# The applicability and rationality of "credit score based on digital footprint"

Qianhui Yang

The University of Waikato Joint institute at Zhejiang University City College, Hangzhou City University, Hangzhou, Zhejiang, 310000, China

Corresponding author's e-mail: 32111227@stu.zucc.edu.cn

**Abstract.** Because of the application of the internet and the strong penetration of mobile phone usage, financial institutions realize that credit scores can be performed based on the access records of each user, that is, digital footprints. This approach can not only predict consumer default but also make it easier for the unbanked. However, there are also a series of problems with using digital footprints for credit scores. This paper will mainly affirm the applicability of using the digital footprint model for credit scores by comparing the traditional and new credit score models, the FICO scoring system and the user-generated digital footprint model. Finally, describe the problems of the UGDF model and the measures to be taken.

**Keywords:** fintech; credit score; digital footprint; consumer default; big data.

## 1 Introduction

Digital footprint, sometimes also called a digital shadow or an electronic footprint, refers to the trial of data that users leave when using the internet [15]. It can grow in various ways, such as making comments on social media, shopping online, mailing, or other online activities. With the strong population of mobile phones and the application of the internet, financial institutions realized that credit scores can be performed based on digital footprints. In the past, financial institutions can only extend credit to customers who have bank accounts [9], so those who didn't have bank accounts couldn't obtain credit. Now, through UGDF, financial institutions can not only cater to the unbanked but also predict consumer default and reduce credit risk. Such social inclusion has far-reaching consequences for social and economic development. Traditional credit scores are derived from credit analyses performed by various credit bureaus and indicate whether an individual or institution is sufficiently capable to obtain a loan. Previously, these traditional institutions were the only ones capable of analyzing and scoring the financial history of the individual [5]. The World Bank says 1.7 billion people around the world are unbanked. They have no access to get these services at all. However, even in the U.S., according to Morning Consult, about one in 10 people is underbanked and 10% don't have any bank accounts at all [6]. With the development of technology, credit


A. Bhunia et al. (eds.), *Proceedings of the 2023 International Conference on Finance, Trade and Business Management (FTBM 2023)*, Advances in Economics, Business and Management Research 264,
https://doi.org/10.2991/978-94-6463-298-9_60

scoring using digital footprints are slowly beginning to outperform traditional credit scoring methods. Tobias Berg and other scholars published an article in July 2020, discussing the use of “digital footprints” to predict consumers’ credit defaults. They find that even those variables which are very simple and readily available can have predictive power comparable to that of credit scores [7]. However, they also indicate that the digital footprint is a supplement to credit bureaus rather than a substitute. In China, there is no universal credit scoring system. At present, China still mainly uses traditional information such as user personal information and credit records to score. The average default rate of housing mortgage loans of the Bank of China is much higher than that of the whole world. If the digital footprint is used for credit scoring, it has the potential to reduce default rates [12]. Then why credit scoring based on digital footprint is superior to traditional methods? Because its data sources are very rich, while traditional scoring methods only rely on a single data source, that is, personal credit records [5]. What’s more, the data of traditional methods are usually those updated over a long period of time, which cannot reflect the immediate changes in real-time. Instead, digital footprints are updated every hour to more fully reflect a user’s creditworthiness [12]. In this article, the applicability of the method of credit score using digital footprint will be explained in detail in two parts. By introducing and comparing the two models, it is concluded that there are improvements in credit scoring. Although the method of credit scoring using digital footprints is indeed highly practical, it also has certain problems. Through these problems, the article will give corresponding measures.

## 2 Progression from the traditional model to the new model

Traditional credit scoring models, such as the FICO system, 5C model, VantageScore, etc., use financial history to calculate an initial credit score. This results in the exclusion of "credit unscored" people, those with too little financial history, and "credit invisible" people without any financial history [21]. This situation is not conducive to social stability. However, with the popularity of mobile phones and the development of technology, we are living in a digital age of social networking and online information exchange. According to the latest research from the organization, as of June 2021, there will be 3.95 billion smartphone users, and the total global population will be 7.9 billion, which means 50% of the world's population owns smartphones [1]. As long as we're online, we're leaving a digital footprint that consists of all the data associated with our name that can be traced back to us. Digital footprints can be considered as an impression of our identity [7]. In this way, it looks suitable to use numerous digital footprints to make the credit score. Until now, some economists have proposed the UGDF model---user-generated digital footprint model [19]. Next, the traditional model with the FICO system as an example and the new model with the UGDF model as an example are introduced in detail respectively, and the progression from the traditional model to the new model will be obtained by comparison.

