Dean&Francis

ISSN 2959-6130

A study on the application of digital footprints in credit scoring

Jiawei Wang

City University of Macau, faculty of data science, Macau, 999078, China

Abstract:

In the modern economic model, credit scores are an indispensable part of the credit system, which can predict whether people can get loans and how much loans they can get. This paper will define digital footprint and credit score, disband and analyze the specific process of credit scoring based on digital footprint, and discuss the similarities and differences and advantages of emerging digital footprint credit scoring methods compared with traditional digital scoring methods. In addition, the credit analysis is carried out based on the loan data of the states in the United States in 2015 obtained from lending club, and on this basis, the empirical research on the credit score of the states is further divided according to the red and blue continents.

Keywords: credit scoring, digital footprint,big data,-credit risk

1. Introduction

Credit scoring is an indispensable part of the modern economic system, which can foretell whether people can get loans and how much loans they can get, etc., and the traditional credit scoring methods can no longer meet the growing demand for credit scoring, and the technique of applying digital footprints to credit scoring was born in this context.

Research Status:Mikella Hurley* & Julius Adebayo (2016) compared the use of digital footprints for credit scoring with traditional credit scoring methods, pointed out the risks of using digital footprints for scoring, and proposed technical and policy improvements to improve the credit scoring system using digital footprints. Andreas Fuster (2019) discusses the advantages of using digital technology for lending versus traditional lending, and Sandra C Matz (2019) shows that people's privacy is not protected when digital footprints are being used on a large scale, and that there is still a need to regulate the use of digital footprints with appropriate legal constraints.Tobias Berg (2020) used digital footprints to determine the likelihood of a person's successful loan repayment after borrowing, and the accuracy of the data derived from digital footprints was even higher than that given by the official credit assessment bureaus, demonstrating the accuracy of digital footprints for assessing an individual's propensity.Christophe Croux (2020) analyzed lendingclub data to derive the factors that specifically affect loan repayment rate and the corresponding influence coefficients.Ahmad Amine Loutfi (2022) argues that some of the models that use digital footprints for credit scoring only focus on the accuracy of the final result while ignoring the other factors, thus setting up a highly deployable framework.Lili Dai et al. (2023) further develops a framework for the use of digital footprints for credit scoring on the basis of the the role of digital footprints in financial lending and found that digital footprints not only improve the accuracy of credit scoring, but also have a great effect on debt collection after lending, with a 26.5% increase in repayment rate. Nowadays, digital footprints are still not exploited on a large scale, and there is still huge room for development.

2. Definition of Credit Score and Digital Footprint

2.1 Definition of credit scoring

Credit scoring refers to the evaluation of an individual by a credit scoring agency that is derived from the analysis of information from various sources and is expressed in the form of a score. It aids major financial institutions in making decisions about lending, and the technique determines whether or not to grant a person a loan, how much, as well as the interest rate of the loan, the repayment period, and so on. Credit scoring dates back to the 1930's and was introduced during World War II, replacing the traditional subjective assessment by creditors to decide whether or not to grant a loan. Today, credit scoring is not only used to help financial institutions assess whether a loan can be repaid, but also to help lenders assess how much benefit they can derive from a customer's loan. What's more, credit scoring has been applied to people's daily life, such as sharing bicycles and renting rechargeable batteries. Credit scoring has become an integral part of the modern economic system, and changes to credit scoring mechanisms have been on the agenda in recent years, with digital footprint-based scoring being particularly effective.

2.2 Definition of digital footprint

The concept of digital footprint has emerged in the last decade and has been applied in many fields. In fact, as early as in the web 1.0 era, Negroponte put forward the concept of "Slug trail", this concept is the predecessor of the digital footprint. After continuous development, only in the era of web3.0 formed the current digital footprint. Its concept is the data traces left by people when they use the Internet. This includes websites visited, emails sent, and information submitted online, among others [7]. Immediately after it was developed, the digital footprint was utilized in a number of areas, with the most widely known application being the product/video recommendation algorithm, which automatically suggests to the user what the user is interested in next, based on what the user has previously searched for. Similar algorithms use only the user's previous searches/viewing history to achieve a roughly accurate content push. A digital footprint for credit scoring, on the other hand, requires a much more diverse and rich set of information. This can be used to improve the accuracy of credit scoring.

