Interpretation and Prospect of Personal Credit Evaluation Based on Big Data

Xingyu Chen[1*]

[1]School of Economics, Xi'an University of Finance and Economics, Xi'an, China

Abstract. In the current booming development of financial technology, in order to improve the quality of personal credit evaluation, enhance platform prediction efficiency, and improve the credit evaluation system, this article deeply explores effective methods for building a credit evaluation service platform. Compared with traditional financial data, these platforms fully leverage the significant characteristics of big data, such as large scale, rapid circulation, and diverse types. Through a comprehensive investigation of the current situation of personal credit evaluation in China and in-depth analysis of the key role played by big data in credit evaluation, important achievements have been made in the research. It not only lays a solid theoretical foundation for building a reliable and accurate AI driven evaluation system, but also provides valuable decision-making references for policy makers, helping them make scientific decisions in balancing technological innovation and data ethics, innovating data collection and usage methods, promoting cross platform data sharing, and effectively addressing opportunities and challenges in the field of credit evaluation.

1 Introduction

With the advent of the Internet and the advent of the digital age, the birth of big data has brought significant changes, and the multi-dimensional acquisition and computation of personal behavioral data is leading to a revolutionary shift in integrity from static wealth control to dynamic behavioral prediction. In terms of personal credit scores, financial institutions have turned to a new type of digital credit reporting based on traditional credit reporting models such as real estate, payroll, bank statements, etc., where big data can predict personal credit through everyday behavior, such as WeChat Payment Points, Payment Sesame Credit.

The core of this change is the logical transformation of using alternative data to construct credit scores, from repayment proof to repayment preparation prediction. The construction of a personal credit evaluation system is deeply influenced by regional institutional and cultural differences. In developed countries, traditional models represented by the FICO score in the United States rely on strong data such as social security numbers and tax records, and their institutional foundation is a highly mature citizen information management system [1]. Although the role of big data has been widely recognized in academia, there are still

* Corresponding author: 15319170733@163.com

obstacles to its application in improving prediction accuracy: the focus on algorithm optimization in technology has overlooked the compliance restrictions of some data under legal protection, such as the Personal Information Protection Law [2-4]. Secondly, there is an emphasis on international financial cooperation and information sharing in the institutional framework to promote the development of the personal credit market, but the opposition between data and law in developing countries has not been resolved, and the failure of the AHP credit evaluation model in China's P2P online lending market has not been addressed [5]. Thirdly, some experts subjectively emphasize the risks of financial institutions when conducting surveys, resulting in a lack of integration and analysis of the three-dimensional framework of technology, institutions, and society [6]. For example, the decline in data quality of credit cards has led to errors in their actual application [7]. This article breaks through the limitations of a single discipline and systematically reveals the importance of digital transformation in personal credit evaluation. This is not a simple technical suggestion. In the second section, this article systematically reviews the current application status of big data in personal credit evaluation, reveals its data sources, and proposes its core contradictions.

2 Literature review

2.1 Big data sources in personal credit evaluation

Alternative data refers to data information that is not limited to traditional financial consumer credit data and has the nature of credit supplementation. It can collect diverse information from various non-traditional sources, such as social media activities, mobile device usage, consumer behavior, etc. Enrich traditional credit reporting methods by obtaining data on users' daily behavior.

In the modern consumer evaluation system, alternative data plays an important role. Alipay's Sesame Credit not only considers users' financial transaction records but also analyzes users' daily consumption records to establish credit ratings. This approach can more accurately reflect the integrity of users and provide reliable assistance to financial institutions [5]. WeChat's payment system uses social data to evaluate credit levels, but there are strict restrictions on the use of personal information within the current legal framework. How to ensure that data collection is not restricted by law has also sparked deep thinking [8]. The application of IoT data, such as Kenya's mobile lending platform M-Shwari, infers users' stability and lifestyle habits based on information such as phone usage frequency. By analyzing individuals' daily lives, credit scores can be improved. Due to differences in economic development and infrastructure construction in different regions, adjustments and optimizations need to be made based on the actual situation of each region in practical use [9]. Social media data in credit scoring programs can be used for credit assessment, providing valuable insights into users' economic and professional profiles by utilizing four data features from social media, including demographics, personality, psycholinguistics, and social networks. For example, demographic information such as education level, salary, occupation, and employment history reflects financial stability and income potential, both of which are key factors in credit scoring. In addition, adopting social media data focuses on minimizing the risk of misleading or overly arbitrary data input, which is different from other platforms where social participation may not be directly related to financial behavior. Demographic data helps us understand the characteristics of target borrowers, providing insights into personality and psycholinguistics through user text content, as well as analyzing personality traits to reveal individuals' attitudes and motivations. In addition, psycholinguistic analysis can provide a deeper understanding of language patterns and emotional expressions in online

