# Research on Credit Empowerment Based on Big Data and Digital Footprints

Xinyue Zhang *

College of Business, Harbin University of Commerce, Harbin, China

* Corresponding Author Email: 2511184459@qq.com

**Abstract.** Under the background of the deep development of the digital economy, the traditional credit evaluation relies too much on asset mortgage and financial reports, which leads to information lag and makes it difficult to cover the "white credit" group. This limitation promotes the transformation of credit evaluation to "predictive insight" driven by big data and digital footprint. This paper systematically discusses this enabling mechanism, covering the technological evolution path from data fusion, feature engineering, to deep learning, graph neural network, and federated learning, and analyzes its wide application in financial credit, social mortgage-free leasing, and metacosmic decentralized identity. The study found that big data empowerment significantly improved the fidelity of risk identification. By giving financial value to the digital trajectory, it eliminated the mortgage bias of traditional credit and created opportunities for inclusive participation for vulnerable groups. At the same time, credit scoring is evolving from a single economic constraint to an "algorithmic contract" that promotes social transformation and digital survival. Despite the challenges of algorithmic discrimination, model black box and privacy paradox, this paper points out that the future credit system should be transformed to a "trusted AI" framework, combining regulatory Technology and privacy computing technology to achieve deep decoding and governance optimization of credit value on the premise of ensuring data sovereignty.

**Keywords:** Big Data; Digital Footprints; Credit Empowerment; Financial Inclusion; Algorithmic Governance.

## 1. Introduction

In the context of the full development of the digital economy, cutting-edge technologies are profoundly changing the traditional pattern of credit evaluation. For a long time, credit evaluation has been highly dependent on asset mortgages and financial reports. Due to the lagging access to information and limited coverage, a large number of "white credit households" lacking financial history have been blocked out of the formal financial threshold. This limitation has given birth to a strange phenomenon. Many people lack credit certificates in the physical world, but they have left rich behavioral marks in the digital world. With the application of big data and digital footprint technology, credit evaluation has undergone a paradigm change. By transforming real-time and multidimensional behavior data into quantitative features, it has achieved a leap from "data replication in the past" to "future risk prediction".

A series of empirical studies in academia have confirmed the core value of technology empowerment. Tobias Berg and other scholars have found that the prediction ability of default risk has exceeded the score of traditional credit reporting institutions by analyzing the basic digital footprints left by e-commerce platforms, such as equipment types, e-mail service providers and login time. This finding shows that micro behavioral characteristics can effectively reflect the potential characteristics of consumers and show great potential in filling the credit gap [1].

Liu et al. conducted a study that utilized call detail records (CDR) and billing data to assess approximately 300 million "credit-invisible" individuals in China who lacked traditional credit records. They constructed a credit assessment model by analyzing communication behavior indicators such as call frequency, duration, and contact diversity. This model effectively predicted individual default risks and empirically verified the significant correlation between mobile communication data and loan credit (including repayment ability and willingness) [2].

At the same time, the dimension of credit evaluation is also expanding in breadth and depth. María Óskarsdóttir and others used social network analysis technology to prove that the credit status of individuals is highly correlated with the performance of their social circle. This method extends risk identification from isolated individual behaviors to complex association networks and significantly enhances the accuracy of prediction by capturing risk contagion signals [3]. In 2025, the technology trend is moving towards deep semantic understanding. The large language model (LLM) has been used to analyze unstructured text. Through the chain thinking logic, the social interaction or contract terms are transformed into an interpretable scoring basis, which solves the limitations of traditional technology in sensing repayment willingness.

Based on the above background, this paper aims to systematically explore the credit empowerment mechanism driven by big data and digital footprint. By constructing a technical system covering data fusion, feature extraction, model construction and decision-making application, this paper deeply analyzes the logic of transforming digital trajectory into credit features. This paper focuses on the practical application of the system in the fields of finance, social management and metauniverse, and evaluates its effectiveness in improving efficiency, accuracy and social inclusiveness. Finally, this paper makes a critical discussion on the current challenges and looks forward to the future development path.