## 2.1 The FICO system

FICO, originally Fair, Isaac, and Company, is a data analytics company based in Bozeman, Montana, specializing in credit scoring services. It was founded in 1956 by Bill Fair and Earl Isaac [27]. FICO is an indicator to measure consumer credit risk, that is, lenders use a borrower's FICO score, along with other details from a borrower's credit report, to assess credit risk and decide whether to extend credit [13], which has already become a fixture of American consumer lending [27].

### 2.1.1. Calculation mode.

The FICO score is determined by five parts, namely payment history, amounts owed, length of credit history, new credit, and credit mix [16]. Among them, payment history accounts for 35%, amounts owed account for 30%, length of credit history accounts for 15%, new credit accounts for 10%, and credit mix accounts for 10%. Payment history, which is the largest factor in the FICO scores, records how an individual has paid over the credit term and whether they have paid their credit account on time. Amounts owed, referring to how much debt an individual bears, can predict future credit performance. However, compared to the debt amount, the credit utilization ratio is more important [19]. The length of credit history means how long has the account been opened. The longer the credit history, the higher the credit score. Credit mix, that is, the diversity of accounts. The stronger the credit mix, the higher the credit score. As for the new credit, if a user opens a large number of accounts in a short period of time, it is very likely that he will engage in extremely risky activities, and his credit score may drop. Traditional credit institutions such as banks obtain the final FICO score of users based on the five categories of users. Table 1 shows the credit ratings corresponding to different credit score ranges.

**Table 1.** What a FICO score means [13]

| FICO Score | Rating | What the Score Means |
|---|---|---|
| <580 | Poor | Well below average<br>Demonstrates to lenders that you’re a risky borrower |
| 580-669 | Fair | Below average<br>Many lenders will approve loans |
| 670-739 | Good | Near or slightly above average<br>Most lenders consider this a good score |
| 740-799 | Very Good | Above average<br>Demonstrates to lenders you’re a very dependable borrower |
| 800+ | Exceptional | Well above average<br>Demonstrates to lenders you’re an exceptional borrower |

### 2.1.2. Disadvantages.

Firstly, due to the different proportions of different categories, even if an individual loses a lot of points in a certain category, it may not affect the whole, so the personal credit score is not very accurate. For example, a late payment on a credit report does not mean that you cannot get a "perfect score", because the payment information record is only one piece of information to calculate the fico score [20]. And acts of paying bills on time or contacting creditors can potentially enhance payment history to compensate for late payments [20]. Secondly, over-reliance on credit history, and credit card data, results in narrow coverage. Thirdly, it can only represent the past behavior of the borrower but cannot predict the performance of the future borrower.

## 2.2 The UGDF model

UGDF is a user-generated digital footprint, which uses the digital footprint for credit scoring. But actually, the UGDF model is not a specific name, but a very general concept. UGDF models use data related to digital footprints [18], use these as research data, and then use machine learning, and other algorithms for predictive training. This research method is like the common linear regression in China, where data and related indicators are replaced to establish a linear regression model.

