The Impact of Social Network Data on Personal Credit Risk Assessment

Qianzhenghua Gao[1*]

[1] School of Automation and Software, Shanxi University, Taiyuan, China

Abstract. In the digital age, personal credit risk assessment has become a hot topic of discussion and research in the field of technology finance. With the popularity of social networks and the development of big data technology, personal credit risk assessment models using alternative data have emerged as a solution to the problem of personal credit risk assessment, with social network data providing an important supplement to personal credit risk assessment. This article aims to explore the application and impact of social network data in personal credit risk assessment by integrating, summarizing, and analyzing relevant literature. Through research, it has been found that social network data can be used to analyze information such as borrowers' consumption habits and social behavior, thereby helping lending institutions to more comprehensively and effectively assess personal credit risk. This article also raises two issues in the application of social network data in personal credit risk assessment, aiming to provide suggestions for financial institutions on using social network data to promote the healthy development of credit finance.

1 Introduction

Personal credit loans have always been one of the core businesses of many small and medium-sized enterprises and banks in the financial field, and risk assessment of personal credit is the core process of the personal credit business. Whether the loan risk can be accurately assessed is a key issue in this industry. Traditional personal credit evaluation mainly relies on structured data such as personal financial data and personal credit history. These structured data and a single data source are difficult to cover the multi-dimensional characteristics and dynamic behaviors of lenders, resulting in a single and one-sided evaluation dimension. Cheng Liu et al. pointed out the limitations of the traditional single model in personal credit scoring by establishing the SVM-Logistic composite model, especially in the face of big data such as Internet data and social network data [1]. For a long time, lending companies have been hoping for more accurate evaluation methods. With the development of social sciences and finance, as well as the arrival of the big data era, social network data rich in behavioral characteristics and social relationships, such as personal digital footprints, and sentiment analysis data based on them, have gradually entered the field of social credit risk assessment and been incorporated into the risk assessment system.

* Corresponding author: gaoqianzhenghua@sxu.edu.cn

Most existing research focuses on designing different types of social network data as variables in new risk assessment models. Yanyong Wang et al. studied the social network data generated by Weibo users when using Weibo (including demographic attributes, user relationship structure, and tweet content), and used support vector machines (SVM), naive Bayes (NB), logistic regression (LR), and AdaBoost classification algorithms [2]. Modelling based on practical problems, analyzing social network data of personal credit, and demonstrating the feasibility and advantages of social network data in personal credit evaluation. This article aims to sort out the application of social network data in personal credit risk assessment through the organization and summarization of existing literature. In recent years, there have been frequent data breaches caused by hacker attacks or system vulnerabilities. The number of publicly reported data breaches worldwide in 2024 has increased significantly compared to 2023. In this situation, can social network data be effectively utilized while protecting user data privacy? How to evaluate whether the social network data provided by the lender is reliable? Studying such issues is of great significance for the future development of credit risk assessment.

2 The application of alternative data in the financial field

Alternative data, also known as non-traditional data, is non-traditional data that differs from traditional data such as financial statements and production data but can provide valuable information for financial institutions. In the financial field, traditional data types have disadvantages such as low processing efficiency and slow response speed. Alternative data has three major advantages when applied: firstly, data diversification. Alternative data includes many data that cannot be collected and analyzed by traditional data. For example, the application of satellite signal data in predicting agricultural activities can be achieved through remote sensing technology to predict the yield and income of crop production areas; Payment information data can reflect the borrower's consumption ability and habits. Secondly, the data flow rate is high. Traditional data types may experience lag in data and information due to delayed disclosure times, while alternative data is almost real-time. For example, in online shopping, after payment is completed, the backend can receive transaction information in real time. Thirdly, it has a large volume, and the scale and transmission volume of alternative data are both significant, constantly accumulating at an astonishing speed [3].