## 2. Technical Method System

In the context of the digital economy, big data has gone beyond simple information stacking and evolved into a deep slice with multidimensional interaction, real-time update and unstructured features, which can accurately capture the credit default signals of micro individuals. As a microcosmic expression of big data, digital footprint is regarded as the "core fuel" of the modern credit system. These footprints can be divided into two categories: one is the active footprints of users in the scenarios of e-commerce consumption and social messaging; Second, the system automatically generates passive footprints such as geographical tracks and equipment metadata. These continuous series data make the credit evaluation from a single financial audit to a comprehensive personality measurement.

The implementation path of credit empowerment is similar to a set of sophisticated data alchemy, and its technical architecture follows the logical evolution from bottom-level perception to high-level decision-making.

From the original track to the credit signal in the data layer, the core of research is the fusion and governance of cross-domain data. Integrating heterogeneous data from e-commerce, social networking, government affairs and the Internet of Things through ETL Technology, effectively breaking the data island. Subsequently, at the feature level, researchers carried out key feature engineering, extracted key variables such as residential stability, social group reputation and consumer prudence from the massive footprint, and transformed the fragmented trajectory into an interpretable credit signal.

From traditional algorithms to intelligent evolution, the technology path has experienced a leap from classic to frontier in the selection of models. Represented by the research of Khandani et al., traditional machine learning, such as logistic regression, significantly improves the accuracy of default prediction by integrating consumption flow and transaction frequency. Although the traditional model has excellent explanatory ability, it still lags behind in dealing with nonlinear sudden risk. With the increasing complexity of associated risks, the deep learning model began to shine brightly [4]. The graph neural network (GNN) framework proposed by Yang et al. Has greatly improved the recognition rate of associated default by deeply depicting the digital correlation between the supply chain and the Guarantee Circle [5].

Privacy protection and semantic insight for the problem of data privacy and islands, federated learning technology provides a new path of collaborative modeling of "data is not separated from the domain". The federal forest architecture proposed by Liu et al. Has achieved high modeling accuracy

under the premise of ensuring compliance and solved the bottleneck of cross-agency risk control [6]. In addition, credit empowerment has been extended to a sense of human subjective will. Zhou et al. Used natural language processing (NLP) to mine the emotional color in lending texts, revealing the deep connection between psychological traits and willingness to default [7].

Finally, the model output is transformed into an accurate decision-making tool. This not only includes the standardized score card used to quantify the risk but also derives the risk pricing model of travel alienation, which realizes the optimal allocation of financial resources. At the same time, the anti-fraud system, combined with real-time behavior trajectory, can quickly identify identity theft, ensuring the security and robustness of the credit empowerment process. Table 1 compares the technical differences between traditional assessment and big data driven.

**Table 1.** Technical path comparison table.

| Dimension | Traditional Credit Assessment System | Big Data & Digital Footprint System | Empowerment Manifestation |
|---|---|---|---|
| Data Sources | Financial statements, bank transactions, collateral information | Consumption footprints, social trajectories, geographic location, and device information | Eliminating access barriers for "credit invisible" individuals |
| Data Characteristics | Structured, low-frequency, static historical data | Unstructured, high-frequency, real-time dynamic data streams | Achieving millisecond-level early warning of credit risks |
| Core Algorithms | Logistic regression, linear scoring cards | Deep learning, graph neural networks, federated learning | Capturing nonlinear complex default correlations |
| Assessment Logic | Retrospective based on historical repayment facts | Predictive based on behavioral patterns and social capital | Cross-scenario application of credit as a social asset |

## 3. Application Scenarios

### 3.1. Financial Services

The financial sector is the most intuitive and mature testing ground for big data credit empowerment. In credit decision-making, by integrating multi-dimensional digital footprints, financial institutions can provide accurate credit portraits for small and micro enterprises that lack collateral, so as to achieve effective coverage of the sinking market. Anti-fraud and risk control use the relevance of social networks and device metadata to identify potential risks such as gang fraud or identity fraud in real time. More revolutionary, this empowerment has realized dynamic quota management and differentiated pricing: credit is no longer a fixed label, but a dynamic indicator of fluctuations based on individual real-time behavior data. For example, the system can adjust the credit line in real time according to the high-frequency fluctuation of the merchant flow, so as to achieve the millisecond matching between the capital supply and the business demand.

