关变量）和标准化处理（采用 MAX-MIN 或 Z-score 方法），提高了数据的可用性。京东金融还对数据进行分类分级管理，对身份证号码等敏感信息采用高强度加密存储，平衡了数据利用与安全保护。

3.2 算法模型的优化与迭代

互联网金融机构根据业务场景选择合适的算法模型，并建立持续优化机制。恒丰银行针对不同信贷产品客群，采用逻辑回归和 XGB 集成算法回溯建模，在测试集和跨时间样本集上均表现优秀。芝麻信用则综合运用逻辑回归、决策树、随机森林等多种算法，通过加权计算得出最终信用分值，其公式为： $score=(1-P)*600+300$ ，其中 P 为预测违约概率。模型评估体系日益完善，主要采用混淆矩阵（查准率/查全率）和 ROC 曲线（AUC 值）作为核心指标。一般要求 AUC 值至少在 0.6 以上，优质模型可达到 0.8 以上。模型迭代机制确保了评估效果的稳定性，芝麻信用每周复盘坏账数据，分析模型失效维度并及时调整参数，京东金融的天机风控系统则能根据新数据自动更新模型。

3.3 风控流程的智能化重构

大数据技术实现了信贷风控全流程的智能化升级。在贷前环节，恒丰银行通过风控处理平台实时调用基石底座数据产品，根据准入策略和授信策略完成客户快速审批；在贷中环节，京东金融的天机风控系统实时监控交易行为和用户互动，识别异常模式；在贷后环节，恒丰银行定期对存量客户调用数据产品，识别潜在风险客户并触发预警策略。智能反欺诈成为信用评估的重要补充。京东金融的天机风控系统通过深度神经网络和集成学习方法，分析异常 IP 地址、设备指纹和用户行为模式，成功拦截大量伪造身份信息的恶意贷款行为。这种实时风控能力显著降低了欺诈损失，提升了信用评估的可靠性。

大数据技术还推动了风控流程的动态化与个性化升级。以微众银行为例，其基于用户实时生成的地理位置、消费时段、设备更换频率等多维度数据流，构建动态风险评估引擎，为不同信用等级用户设置差异化阈值，对高信用用户简化审批流程，对风险波动较大的用户加强监控频率。同时，引入知识图谱技术构建用户关系网络，精准识别“一人多号”“团伙伪装”等欺诈模式。例如，当系统发现 10 个以上账号共享同一设备且交易集中于高风险商户时，会自动触发联合风控机制，进一步提升了风险拦截的时效性与精准度。

4 大数据信用评估的应用效果分析

4.1 风险识别能力显著提升

大数据信用评估模型在风险识别方面表现出优异性能。恒丰银行引入大数据分析和机器学习技术后，信用风险识别度和处理速度大幅提高，2023 年网贷业务月均不良新增额

下降超过 50%，年化不良下迁率下降 2.17 个百分点，领先同业水平。这一效果源于大数据模型对潜在风险信号的精准捕捉，如芝麻信用通过分析用户的消费习惯和履约记录，能够提前识别高风险客户。中国人民大学的研究进一步验证了大数据模型的优势。通过对比 GBDT、XGBoost 算法与传统 Logistic 模型在消费金融数据上的表现，发现机器学习模型在拒绝率与违约率的平衡上更具优势，能够在相同拒绝率下降低违约风险，或在相同风险水平下扩大授信规模。

4.2 服务效率与覆盖范围扩大

大数据技术大幅提升了信用评估的效率，实现了信贷业务的敏捷迭代。恒丰银行通过优化数字化信贷平台，实现了字节、借呗等贷款产品的整装式上线和快速迭代，打造了“自主获客”“智能风控”“数字运营”三大数字化核心能力。信用评估时间从传统的数天缩短至分钟级，显著改善了用户体验^[3]。大数据信用评估扩大了金融服务的覆盖范围，推动了普惠金融发展。传统征信体系难以覆盖的蓝领、自由职业者等群体，通过其在互联网平台的行为数据获得了信用评分，从而获得合理的信贷服务。芝麻信用分 350-950 的分值范围，为不同信用水平的用户提供了差异化服务可能。恒丰银行的花呗联合贷款产品通过大数据模型实现数百万级新客准入，体现了大数据技术在普惠金融中的价值。