communication, and the nature and quality of network connections can help to gain a deeper understanding of their social networks [10]. After comparison, it was found that different alternative data have significant differences in their value in credit evaluation. Consumer data can directly display whether the agreement can be followed as scheduled and has the best effect in predicting repayment willingness. Mobile data has advantages under inclusive finance conditions and is more sensitive to income stability prediction. Social data may have biases due to cultural differences, and its accuracy improvement is limited and volatile.

2.2 Unstructured data: providing new dimensions for credit evaluation

With the development of information technology, the processing technology of unstructured data is gradually changing the way credit evaluation is conducted, providing more comprehensive and accurate tools.

In credit evaluation, there are two types of unstructured data. One is text data, which in the financial industry automatically extracts key information from contracts through text data annotation and natural language processing technology, reducing manual processing workload and improving work efficiency. Another aspect is that images and behavior logs can be used to record users' online behavior, improving service efficiency and providing customers with more personalized and convenient service experience. The application of textual data mainly provides key information related to personal credit by analyzing users' communication records, social media, and other textual information. If negative keywords frequently appear in a user's daily records, it may indicate doubts about the user's trustworthiness [7]. In addition to keyword recognition, sentiment analysis is also one of the commonly used technologies. By analyzing users' statements and comments on the Internet, can speculate whether they are trustworthy, which helps financial institutions better understand customer behavior and potential risks. As another important unstructured data source, images and behavior logs are also gradually applied to the field of credit evaluation. Micro expression analysis is carried out on users to capture their real reactions in different situations to identify the uncertainty of fraud [8]. By tracking the traces left by users on the Internet, users' interests and habits are inferred. This information helps to build a more detailed model, thus improving the accuracy of credit evaluation. Although unstructured data has opened new perspectives and methods for credit evaluation, it still needs to overcome the challenges of privacy protection and immature technology in practical applications.

2.3 Structured data: balancing tradition and innovation

Structured data plays an important role in credit risk assessment. Data that strictly follows data models, is easy to search and organize, and is usually stored in relational databases. However, traditional structured data is still the basis for banks and other financial institutions to conduct credit assessments. Traditional structured data provides direct and reliable credit information, which is an important basis for evaluating whether the borrower can repay and whether they are willing to repay. With the development of technology, the connotation of structured data is constantly expanding. By using logistic regression models to analyze the continuity and quantity of personal provident fund contributions, the stability of their work and economic sources can be evaluated [4]. Furthermore, personal tax payment situations can be used to determine whether the borrower's financial situation meets the standards. However, in developing countries, the imperfect data sharing mechanism limits the application scope of structured data, which affects the accuracy of credit evaluation. In the future, optimizing data collection and processes to improve data accuracy can also be used to find the best balance between traditional structured data and innovative points, promoting the continuous progress of personal credit evaluation.

3 Current status and challenges of existing research

In the field of credit evaluation, with the development of big data technology, traditional statistical models and new machine algorithms are constantly evolving, but each method has its unique comfort zone and challenges.