With the advent of the digital age, the issue of personal credit risk assessment in the financial field has become a hot research topic. The author believes that social network data, as a type of alternative data, can effectively reflect the advantages of alternative data in the application of personal credit risk assessment. At the same time, according to the data of China Internet Network Information Center (CNNIC), as of March 2024, the number of Internet users in China has reached 1.092 billion, and the mutual penetration rate of the Internet in China has reached 77.5%. The global internet user base has reached billions, accounting for 65.4%. Such a large number of netizens will generate massive amounts of data, including typical social network data such as social media interactions, online activity, and comments. These data provide a foundation for financial institutions to conduct a multidimensional analysis of lenders. For example, Huangjing et al. conducted research on online lending based on the XGBoost algorithm, using over 680000 pieces of data from 26000 users on the Rong360 online lending platform. They used the XGBoost algorithm to establish a predictive model for users' "whether to borrow again". Research has found that in online lending, the "hard information" (personal basic information) provided by users no longer has a signal revealing effect on predicting whether to borrow again, while the "behavior trajectory" information and "social interaction" information of users' borrowing and consumption on the platform have a more revealing effect on trust signals [4].

3 Application of social network data in personal credit risk assessment

Social network data, as a type of alternative data, refers to the data generated by users engaging in transactions, making comments, and other interactions through social media platforms. This data provides valuable information for financial institutions beyond traditional data and credit records. The traditional personal credit risk assessment system and credit scoring system overlook some aspects of the borrower's credit situation that cannot be reflected by traditional data, such as the borrower's active participation in social media interactions and positive comments. Especially for "credit white accounts" who have not undergone credit risk assessment before taking out a loan, the use of social network data can evaluate the borrower's credit status from aspects such as the borrower's relationship intimacy, friend credit level, community participation, and personal digital footprint.

3.1 Application of big data technology in personal credit risk assessment

Social network data has the characteristic of large volume and belongs to a type of big data. When processing this type of data, the support of big data technology is required, and processing large amounts of data is a prerequisite for applying social network data. Big data technology provides the ability to process and analyze massive amounts of data. Having big data processing technology can make it more convenient and accurate for financial institutions to evaluate lenders from multiple dimensions, which is crucial for credit risk assessment. At present, many scholars have confirmed the feasibility of applying big data technology in personal credit risk assessment. For example, Mar í a Ó skards ó ttir et al. used neural networks, random forests, and other algorithms to design a credit scoring system using mobile phone data and social network data. The experimental results confirmed the inclusiveness of mobile phone data and social network data in finance, demonstrating the feasibility of applying big data technology in personal credit risk assessment [5].

3.2 Application of digital footprint data in risk assessment

Digital footprint is a product of the digital age. With the popularity of social media platforms and the steady increase in the number of netizens, online users have generated many traceable digital traces when using digital products, including device information, user IP, etc. Such traces are called digital footprints. Digital footprint data is closely related to Such traces are called digital footprints. social network data. By analyzing the digital footprint of borrowers, one can grasp their online behavior, payment behavior, and so on. Orlova Ekaterina V designed a borrower clustering model based on hierarchical clustering method and k-means method, and proposed a new model that comprehensively considers three core indicators: borrower online behavior, financial behavior, and digital footprint data. The model combines machine learning methods to evaluate borrower credit, thereby improving the accuracy and predictive ability of credit evaluation [6]. The accuracy of the model increases with the increase of sample size. This study also demonstrates the role of social network data in improving the accuracy of personal credit risk assessment. Through social network data, financial institutions can supplement lenders' information and make better judgments.

3.3 Analysis of sesame credit application

In the research report by Xie Li et al., the ideas of Sesame Credit in utilizing borrower social network data were summarized [7]. Unlike the traditional credit reporting approach, the data used by Sesame Credit not only includes basic data such as gender, age, and family status, but also multidimensional data such as credit card account usage duration, consumption level, network credit rating, and rental information, in order to more accurately evaluate borrowers. Sesame Credit has been widely used in the lives of young people in China, and its scoring mechanism and effectiveness have received unanimous praise from lenders.