### 2.2.1. Background of the UGDF model.

Advances in artificial intelligence and machine learning enable scoring algorithms to calculate credit scores using non-financial data such as digital footprints and psychometric data from mobile devices. Big Data Brings Breakthrough Changes to Traditional Credit Scoring. Campbell-Verduyn et al. (2017) [2] discuss how big data is penetrating the financial services industry through credit bureaus and fintech companies that are using big data in their algorithms. In the era of big data, the data used for credit scoring has expanded from credit history to social network usage records and digital footprints [8]. One company, Lenddo, reportedly bases its credit scoring on information from users' social networking profiles, such as education, employment history, number of followers, the status of friends, and information about those friends [22]. Meanwhile, a growing number of startups like Lenddo use data exclusively from social networks. The companies claim that social network-based credit scores broaden opportunities for more people and could benefit low-income individuals who would otherwise struggle to access credit. The paper of Wei et al. (2016) [8] and Kshetri (2016) [26] also affirm that big data can provide credit scoring for potential borrowers with limited financial histories, thereby enabling them also to access financial services. While the FICO system remains at the heart of contemporary lending decisions, many traditional scoring agencies are also developing new models based on non-traditional data. For example, Experian has used big data to develop a "universal customer profile" that combines information from the online and offline activities of thousands of customers [23]. After testing the use of non-traditional data to provide customers with little loan history, FICO finally developed a new system with Equifax called FICO Score XD, which uses consumers' wired information for credit scoring. In tests, FICO XD allowed more than half of previously unscorable credit card applicants to score, said Jim Wehmann, FICO's executive vice president of scoring. In a test with a large auto lender, the new score increased loan approval rates by 24%, TransUnion said [3]. Therefore, with the support of UGDF, the evaluation seems to be better. However, although alternative data like

UGDF can go a long way toward increasing financial inclusion, it's not perfect. So it still needs to take time to tell if the use of digital footprints for credit scoring can be fully applied to everyone.

#### 2.2.2. Application.

Now many financial institutions mainly use UGDF data instead of traditional data to build models when using digital footprints for credit scoring. For example, Swedish payments provider Klarna trained its model on new UGDF variables, such as time to execute a transaction, OS version, and so on. In addition, KrediTech enhances its model with behavioral UGDF variables. For example, in China, JD. com’s credit score database based on consumers’ online shopping records and personal demographic data. JD.com has developed a payment method called Baitiao, that is, buy first, pay later. By leveraging its vast data capacity and customer base, JD.com can rate consumers based on their online shopping and browsing history. It also uses standard modeling techniques to assign a credit score to each customer. Based on JD credit data, a consumer’s credit score is generated from five aspects, respectively demographic information, consumption preference, asset level, social network, and delinquency records [11]. The part of demographic information consists of age, gender, education, and so on. Consumption preference is measured by shopping records, which can also affect one’s credit situation. What’s more, the higher the asset level, the higher the credit score. Adversely, the more delinquency records, the lower the credit records. And the one who has a more active social network, he/she may have a higher credit score. Using JD Credit Data, the baseline model can generate an AUC score of 70% [11]. The AUC measure is used to judge the discriminative power of the default prediction model. According to Iyer, Khwaja, Luttmer, and Shue (2016) [14], an AUC of 70% or more is considered a good result in an information-rich environment. Therefore, it is confirmed that JD. Com is good at evaluating the rate of credit default behaviors. What’s more, a German e-commerce company also made a test. It divided a data sample containing 270399 transaction records into two groups [24], namely group A and group B. When analyzing group A, the researchers use both credit bureau score and digital footprint variables, which is called multivariate regression analysis. And finally, the result of the AUG value is 73.6%. However, when analyzing group B, the researchers use only digital footprint variables, without credit bureau score, which is called univariate regression analysis. And the result of the AUG value is 69.6%. Therefore, it can be concluded that digital footprint data is not a complete replacement of traditional credit data, but a supplement to make scoring results more accurate.

### 2.3 Progression

There are two main progressions. Firstly, increasing financial inclusion. The FICO system excludes “credit invisible” or “thin file” consumers because they do not have enough traditional data to describe their financial payment history [4], especially for those unbanked poor people. Secondly, evaluation results are more accurate. Actually, the UGDF model will not completely replace traditional credit models like FICO. On the contrary, based on the FICO system, the participation of UGDF makes the evaluation results more accurate. TransUnion said its new score combines a deeper analysis of a consumer's traditional payment history with alternative data to more accurately assess risk. Mr. Mondelli, president of TransUnion, said that while the traditional score records whether you make minimum required credit card payments on time, the alternative data includes the size of your monthly payments and whether payments increase or decrease over time [3].

