Exploring Personal Digital Footprints and Credit Risk Assessment

Hongsen Xu[1*]

[1]Hangzhou No.2 High School, Hangzhou, China

Abstract. Traditional credit evaluation for a long time depends on financial statements and credit reports; it does not include people who do not have an existing credit history. Digital footprint is a form of unstructured data. The form includes the social and e-commerce behavior, the transaction with the finance and the geolocation data, which could be analyzed to determine the repayment capacity and willingness. In this paper, I review the types and characteristics of personal digital footprints, and study how they are applied in credit risk models, and explain their effectiveness with behavioral economics and information asymmetry theory. A survey of national studies and cases, followed by national and international studies and cases, is conducted. Evidence from the FDIC and others is cited showing that even small digital footprint variables can produce AUC predictions as good as traditional credit scores. In short, the future developments of standardization technologies on federated learning and online learning systems as well as multimodal data fusion, can be anticipated. The studies in this paper can provide theoretical and practical help to the scientific application of digital footprint on credit risk assessment.

1 Introduction

Credit risk evaluation is the key to the bank in the process of issuing credit loans. The main goal is to understand the debtor's willingness and debt repayment ability. The traditional systems mainly use structured information to form scoring models, such as financial statements, proof of earnings, balance sheets, and credit reports. However, "credit invisible," groups like freelancers, new users, and consumers from emerging markets who have limited or no credit history whatsoever and are hard for us to assess correctly. And it is also late on traditional data update that does not reflect the current situation of the borrower [1]. And there may be misstatements or omissions in the financial statements [2,3].

Now, with the popularity of internet-based technologies and mobile devices, a lot of people's activities are being done online, which generates a great deal of data about their behavioral trajectories called "personal digital footprints". They can be multiple types like social network data, e-commerce transaction information, online search records, geographical location, and app usage. As compared to conventional info, these data are greater in volume, newer in time, and fuller in dimension. such as people's interest, attention and activity, and interpersonal relationship situation on social platform reflects his/her mental

* Corresponding author: xuaq1668816@gmail.com

situation and life stability. E-commerce purchase records can reflect people's consumption, how much money they have, and if they know about money. The use frequency and payment habits, and transfer behavior of financial applications represent the flow of funds and repayment ability [4]. These are alternative information that can not only narrow the credit information gap, but also make up for the weak links of traditional credit information from a behavioral and psychological perspective [5].

Theoreticagap butital footprints make up for the problem of information asymmetry in credit evaluation [6]. Reseaperspective expression in behavioral economics proves that it is important to predict people's future default behavior because it can show people's psychological characteristics, risk preferences and consumption habits [7]. Recently, with the rapid growth of artificial intelligence and machine learning technologies, it has become an impulse to mine and model unstructured datasets, and they have been used in applications such as the P2P platform, digital bank, microfinance [5]. And also, this is supported by the FDIC's empirical evidence: Simple digital footprint variable alone achieved model effects for those with credit score not covered [8]. In 2024 Wang Zhiqin noted that the introduction of SHAP, LIME and other interpretable AI can improve the transparency of models in 2024 Wang Zhiqin noted that the introduction of SHAP, LIME and other interpretable AI can improve the transparency of model [9]. Alternative data and ML techs have major effects on default in dev reg according 2025 SEEJPH study [10].

Though digital footprint does come up short, there are still applications that can be achieved; yet, there are several obstacles that remain: data privacy, data compliance, difficulty integrating data on different platforms, low timeliness due to changes in behavior, poor model generalization abilities, black box model with poor interpretability, potential algorithm bias. Therefore, on this basis, this paper will conduct research from the following aspects. Firstly, sorting out the types and features of digital footprints. Secondly, studying its mechanism and theoretical foundation in intelligent credit evaluation. Thirdly, evaluating and analyzing domestic and foreign use and empirical cases; Finally, summarizing research limitations and proposing application standards and future suggestions. for establishing a scientific, transparent and compliance-digital credit evaluation footprint, it is aimed for this article to be a reference

2 Overview of personal digital footprint

2.1 Definition and types of personal digital footprint

Personal digital footprint means all recorded data information produced after an individual's online activity. It can show the features of such behavior such as personality, interest, hobbies, habits, as well as personal social circle especially into.