To address the challenges of big data, traditional statistical models have improved their compatibility through technological improvements. In the field of credit evaluation, Lasso regression is used to process high-level structured data, and a concise and explanatory credit scoring model is constructed. The most critical risk factors are explored through feature selection, ultimately forming a concise and explanatory model. This model can not only improve the accuracy of risk prediction but also facilitate business personnel to understand and interpret model decisions, which is beneficial for financial institutions to formulate more accurate risk control strategies. However, it may not effectively integrate unstructured features. The dynamic weight adjustment of the credit card scoring model attempts to include macroeconomic indicators (such as unemployment rate) [4], to improve adaptability to economic fluctuations. This adjustment mechanism helps to more accurately reflect changes in external economic conditions, thereby enhancing the predictive accuracy of the model. However, compared to machine learning models, this method still lacks adaptability and is less flexible. However, facing the problem of exploding dimensions, manually selecting and constructing features is not only time-consuming and laborious, but also prone to missing important information. High-angle data is subject to LASSO regulatory control, but only supports structured data, and WeChat cannot understand non-traditional features such as payment history.

Supervised learning algorithms, represented by XGBoost, have become mainstream tools in credit evaluation. These algorithms significantly improve the accuracy of repayment intention prediction by analyzing payment behavior, consumption frequency, and other characteristics. For example, XGBoost can efficiently process large-scale datasets and has good interpretability, making it widely popular in practical applications and providing high stability.

In the field of credit evaluation, traditional methods have improved their adaptability and compatibility to a certain extent through technological improvements but still face challenges such as dimension explosion and big data bottlenecks. At the same time, machine learning technology has gradually become a mainstream tool with its powerful data processing capabilities and prediction accuracy. However, the black box nature of machine learning models and the weight allocation problem of hybrid models remain key challenges that need to be further addressed.

4 Conclusion

For a long time, influenced by factors such as policies and regulations, economic environment, and the credit review system of financial institutions, the expansion of data dimensions has shifted credit evaluation from static asset verification to dynamic behavior prediction. Machine learning technology has become the core tool for improving accuracy, but interpretability and generalization ability are still bottlenecks. To promote the development of the field of personal evaluation, need to start from several aspects: focusing on researching data fusion, algorithm optimization, and interpretability exploration; Drawing on advanced research methods and technologies from abroad, aim to improve the research level of ethical issues such as privacy protection and algorithm fairness in China; Adapt to local conditions, strengthen model research on credit evaluation markets in developing countries, and research should not overly rely on idealized data.

In the field of credit evaluation, data governance is carried out, and an alternative data system is constructed based on laws and regulations to define the scope of different risk levels. An evaluation framework is constructed to include sudden economic variables and reduce the risk of estimation and prediction.

References

1. X. Li, Y. Zhong, An overview of personal credit scoring: Techniques and future work, Int. J. Intell. Sci. **2**, 181-189 (2012)
2. F. Wu, X. Su, Y.S. Ock, Z. Wang, Personal credit risk evaluation model of P2P online lending based on AHP, Symmetry. **13**, 83 (2021)
3. Z. Ma, W. Hou, D. Zhang, A credit risk assessment model of borrowers in P2P lending based on BP neural network, (2021)
4. S. Li, H. Lai, Z. Zongfang, A review of research on personal credit risk assessment, Proc. Third Symp. Risk Anal. Risk Manag. West. China. 149-154 (2013)
5. People's Bank of China Inclusive Finance Working Group, China inclusive finance report, (2023)
6. A. Devereaux, L. Peng, Give us a little social credit: to design or to discover personal ratings in the era of Big Data, J. Int. Econ. **16**, 369-387 (2020)
7. W. Tian, S. Li, Research on personal credit risk assessment based on data fusion, Acad. J. Bus. Manag. **5**, 62-74 (2023)
8. J. Jiang, L. Liao, X. Lu, Z. Wang, H. Xiang, Deciphering big data in consumer credit evaluation, J. Empir. Financ. **62**, 28-45 (2021)
9. G.T. Chi, W. Xu, X.F. Sun, Research on personal credit card risk evaluation system and model, J. Tongji Univ. (Nat. Sci.) **4,** 557-563 (2006)
10. A. Alamsyah, A.A. Hafidh, A.D. Mulya, Innovative credit risk assessment: Leveraging social media data for inclusive credit scoring in Indonesia’s Fintech sector, J. Risk Financial Manag. **18**, 74 (2025)