3. The specific process of digital footprint credit scoring

3.1 The process of credit scoring with digital footprints

Data sampling: before using digital footprints for digital scoring, we first need to obtain data, the target data need to be selected in strict accordance with the criteria; first, we need to obtain the credit score given by the traditional credit score bureaus, based on which the data also need to contain a variety of information [3], all the factors related to credit scoring need to be included in the data obtained such as the price of the equipment used, the type of system used, and even the brand of the mailbox used and whether the mailbox number with a name need to be included in the information obtained, the more accurate the resulting credit score. All factors related to credit scoring need to be included in the acquired data, such as the price of the equipment used, the type of system used, and even the brand of the e-mail used and whether the e-mail number contains a name or not, and the more types of information acquired, the more accurate the credit scores will be.

Data Adaptation: Since digital footprints are the information that users leave behind when they use the Internet, the type of information we obtain can be varied. Some of the data can simply be used in the same way as the type of data we originally obtained, such as age, credit bureau score, etc., but most of the digital footprint data needs to be quantified or segmented into blocks before it can be put to use, such as gender, type of device, time of purchase, etc.

Variable correlation measurement [4]: Due to the mixed sources of digital footprint information, it can not be used directly as standard data, such as the correlation between the variables is too large will lead to large multicollinearity, when the valuation accuracy is greatly reduced. Therefore, it is necessary to detect the correlation between the variables first, and then filter these variables into the calculation of credit score.

Modeling: This process involves the use of regression analysis to model a variety of data to find the most accurate valuation of a credit score, which is the most central part of calculating a credit score with a digital footprint.

Conclusion drawn: the final estimated credit score can be obtained by multiplying each variable with the coefficients assigned after calculation through the data model.

4. Using the data in lending club as an example, explaining the role of digital footprints in credit scoring

This study intercepts data from lending club 2015 loans and conducts a series of analyses to specifically demonstrate the impact of digital footprints on credit score prediction in conjunction with the process of applying digital footprints to credit scores.

4.1 Data Acquisition

The data for this study was obtained from the specific loan data of lending club[2] on Kaggle website which is open to the public in the year 2015, a total of 42,538 loan records were recorded containing various lender information such as loan amount, credit bureau credit scores, and the purpose of the loan.

4.2 Data processing

The topic of this research is the study of the application of digital footprints in credit scoring, so the variable of loan address in all the data is chosen to analyze what effect the difference of loan address will have on the repayment rate.

4.3 Relevance Quiz

Since the loan address is a single variable, no correlation test is required.

4.4 Comparison of methods

In this study, all the loan information was divided into fifty data blocks according to the continent from which the loan originated, and then the number of people who took out loans, the number of people who repaid the loans when due, the number of people who repaid the loans late, and the number of people who were confirmed to be unable to pay back the loans were counted in each of the data blocks separately. These four data were further processed to derive the loan repayment rate for each state.

Table 1: Repayment Rates by State

name of continent	Total number of borrowers	Number of persons outstanding	Proportion of persons not reimbursed		name of continent	Total number of borrowers	Number of persons outstanding	Proportion of persons not reimbursed
ME	3	0	0		WI	516	77	0.149224806
DC	224	17	0.075892857		NC	830	127	0.153012048
WY	87	8	0.091954023		KY	359	55	0.153203343
ID	9	1	0.111111111		AZ	933	144	0.154340836
DE	136	16	0.117647059		MI	796	123	0.154522613
KS	298	36	0.120805369		NJ	1988	310	0.155935614
VT	57	7	0.122807018		WA	888	143	0.161036036
WV	187	23	0.122994652		MD	1125	186	0.165333333
LA	461	58	0.125813449		UT	278	46	0.165467626
MA	1438	185	0.128650904		HI	181	30	0.165745856
TX	2915	377	0.129331046		NM	205	34	0.165853659
CO	857	111	0.129521587		CA	7428	1233	0.165993538
RI	208	27	0.129807692		GA	1503	250	0.166333999
CT	816	106	0.129901961		IA	12	2	0.166666667
AL	484	63	0.130165289		OR	468	78	0.166666667
AR	261	34	0.130268199		MT	96	17	0.177083333
VA	1487	194	0.130464022		MO	765	140	0.183006536
OH	1329	174	0.130925508		FL	3104	571	0.183956186
PA	1651	225	0.136281042		TN	32	6	0.1875
NY	4065	558	0.137269373		SD	67	13	0.194029851
OK	317	44	0.138801262		AK	86	17	0.197674419
IL	1672	234	0.139952153		MS	26	6	0.230769231