In the core source of financial empowerment, the research of Björkegren and its collaborators provides a new perspective for the financing problems of small and micro enterprises in emerging markets. The core breakthrough of this research is to completely break away from the dependence on traditional financial statements and turn to mining 'behavior metadata' in borrowers' smartphones. The researchers creatively extracted multi-dimensional alternative indicators, including the diversity of call contacts, the entropy of geographical movement trajectory (used to measure the regularity of activity space) and the stability of call recharge behavior, and used these data as proxy variables to measure the robustness of individual businesses. The empirical analysis shows that this evaluation model based on digital footprint shows significant prediction efficiency: after the introduction of metadata features, the prediction accuracy (AUC value) of the model for default behavior has increased to 0.71-0.77 compared with the benchmark credit model (AUC 0.51-0.57), an absolute

increase of about 0.14-0.20, and a relative increase of 24-51%. This case profoundly proves the great potential of digital footprint in alleviating information asymmetry [8]. Especially in the long tail financial market, which lacks standardized financial data, through the behavior "mirror" of non-financial dimension, financial institutions can restore the borrower's real repayment ability at a very low cost, so as to achieve more accurate credit empowerment.

### 3.2. Social Management

With the deep popularization of digital footprint, the scope of credit empowerment is extending from a single financial field to a broad sense of social governance, successfully realizing the cross-border transformation from financial credit to social capital. In the scenario of the sharing economy, the credit scoring system based on digital footprint has spawned a large-scale "mortgage-free lease" model. This innovative measure not only significantly reduces the friction cost of market transactions but also reshapes the mutual trust mechanism between strangers to a large extent. At the same time, the government service sector is also actively exploring credit applications. The "Xinyi+" series of services launched by governments at all levels has gradually institutionalized the incentive mechanism for keeping promises by linking personal credit rating with the allocation of public resources, such as administrative approval and medical green channels. In addition, in the field of credit investigation and business cooperation, the credit back adjustment system and supply chain financial model based on behavioral big data have become the core tools for enterprises to identify cooperation risks and optimize the efficiency of industrial chain cooperation.

Sesame Credit, as an independent third-party credit reporting agency under Ant Group, has built a credit evaluation system covering financial and social governance by analyzing users' digital footprints. Its core innovation is to export the financial credit evaluation ability to the shared economy scenario, and cooperate with ofo, harrow bicycle and other companies to launch the "pledge free ride" service. By 2018, 48million users had been exempted from the deposit of 9.8 billion yuan, significantly reducing the market friction cost. At the same time, this model extends to the field of government services, forming a series of "Xinyi+" applications and realizing the cross-border transformation of credit empowerment [9].

### 3.3. Emerging Fields

With the continuous progress of technology, the application of credit empowerment is expanding to more forward-looking frontier fields. In the ecosystem of metauniverse and Web3.0, credit evaluation based on decentralized identity (did) has become the cornerstone of constructing a virtual social contract. Through in-depth analysis of the interactive behavior footprint on the chain, researchers can establish an efficient credit system, which can directly improve the speed and efficiency of digital asset circulation. At the same time, against the background of advocating global sustainable development, the integration of carbon footprint and green credit has become an important trend. The trajectory of individual low-carbon life is translated into quantifiable green credit scores, which are linked to green financial incentives or government policy awards and subsidies. These changes show that credit empowerment is evolving from a single economic management tool to a core rule that drives the green transformation of society and supports digital survival.