4.3 业务规模与经济效益增长

在风险可控的前提下，大数据信用评估支持了互联网金融业务的规模扩张。恒丰银行 2023 年网贷业务规模较上年增长超 10%，在规模增长的同时实现了风险指标的多维度下降，验证了大数据模型在平衡增长与风险中的作用。这种良性循环源于大数据模型对优质客户的精准识别，恒丰银行通过模型筛选出数百万级存量优质客户进行提额，既提升了客户满意度，又增加了业务收入。从行业层面看，大数据信用评估推动了互联网金融信用评级市场的快速发展。头部机构通过技术创新构建竞争优势，市场集中度呈现逐步提高的趋势。这种技术驱动的市场扩张不仅提高了金融资源配置效率，还为实体经济注入了活力，特别是在促进消费升级方面发挥了重要作用。

5 大数据信用评估面临的挑战与对策

5.1 主要挑战

大数据信用评估面临四大核心挑战。数据质量与标准化方面，尽管数据海量，但虚假交易、数据缺失等问题易引发模型偏差，且不同机构数据标准不统一，增加了跨平台合作难度，芝麻信用需通过变量筛选剔除区分度不足的行为数据。隐私保护与合规风险突出，金融数据敏感性强，《个人信息保护法》等法规对数据收集、使用和共享提出严格要求，如何在合规前提下挖掘数据价值成为机构难题。算法存

在偏见且可解释性不足，机器学习模型尤其是深度学习模型的“黑箱”特性，不仅影响用户信任，还可能隐藏历史偏见导致不公平评估，芝麻信用在模型设计中需兼顾多维度变量以增强可解释性。模型鲁棒性与适应性不足，金融市场环境变化及欺诈手段升级易导致模型性能衰减，即便表现优异的模型（如京东金融案例所示），也需应对新型欺诈挑战，提升鲁棒性与适应性成为技术团队的长期任务。

5.2 解决对策

完善数据治理体系，建立覆盖采集、清洗、存储、使用的全流程机制，提升数据质量并推动行业标准建设，如恒丰银行通过与同业、卡组织等生态合作，构建多元数据来源，为模型提供高质量支撑。创新隐私计算技术，采用联邦学习、多方安全计算等，在不泄露原始数据的前提下实现模型共建，京东金融的消费数据 API 方案借助联邦学习，让合作机构在保护隐私的同时联合训练模型，兼顾合规与效果。构建可解释的 AI 模型需平衡预测准确性与可解释性，可结合逻辑回归与机器学习模型，用简单模型解释复杂决策，芝麻信用通过五大维度划分，使用户能理解评分逻辑。建立动态优化机制，通过实时监控和定期评估调整模型性能，恒丰银行的“数据-模型-策略”闭环及芝麻信用的周度复盘机制均为有效方法，同时需加强跨行业合作，共享欺诈案例以提升模型适应性。

6 结语

大数据技术为互联网金融信用评估带来了革命性变革，通过数据维度的扩展、算法模型的优化和风控流程的重构，显著提升了信用评估的准确性、效率和覆盖范围。实证案例表明，采用大数据信用评估模型的金融机构能够在业务规模增长的同时有效控制风险，实现可持续发展。然而，大数据信用评估仍面临数据质量、隐私保护、算法偏见等多重挑战，需要通过技术创新与制度规范相结合的方式加以解决。随着技术的不断进步和监管的日益完善，大数据信用评估将在推动普惠金融发展、防范金融风险方面发挥更大作用，为数字经济时代的金融体系稳定提供坚实支撑。未来研究可进一步探索人工智能、区块链等新技术与信用评估的融合应用，以及全球数据治理框架下的信用评估体系构建。

参考文献

- [1] 杜华林.大数据技术时代互联网金融的风险防控方法[J].财富生活,2025(5):130-132.
- [2] 尹斌.互联网工具与技术在中小企业信用管理中的应用[J].智能城市应用,2024,7(1):107-109.
- [3] 周钰杰.浅谈人工智能算法在互联网金融领域的应用与优化[J].中文科技期刊数据库(全文版)经济管理,2024(5):0047-0050.
- [4] 刘凌子.银行客户行为分析对信用风险评估的精准度提升研究[J].中文科技期刊数据库(全文版)经济管理,2025(1):116-119.