4 Challenges of social network data in personal credit risk assessment

When using users' social network data for credit risk assessment, the quality and reliability of the data are also factors that financial institutions need to evaluate and confirm. If users conceal problem data, set it as non-public, or even forge personal media data. These issues will all affect the quality of data, thereby reducing the accuracy of personal credit risk assessment. Yanhao Wei et al. compared models using and not using network data and found that models using network data had higher accuracy in credit rating lenders. The study also analyzed the impact of users strategically packaging and building their social networks on rating accuracy [8]. Chen Meiqi et al. pointed out the limitations of sentiment analysis of textual information in understanding bank risk through text analysis of bank risk attitudes, especially in dealing with the differences and complexity of bank annual reports, analyst reports, and credit rating reports [9]. This indicates that although social network data is valuable, further processing and validation of its quality and relevance are needed when applied to personal credit risk assessment.

The privacy issue of data is also a current contradiction. To evaluate the credit situation of borrowers more accurately and effectively, more personal information is needed, and it is difficult to avoid personal privacy issues from this information. Financial institutions may face legal risks when obtaining personal information from borrowers. For example, the "Jirong Borrowing" app under Jiayin Technology was notified by the National Cybersecurity Center for not listing the purpose, method, and scope of collecting and using personal information in its privacy policy, and for not providing users with a way to withdraw their consent. At the same time, opaque information collection, information security risks, and improper use of user information by institutions can also lead to a crisis of trust in financial institutions by users. At the same time, Miguel malheiros and others simulated the new credit card application form in their research using the method of double-blind Research (the experimenters who processed the participants were told to be identical), including sensitive questions such as "friends' names and contact information", and found that the response rate of such sensitive questions was only 69%. Finally, the conclusion is drawn it is acceptable to collect data items that are not traditionally considered relevant if the lender can reliably establish the correlation and assure the applicant that they will receive a fair assessment [10]. This study provides a legal way to obtain the lender's private data, but it needs to establish sufficient trust between financial institutions and lenders. In short, the privacy issue of personal data of borrowers is also one of the problems and challenges that need to be addressed at present.

5 Conclusion

After analyzing relevant literature, this article systematically summarizes the application and impact of social network data in personal credit risk assessment. Research has shown that current models that include social network data are superior to traditional credit risk assessment models both in theory and in practical applications. Social network data has become an important supplementary information required by financial institutions for loan risk assessment of borrowers, especially new borrowers without credit records. At the same time, the introduction of social network data has provided financial institutions with more risk assessment perspectives. However, this process is also accompanied by new challenges such as data privacy protection, data quality control, and the continuous advancement of data analysis technology. Looking ahead, with the continuous development of big data analysis technology and the improvement of personal data privacy regulations, alternative data such as social network data will have deeper and wider applications in the financial field. For future research, hope to develop more effective data protection technologies that can efficiently utilize social network data while protecting privacy.

References

1. Y. Liu, Development status and future trends of alternative data in financial field. China High-Tech. **19**, 104-105+130 (2021)
2. C. Liu, D. Wang, W. Wang, Z. Ji, Personal credit evaluation under the big data and internet background based on group character. Proc. 2019 Int. Conf. Model. Simul. Big Data Anal. (2020)
3. M. Óskarsdóttir, C. Bravo, C. Sarraute, J. Vanthienen, B. Baesens, The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. Appl. Soft Comput. **74**, 26-39 (2020)
4. E.V. Orlova, Methodology and models for individuals' creditworthiness management using digital footprint data and machine learning methods. Mathematics. **15**, (2021)
5. Y. Wei, P. Yildirim, C. Van den Bulte, C. Dellarocas, Credit scoring with social network data. Mark. Sci. **2**, (2015)
6. Y. Wei, P. Yildirim, C. Van den Bulte, C. Dellarocas, Credit scoring with social network data. Mark. Sci. **2**, (2016)
7. M. Chen, J. Ma, L. Wei, Difference analysis in bank risk attitudes: A triple perspectives of bank managers, financial analysts, and credit raters. Procedia Comput. Sci. (2023)
8. A. Onan, S. Korukoğlu, H. Bulut, A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. Expert Syst. Appl. (2016)
9. T. Berg, V. Burg, A. Gombovi, M. Puri, On the rise of FinTechs: Credit scoring using digital footprints. Rev. Financ. Stud. **7**, (2020)
10. Y. Wang et al, Applied analysis of social network data in personal credit evaluation. Lect. Notes Comput. Sci. 10970 (2018)