In the process of exploring the digital civilization, the research innovatively proposed a credit evaluation mechanism based on "soul binding token" (SBT). The research team has built an unforgivable dynamic scoring framework by systematically analyzing individual governance participation, asset holding history and the interactive footprint of decentralized Finance (DFI) in the metauniverse. This case fully proves that in a decentralized environment, the digital footprint based on blockchain has natural transparency and traceability. This feature effectively solves the pain point that it is difficult to establish a reputation in anonymous scenes, and provides a solid and unforgeable technology base for new credit scenes, such as metauniverse and carbon footprint tracking [10].

## 4. Enabling Effects and Advantages

The most intuitive performance of big data empowerment is to realize the automatic upgrading of the credit process and significantly reduce the marginal cost of credit evaluation. Compared with the limitations of relying on manual investigation and taking several days or even weeks in the traditional mode, the model based on digital footprint can provide second level decision response. This improvement in efficiency not only reduces the operating costs of financial institutions, but also accurately meets the practical needs of small and Micro customers for "fast, frequent and urgent" capital flow, and makes credit evolve from static access vouchers to efficient liquidity resources.

The core technical advantage of credit empowerment is to improve the fidelity of risk identification by using a high-dimensional data matrix. The big data system introduces alternative features such as geographical stability, consumption preference, social influence and psychological characteristics, which make up for the evaluation errors caused by single variables and missing data in the traditional model. Through multi-faceted footprint cross-validation, the system can accurately outline the borrower's real willingness and ability to repay, even when part of the information is damaged. This not only reduces the bad debt rate but also effectively solves the problem of "wrong killing" of high-quality customers caused by information asymmetry.

As one of the most valuable aspects of digital finance, credit empowerment has initiated a profound change in the "availability" of credit by injecting financial attributes into the digital footprint. "White credit households" and start-ups, which were originally excluded from the banking system, were able to obtain financing support by virtue of the business traces and reputation accumulation deposited in the digital space. This model gradually dispels the dependence bias of traditional credit on collateral, provides flowing capital leverage for the bottom of society, and ensures that the digitally vulnerable groups can equally share the dividends of digital civilization.

From the perspective of long-term evolution, credit empowerment has realized the transformation from simple "risk interception" to active "behavior guidance" through the positive incentive mechanism. When credit scores are widely infiltrated into multiple life scenes such as mortgage-free leasing and public services, individuals can intuitively feel the premium dividends brought by credit. This visible convenience forms a positive psychological hint and behavioral constraint, which prompts individuals to be more prudent and compliant in the digital space, and then promotes the transformation of social trust from the traditional blood and geographical relationship to the paradigm of "algorithm contract" based on big data.

## 5. Challenges and Development

### 5.1. Critical Challenges and Risk Critique

While eliminating the exclusion of traditional finance, big data credit assessment may also breed more covert algorithm discrimination. Due to the historical bias of model training data, the algorithm may mark low-income, ethnic minorities, or elderly groups as high-risk, leading to systematic unfairness. This phenomenon has further exacerbated the digital divide, making people who cannot access the digital world completely marginalized in the "credit society". In addition, the society is facing a serious privacy paradox: individuals are forced to give up their privacy in order to obtain financial convenience, resulting in their behavior trajectory being monitored in real time, and finally moving towards a "transparent" society that has lost its freedom of behavior.

At the technical implementation level, data quality has always been a stubborn problem that plagues the accuracy of the assessment. Noise data, false information and speculation such as "swiping orders and speculating letters" seriously interfere with the authenticity of the signal. The core contradiction lies in the model black box problem. Although the prediction accuracy of complex models (such as deep neural networks) is very high, their unexplainability infringes the user's right to know and appeal, and the user cannot obtain a reasonable explanation in the face of credit refusal. In addition, models often have the risk of overfitting. Algorithms that perform well in specific

consumption scenarios face a severe test of their inclusiveness and stability when migrating to cross-industry or macro fluctuation environments.