# 3 Problems and solutions

Judging from the progress from the traditional model to the new model, credit scoring based on digital footprint is indeed more practical and more financially inclusive. However, this scoring is not perfect. There are risks of inaccuracy and privacy leaks.

## 3.1 Privacy leaks

A digital footprint is the collection of data on any activity of an individual on the Internet, including personal details such as e-mail or phone number which are most likely to leak. Some hackers or criminals illegally access computer networks and download large amounts of information for crimes. The most frightening thing is that criminals may commit crimes in your name. However, those who are only concerned about the score drop in the event of a breach may be disappointed because even if the data breach itself does not cause a score drop, the scammers using the leaked data to create a new credit line may cause a score drop [17].

## 3.2 All visits are logged

Digital footprint records the trail of data the user leaves when using the internet, so it’s keeping the track of everything that goes online. For example, those who spend at night get lower ratings, as it is generally believed that night is when crime is highest. However, there are always such a group of people who like to spend money at night purely due to living habits, so how can these records determine their low credit score? The Internet can't erase the record from the computer but might make up the credit score from somewhere else.

## 3.3 High cost

For some small new financial technology companies, it is difficult to construct a UGDF model, because it requires a large number of data sources, that is, to satisfy data accessibility, data coverage, data timeliness, and data authenticity [18]. So it requires a lot of data, as well as a high cost. Firstly, it has the cost of data collection. There have three data sources, namely fully owned, purchased from third parties, and collected from public resources. UGDF model needs to have different variables, and different variables will fluctuate with market supply and demand fluctuations. Secondly, it needs the cost of algorithm acquisition. Using digital footprints for credit scoring is a quantitative process, so it requires the use of some algorithms, and these algorithms are patented by certain developers and are not publicly available [25]. Finally, it needs calculating facility cost. Therefore, the addition of these three costs is not a small fee, which is very difficult for some newly established companies.

Although these problems cannot be avoided, the government and fintech companies also need to take certain measures to improve. The government should formulate relevant laws and regulations to strictly prohibit third parties or companies from leaking users' personal information to criminals. At the same time, the government should regulate the situation where third parties charge companies large amounts of data collection fees and encourage companies not to rely too heavily on digital footprints. As for the companies, Companies need to regularly update and encrypt internal network traffic tools to prevent vulnerabilities from being attacked by hackers and attackers [10].

# 4 Conclusion

More and more traditional financial institutions are experimenting with digital footprints for credit scoring. This is not only conducive to expanding financial inclusion and enhancing social equality but also makes the scoring structure more accurate. But it is undeniable that there are many loopholes in this scoring method, the most serious one is the problem of privacy leakage. Therefore, more research is needed to demonstrate whether digital footprint variables can be applied universally to all consumers.

# Reference

1. After 27 years, global smartphone users reached 4 billion, accounting for 50% of the total population. (n.d.). https: // baijiahao. baidu. com/s? id = 1703794472463576439&wfr =spider&for=pc
2. Campbell-Verduyn, M., Goguen, M. and Porter, T. (2017), Big data and algorithmic governance: the case of financial practices, New Political Economy, Vol. 22 No. 2, pp. 219-236.
3. Carrns, A. (2015, October 9). New Credit Score Systems Could Open Lending to More Consumers. The New York Times. https://www.nytimes.com/2015/10/10/your-money/new-credit-score-systems-could-open-lending-to-more-consumers.html
4. Chopra, S. (2021). Current Regulatory Challenges in Consumer Credit Scoring Using. Alternative Data-Driven Methodologies. Vanderbilt Journal of Entertainment & Technology Law, 23(3), 625. https://scholarship.law.vanderbilt.edu/cgi/viewcontent.cgi?article = 1044 &context=jetlaw
5. Credolab. (2021). Why is the traditional credit rating system losing steam? https: //www. credolab.com/blog/why-is-the-traditional-credit-rating-system-losing-steam
6. Credolab. (2022). How to combine digital footprint lookups with behavioural data to turbocharge credit scoring. https://www.credolab.com/blog/how-to-combine-digital-footprint-lookups-with-behavioural-data-to-turbocharge-credit-scoring
7. Cveticanin, N. (2023, January 20). What is a Digital Footprint: Everything You Need To. Know. Dataprot. https://dataprot.net/articles/what-is-a-digital-footprint/