MN	652	92	0.141104294		NV	527	125	0.237191651
SC	489	70	0.143149284		IN	19	7	0.368421053
NH	188	27	0.143617021		NE	11	6	0.545454545

Drawing conclusions:

Based on the analysis and processing of all the loan information of lending club 2015 we can finally get the following results:

From the results of the data obtained we can see that the repayment rates are very significantly different from one loan address to another, with the highest repayment rate being as high as 100 percent in Maine ME and the lowest in Nebraska NE.

Only forty-five percent repayment rate. After removing the extreme case of a sample size of less than fifty, the highest repayment rate of ninety-two percent in Washington D.C. and the lowest repayment rate of seventy-six percent in Nevada NV. have a striking sixteen percent interpolation, and in this regard we can conclude that the address of the loan application in the digital footprint has a strong predictive effect on the repayment rate. We can conclude that the address of the loan application in the digital footprint is a strong predictor of repayment rates.

We can explore this result further by further dividing the sample of continents used in this loan data by party support into red, blue, and swing states, with red continents referring to those that support the Republican Party and blue continents referring to those that support the Democratic Party.

The swing party is the continent that has not yet made a clear preference, and then calculate the repayment rates for each of these three broad categories.

Table 2: State Repayment Rates by Party Affiliation

blue continent	Number of loans	Number of repayments	outstanding rate	red continent	Number of loans	Number of repayments	outstanding rate	swing state	Number of loans	Number of repayments	outstanding rate
ME	3	0	0	GA	1503	250	0.16	ID	9	1	0.11
CT	816	106	0.12	WV	187	23	0.12	CO	857	111	0.12
VT	57	7	0.12	SC	489	70	0.14	VA	1487	194	0.13
RI	208	27	0.12	TX	2915	377	0.12	OH	1329	174	0.13
NY	4065	558	0.13	AL	484	63	0.13	AZ	933	144	0.15
MA	1438	185	0.12	MS	26	6	0.23	IA	12	2	0.16
NJ	1988	310	0.15	LA	461	58	0.12	FL	3104	571	0.18
MD	1125	186	0.16	AR	261	34	0.13	AK	86	17	0.19
DE	136	16	0.11	MO	765	140	0.18	NV	527	125	0.23
CA	7428	1233	0.16	TN	32	6	0.18	stagger	8344	1339	0.16
IL	1672	234	0.13	KY	359	55	0.15				
MN	652	92	0.14	OK	317	44	0.13				
WA	888	143	0.16	NE	11	6	0.54				
NM	205	34	0.16	KS	298	36	0.12				
OR	468	78	0.16	SD	67	13	0.19				
HI	181	30	0.16	MT	96	17	0.17				
PA	1651	225	0.13	UT	278	46	0.16				
NH	188	27	0.14	WY	87	8	0.09				
DC	224	17	0.07	NC	830	127	0.15				
MI	796	123	0.15	IN	19	7	0.36				
WI	516	77	0.14	republican	9485	1386	0.14				

Democratic Party	24705	3708	0.15								

Based on the data generated we can see that Red Continent has the highest repayment rate, Blue Continent the next highest and Swing State the lowest. The above differences may be related to local economic policies and so on.

5. Advantages of Digital Footprinting for Credit Scoring

5.1 Low ease of access to data

First, as opposed to the traditional method of obtaining credit scores from the credit bureaus, which requires businesses to pay the credit bureaus a significant fee, businesses

There is no way to bypass the credit bureaus to obtain accurate credit score data. Unlike digital footprints, a company can set up user terms and conditions based on a web page, and as long as the user agrees to the collection of information, the company can directly access all the digital footprints that the user has left on the Internet for the cost of designing and setting up the data collection system on the web page, which can be used on an ongoing basis. The cost of setting up the system is essentially negligible compared to purchasing a credit score from a credit bureau.