At present, the lag of regulation has become a global common problem, the speed of technological innovation is far faster than the formulation of regulations, and there is a lack of unified data use standards. From a macro perspective, the large-scale application of the homogenization algorithm may lead to procyclical risk: when the algorithm collectively sends a contraction signal, it may induce a systematic credit crunch, thereby amplifying financial fluctuations.

### 5.2. Future Outlook

Future research should focus on innovative applications of regtech. From the perspective of economics, future research should establish a multi-dimensional signaling model. Digital footprint is no longer a simple risk label, but a multiple signal transmitted by individuals to the market. Given that there may be noise or camouflage signals in the footprint, the task of the credit model will evolve into decoding the real "creditworthiness" of individuals from massive and contradictory digital signals. This is not only a breakthrough in mathematical modeling, but also a profound reconstruction of human behavior logic in the era of digital civilization.

## 6. Conclusion

This paper aims to systematically explore the credit empowerment mechanism driven by big data and digital footprint. The research points out that through the deep integration of multi-dimensional real-time behavior data, big data technology has successfully promoted the transformation of the credit evaluation paradigm and achieved a fundamental change from traditional retrospective statistics to predictive insight.

In terms of technology evolution, the model algorithm has experienced a leap forward in development from traditional machine learning to deep learning, graph neural network and federated learning. This technological progress has not only significantly improved the accuracy of default prediction, but also enhanced the authenticity of risk identification, thus effectively alleviating the long tail group's long-term financial exclusion problem. In terms of application depth, the scope of credit empowerment has gradually expanded from basic financial services to frontier fields such as social governance and the metaverse. This fully proves that digital footprint, as an emerging "social capital", plays a central role in reshaping the mechanism of social mutual trust and reducing transaction costs.

Looking forward to the future, the credit evaluation model needs to be further transformed into a "trusted AI" framework, striving to meet the regulatory compliance requirements for algorithm interpretability while maintaining the prediction accuracy. At the same time, academia and industry should jointly build a multi-dimensional signaling model and deeply integrate regtech and privacy computing technology. Its core goal is to realize the deep decoding of credit value in the era of digital civilization under the premise of fully guaranteeing data sovereignty, and to continuously optimize the relevant governance system.

## References

[1] Berg T, Burg V, Gombović A, Puri M. On the rise of fintechs: Credit scoring using digital footprints. The Review of Financial Studies, 2020, 33(7): 2845-2897.

[2] Liu X, Guo X, Wu Y. A credit scoring model based on alternative mobile data for financial inclusion. Centre for Finance & Investment Working Paper, University of Edinburgh Business School, 2018.

[3] Óskarsdóttir M, Bravo C, Sarraute C, Vanthienen J, Baesens B. The value of social network data in credit scoring. Decision Support Systems, 2019, 116: 1-11.

[4] Khandani A K, Kim A J, Lo A W. Consumer credit risk models via machine-learning algorithms. Journal of Banking & Finance, 2010, 34(11): 2767-2787.

[5] Yang S, Zhang Z, Zhou J, Wang Y, Sun W. Financial risk analysis for SMEs with graph-based supply chain mining. In: Proceedings of the 29th International Joint Conference on Artificial Intelligence, 2020: 4661-4667.

[6] Cheng Y, et al. Federated Forest: Prediction and Training under 20-bit Heterogeneous Data for Privacy-preserving. Proceedings of the 37th International Conference on Machine Learning (ICML), 2020.

[7] Liu Y, Liu Z, Liang Y, Meng C, Zhang J, Zheng Y, Zhang D. Federated forest. IEEE Transactions on Big Data, 2020, 6(4): 634-651.

[8] Björkegren D, Grissen D. Behavior revealed in mobile phone usage predicts loan repayment. The World Bank Economic Review, 2020, 34(3): 618-633.

[9] Kostka G. China's social credit systems and public opinion: Explaining high levels of approval. New Media & Society, 2019, 21(7): 1565-1589.

[10] Venkata B. A Multi-Layered Framework for Decentralized Verification in Web3 Ecosystems, 2024.