First is the social media footprint. It contains a batch of behaviors such as publishing, liking, commenting on social platforms [2]. For example, a person who often posts his or her work achievements on social platform shows that they pay attention to their career and have a can-do attitude; often participating in discussions and joining groups about certain topics shows his interest and possible values [3].

Second, Consumption footprint of E-commerce: Such as shopping record, browsing record, favorite preferences and so on [4]. This paper can study people's consumption ability and habits like what kind of goods they buy, how much it costs, and brands they like from their shopping history [5]. Browsing history is the expression of people's concerns before shopping, as well as their search and inquiry of various commodity information; Collection preference is also the expression of people's focus on certain commodities or certain categories; It can reflect people's consumption plans and capital allocation [6].

Third is the financial transaction footprint, Bank transfer, credit card consumption, investment and financial management data [7]. The bank transfer record is connected to the flow of one's daily money, if can know the bank transfer record, can see if one has a stable and consistent daily fund [8]. Credit card consumption records can be consumed to determine whether there is credit-related behavior, excessive consumption, or timely payment [9]. Investment and financial data is closely related to people's risk tolerance and awareness of wealth planning, what percentage and returns of investment in stocks and funds [10].

2.2 Characteristics of personal digital footprint

In today's digital era, the data volume of personal digital footprint shows the characteristics of a massive amount [1]. Every day, billions of people around the world carry out countless activities on various network platforms, and the data generated by these activities are aggregated into a huge digital footprint data pool [2]. At the same time, data sources are extremely extensive, not only commonly from social media, e-commerce platforms, online banks and other large websites, but also various mobile applications, Internet of Things devices, etc [3]. For example, a smart bracelet can record an individual's movement track, sleep quality, and other data. This information seemingly unrelated to financial credit, may indirectly affect his or her repayment ability and willingness by reflecting his or her lifestyle [4].

Personal digital footprint has the property of real-time, which can reflect people's behavior in real time and the change of their current state [5]. For example, if a person has often searched for small loans on a sudden financial platform, some sudden life change-related information appeared in a social platform, this kind of first time, can be used as a latest reference basis for credit risk evaluation [6]. Dynamics is trying to say that these things and states are always changing, and the evaluation model must catch those changes in time and adjust [7]. The various types of digital footprints have a connection between them and it provide a lot of value for an all-inclusive analysis [8]. Like, consumer posting order behavior on social network is mutually validated with the actual purchase records in the e-commerce consumption footprint, the consumption capacity and preference of individuals can be further confirmed [9]. At the same time, when combining the flow of capital in the financial transaction footprint with the changes in life status on the platform, it's possible to better understand and evaluate the level of individual repayment stability in the face of economic fluctuations [10]. Analyzing all kinds of multi-dimensional digital footprints data can tell us a lot of other data, as well as individuals' possible default, and improve credit risk evaluation accuracy.

3 Correlation analysis between personal digital footprint and credit risk

3.1 Influencing factors of credit risk

Credit risk mainly depends on how able and willing the borrower is to repay debt [1]. Repayment ability encompasses individual income condition, own assets, debt burden and so on, and repayment willingness includes individual credit concept, consumption habit, life planning, and many other factors [2]. Traditional credit evaluation factors are mostly measured through financial statements, income proof, credit report, etc., but these data have certain shortcomings and lags [3].

3.2 Internal connection between personal digital footprint and credit risk

Personal digital footprint provides more rich and real-time information on credit risk [4]. For example, a person's positive attitude towards life, stable working status and good social relations displayed on social media may imply strong repayment intention and relatively stable living environment, which is conducive to reducing credit risks [5]. According to the e-commerce consumption footprint, those with reasonable and planned consumption such as buying daily necessities regularly and only purchasing goods when the price is reasonable through comparison, can be considered as those who have reasonable planning of capital expenditure, which is conducive to one's repayment ability [6].