5.2 Wide data coverage

Compared to traditional credit scoring, digital footprint-based credit scoring uses a much wider range of data. Traditional scoring methods contain a single piece of information, only the necessary economic information, such as the amount of savings, job status, assets owned, etc., while the digital footprint contains a range of information beyond the purely economic category, including a variety of user behavior on the Internet, such as browsing history, social media activities, device models, residential address, online shopping habits, etc., to make a more rigorous assessment of the borrower, a clearer picture of the borrower. This enables a more rigorous assessment of the borrower and a clearer understanding of the borrower.

5.3 High credit score accuracy

Thanks to the information characteristics of the digital footprint and the huge amount of information, the company can make a more comprehensive information portrait of the lender [6], combined with the above analysis of the impact of using the loan address in the digital footprint as a variable on the repayment rate, we can conclude that the digital footprint has a certain predictive effect on the repayment rate. In this regard, it can be concluded that combining the digital footprint as a supplement to the traditional credit scoring will greatly improve the accuracy of credit scoring.

5.4 high speed

A large part of the information source of traditional credit scoring method is similar to fixed assets, fixed deposits and other indicators that will not change for a long time. These indicators can reflect the economic status of the customer, but the update cycle is very long, and cannot accurately show the real-time status of the customer. Unlike digital footprints, which can be updated on a daily or even minute-by-minute basis, credit scores based on digital footprints will be more reflective of a customer's current situation than traditional credit scores.

5.5 Wide user coverage

Credit scoring based on digital footprints provides an opportunity for unbanked users to be able to obtain credit loans. [1] Traditional credit scoring places great emphasis on a loan applicant's bank account deposits, and an unbanked applicant often fails to get a good credit score, or even a credit score at all. Digital footprints do not focus on whether an applicant has a bank account or not, as long as the applicant has not been out of touch with modern society for too long and is using any Internet-related products such as cell phones, computers, etc., then the digital footprints can gather useful information and provide an unbiased score for the applicant to be able to apply for a loan.

6. Summary

Credit scoring, as one of the important indexes in modern economic system, the traditional scoring method has long been unable to meet the growing demand for credit scoring. Combined with digital footprint used in credit scoring as an emerging technology, to a large extent, to solve the shortcomings of the traditional scoring methods, this study combined with case studies and literature review to analyze in detail what are the advantages of digital footprint-based credit scoring compared to the traditional scoring methods and the reasons for it, so that people can have a deeper understanding of the digital footprint for credit scoring.

7. References

[1] Tobias Berg, Valentin Burg, Ana Gombović, Manju Puri, On the Rise of FinTechs: Credit Scoring Using Digital Footprints, The Review of Financial Studies, Volume 33, Issue7,July2020,Pages2845-2897, https://doi.org/10.1093/rfs/hhz099

[2] Andreas Fuster, Matthew Plosser, Philipp Schnabl, James Vickery, The Role of Technology in Mortgage Lending, The Review of Financial Studies, Volume 32 , Issue 5, May 2019, Pages 1854-1899, https://doi.org/10.1093/rfs/hhz018

[3] Ahmad Amine Loutfi,A framework for evaluating the business deployability of digital footprint based models for consumer credit,Journal of Business Research, Volume 152, 2022, Pages473-486, ISSN0148-2963, https://doi.org/10.1016/j.jbusres.2022.07.057.

[4] Credit Scoring in the Era of Big Data Hurley, Mikella Hurley, Mikella Adebayo, Julius 2016 YALE J.L. & TECH. 18 , pp. 148

[5] Christophe Croux, Julapa Jagtiani, Tarunsai Korivi, Milos Vulanovic,Important factors determining Fintech loan default:Evidence from a lendingclub consumer platform,Journal of Economic Behavior&Organization,Volume173,2020, Pages270-296,ISSN0167-2681,https://doi.org/ 10.1016/j.jebo.2020.03.016.

[6] Dai, Lili and Han, Jianlei and Shi, Jing and Zhang, Bohui, Digital Footprints as Collateral for Debt Collection (May 7, 2023). Available at SSRN. https://ssrn.com/abstract=4135159 or http://dx.doi.org/10.2139/ssrn.4135159

[7] Sandra C Matz, Ruth E Appel, Michal Kosinski,Privacy in the age of psychological targeting,Current Opinion in Psychology,Volume31,2020,Pages116- 121,ISSN2352-250X,https://doi.org/10.1016/j.copsyc.2019.08.010.