8. Dwoskin, E. (2015), Lending startups look at borrowers' phone usage to assess creditworthiness, The Wall Street Journal.
9. FinScore. (n.d.). Credit Scoring for All. https://www.finscore.ph/
10. Fruhlinger, J. (2020, February 12). Equifax data breach FAQ: What happened, who was. affected, what was the impact? CSO online. https://www.csoonline.com/article/ 3444488/ equifax-data-breach-faq-what-happened-who-was-affected-what-was-the-impact.html
11. Fu, G., Sun, M., & Xu, Q. (2020). An Alternative Credit Scoring System in China's Consumer. Lending Market: A System Based on Digital Footprint Data. Social Science Research Network. https://doi.org/10.2139/ssrn.3638710
12. Guanghong, F., Minjuan, S., Qing, X. (2020). An Alternative Credit Scoring System in China's Consumer Lending Market: A System Based on Digital Footprint Data. SSRN, 508: 1-51.
13. Hayes, A. (2023, February 18). FICO Score. Investopedia. https://www.investopedia. com/ terms/f/ficoscore.asp
14. Iyer, R., Khwaja, A. I., Luttmer, E. F. P., & Shue, K. (2016). Screening Peers Softly: Inferring the Quality of Small Borrowers. Management Science, 62(6), 1554–1577. https: //doi. org/10.1287/mnsc.2015.2181
15. Kaspersky. (n.d.). What is a digital footprint? And how to protect it from hackers. https://www.kaspersky.com/resource-center/definitions/what-is-a-digital-footprint
16. Lake, R. (2023, March 20). FICO Credit Scores Explained. Investopedia. https://www.investopedia.com/fico-credit-scores-explained-5072985
17. Localmemphis. (2023, Janarary 20). No, your credit score won't automatically take a hit after a data breach [Video]. 24ab. https://www.localmemphis.com/video/news/verify/ money-verify/vfy-data-breach-pkg/536-aa6891b9-24ab-4792-8496-fe40f42e44d0
18. Loutfi, A. A. (2022). A framework for evaluating the business deployability of digital. footprint. based models for consumer credit. Journal of Business Research, 152, 473-486. https://doi.org/10.1016/j.jbusres.2022.07.057
19. myFICO. (n.d.-a). How Owing Money Can Impact Your Credit Score | myFICO. https://www.myfico.com/credit-education/credit-scores/amount-of-debt
20. myFICO. (n.d.). How Payment History Impacts Your Credit Score | myFICO. https://www.myfico.com/credit-education/credit-scores/payment-history
21. Njuguna, R. G., & Sowon, K. (2021). Poster: A Scoping Review of Alternative Credit. Scoring Literature. The Compass. https://doi.org/10.1145/3460112.3471972
22. Rusli EM (2013) Bad credit? Start tweeting. Wall Street Journal (April 1). http://www.wsj.com/articles/SB10001424127887324883604578396852612756398.
23. Tewksbury, M. (2013). The 2013 Big Data Planning Guide for Marketers, Experian Marketing Services. http://www.experian.com/assets/marketing-services/white-papers/big-data- planning-guide-for-marketers.pdf
24. Tobias, B., Valentin, B., Ana, G., Manju, P. (2020). On the Rise of FinTechs: Credit Scoring Using Digital Footprints. The Review of Financial Studies, 33: 2845-2897.
25. Wang, B., & Gong, N. Z. (2018). Stealing hyper parameters in machine learning. IEEE. Symposium on Security and Privacy (SP), 2018, 36–52. https://doi.org/10.1109/ SP. 2018. 00038
26. Wei, Y.'., Yildirim, P., Van Den Bulte, C., & Dellarocas, C. (2016). Credit Scoring with Social Network Data. Marketing Science, 35(2), 234–258. https: //doi. org/10. 1287/ mksc. 2015.0949
27. Wikipedia contributors. (2023, March 20). FICO. Wikipedia. https://en.wikipedia.org /wiki/ FICO