3.3 Theoretical explanation of correlation mechanism

Behavioral economics thinks that people's behavior is due to many reasons like psychology, society and culture [7]. Personal behavior on the Internet, i.e., digital footprints, are external manifestation under the action of such factors [8]. By analyzing these behavioral data, can get to know people's consumer psychology, risk preference, and other characteristics so as to forecast their behavior and performance in credit behavior [9]. For example, people with high T tend to prefer to consume immediately and have high default risk in terms of credit; people with high self-control will rationally arrange their consumption and compensation, making it less likely to default [10].

In the traditional credit relationship, information asymmetric is an old problem, and it is hard for banks and other financial institutions to fully and accurately understand the real situation of borrowers, leading to problems like adverse selection and moral hazard. Personal digital footprint's showing makes the information asymmetry ease some. By analyzing the digital footprints, the financial institution can obtain information about the borrower, especially non-financial information, which is hard to be reflected by traditional data, and then make more comprehensive judgment on the risk of loan of each borrower, and thus making more reasonable credit decisions to reduce information asymmetry causing loss of credit.

4 Conclusion

This paper gives a thorough review of the research and application of personal digital footprint in the aspect of credit risk assessment, and reveals the advantages it carries concerning enriching the modeling feature, bettering the assessment precision as well as its inclusivity. Digital footprint gives new vitality to the traditional credit scoring system through multi-source information, real-time, dynamic and unstructured information, most importantly, it can better cover up no credit investigation subjects and improve the prediction accuracy. A combination of behavioral economics and information asymmetry theory also gives it a strong theoretical basis. From several empirical research and industry practices, can see that the evaluation models like behavior, consumption and finance footprint are used well in a variety situation e.g. P2P, microfinancing.

For the future, people can do some things like improving data privacy protection and compliance by using federated learning and using differential privacy techniques, also, combining XAI tools such as Shapley value and LIME to improve model explainability which gives transparency and regulatory trust. Third building an online learning system to deal with borrowers' behavior change and improve model adaptability; four to promote multi-modal integration (such as the use of biological behavior, psychological evaluation, and sentiment analysis data), enhancing the identification capability of "invisible default" groups, although currently there are still problems such as data heterogeneity, technical

difficulty, and algorithm fairness, digital footprint applications to credit risk assessment have broad prospects, and it is worth further exploration and promotion under the joint advancement of regulatory improvement and technological development.

References

1. T. Bellotti, J. Crook, Support vector machines for credit scoring and discovery of significant features. Expert Syst. Appl. **36**, 3302-3308 (2009)
2. T. Berg, V. Burg, A. Gombović, M. Puri, On the rise of FinTechs: Credit scoring using digital footprints. Rev. Financ. Stud. **33**, 2845-2897 (2020)
3. X. Han, R. Ding, L. Wang, H. Huang, CreditPrint: Credit investigation via geographic footprints by deep learning. arXiv:1910.08734, (2019)
4. M.D. Hoang, L. Le, A.T. Nguyen, T. Le, H.D. Nguyen, Federated artificial intelligence for unified credit assessment. arXiv:2105.09484, (2021)
5. J. Jagtiani, C. Lemieux, The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform. FRB Philadelphia Work. Pap. 18-15 (2018)
6. A.E. Khandani, A.J. Kim, A.W. Lo, Consumer credit-risk models via machine-learning algorithms. J. Bank. Financ. **34**, 2767-2787 (2010)
7. L. Wang, J. Zhu, C. Zheng, Z. Zhang, Incorporating digital footprints into credit-scoring models through model averaging. Mathematics. **12**, 2907 (2024)
8. FDIC, On the Rise of FinTechs - Credit Scoring using Digital Footprints. Federal Deposit Insurance Corporation. (2025)
9. Z.Q. Wang, Artificial Intelligence and Machine Learning in Credit Risk Assessment. Springer. (2024)
10. SEEJPH, Enhancing Credit Scoring with Alternative Data. Society for Economic Education in Japan and the Pacific Region. (2025)