

Zhang, Shaofeng; Xiong, Wei; Ni, Wancheng; Li, Xin

Article

Value of big data to finance: observations on an internet credit Service Company in China

Financial Innovation

Provided in Cooperation with:

Springer Nature

Suggested Citation: Zhang, Shaofeng; Xiong, Wei; Ni, Wancheng; Li, Xin (2015) : Value of big data to finance: observations on an internet credit Service Company in China, Financial Innovation, ISSN 2199-4730, Springer, Heidelberg, Vol. 1, Iss. 17, pp. 1-18, <https://doi.org/10.1186/s40854-015-0017-2>

This Version is available at:

<https://hdl.handle.net/10419/176409>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

RESEARCH

Open Access



Value of big data to finance: observations on an internet credit Service Company in China

Shaofeng Zhang¹, Wei Xiong^{1*}, Wancheng Ni² and Xin Li³

* Correspondence:

wei.xiong@100credit.com

¹100Credit Financial Information Service Co., Ltd., Beijing 100089, China

Full list of author information is available at the end of the article

Abstract

Background: his paper presents a case study on 100Credit, an Internet credit service provider in China. 100Credit began as an IT company specializing in e-commerce recommendation before getting into the credit rating business. The company makes use of Big Data on multiple aspects of individuals' online activities to infer their potential credit risk.

Methods: Based on 100Credit's business practices, this paper summarizes four aspects related to the value of Big Data in Internet credit services.

Results: 1) value from large data volume that provides access to more borrowers; 2) value from prediction correctness in reducing lenders' operational cost; 3) value from the variety of services catering to different needs of lenders; and 4) value from information protection to sustain credit service businesses.

Conclusion: The paper also discusses the opportunities and challenges of Big Data-based credit risk analysis, which needs to be improved in future research and practice.

Keywords: Big data, Credit rating, Information economics, Value of information, Finance

Background

Credit management is the basis of the financial industry (Lin et al. 2015). When lenders provide loans to individuals or companies, they need to assess the borrowers' credit risk to reduce the possibility of bad debt and decide the amount of the loan. Lenders can collect information on the borrower either individually or from other lenders. Information exchange between lenders can occur voluntarily via private credit service or be enforced by regulation via public credit bureaus (Brown et al. 2009). Such information exchange can significantly reduce the cost of the credit market and make it more efficient. In developed countries, the credit system has experienced over a hundred years of development. In China, the public credit system is far behind its economic growth, which causes some deficiencies in the lending market. The need for credit services has led to the appearance of private credit providers in China.

Nowadays, many private credit service providers in China use data collected from the Internet and other data sources. Their practices provide a good example of using Big Data technology to serve the financial industry. In this study, we examine Internet credit services through the practices of 100Credit (www.100credit.com), a new Internet

credit service provider in China, with the goal of deepening our understanding of the value of Big Data to Internet credit services.

The capital market in China around 2010

Since the 1980s, China has gradually established a system of capital markets. Enterprises can raise capital from multiple sources, such as the equity market, bank, and private equity. In 2010, there were 1718 listed companies in China's equity markets, which raised about 1.03 trillion and 1.26 trillion RMB inside and outside of China (China Securities Regulatory Commission 2011). Banks are another very important source of capital for companies not on the equity market. In 2010, Chinese banks held about 30 trillion RMB of loans to large and medium-size companies, and small enterprises in China (People's Bank of China 2011). However, it should be noted that about 75 % of these loans were for large and medium-size enterprises. Most banks target large enterprises for making loans, which are considered to have lower risk. Loans to small enterprises are insufficient and difficult to get.

In addition to the needs of enterprises, individual are also important borrowers. Individuals' capital needs are often represented as housing mortgages, automobile loans, durables goods loans, medical service loans, home remodeling loans, education loans, etc. (Wang 2009). Sometimes, personal loans are used to fund small businesses. From 2001 to 2013, personal loans increased from 643.4 billion RMB (5.74 % of all loans) (People's Bank of China 2002) to 29.66 trillion RMB (41.25 % of all loans) (People's Bank of China 2014). Nevertheless, personal loans provide major business for commercial banks in developed countries (Huang 2007). In comparison, China's personal loans are way behind Chinese people's needs.

One problem for small companies and individuals in getting loans is that they often do not have a long credit history and are considered high risk by banks. In China, a significant portion of such unfilled capital needs are met through private lenders, which base credit on relationships. As reported by the 2012 Blue Book of China Society published by the Chinese Academy of Social Sciences (Dong 2012): China's private lending market was about 740.5 ~ 816.4 billion RMB in 2003 (estimated by the Central University of Finance and Economics), about 950 billion RMB in 2005 (estimated by the People's Bank of China), and more than 4 trillion RMB in 2011 (estimated by the China CITIC bank). In 2010, Chinese Banks tightened their credit requirements on making loans, which significantly increased the capital needs in the private lending sector (Knowledge@Wharton 2011). In private lending, the lenders often set high interest rates to solicit depositors and set high interests on lending, which is illegal in China.

In recent years, the development of the Internet has led to (the growth of) P2P lending, which partially alleviates the problem of capital needs. P2P lending provides an online platform that helps borrowers and investors find each other. Due to the use of online platforms, P2P lending can go beyond social relations and location restrictions in private lending. China's first P2P lending platform appeared in 2006. Before 2010, there were about 10 ~ 20 platforms with a monthly investment of about 0.5 billion RMB. At this stage, most platforms partially played the role of a credit service and specified a credit line for each borrower to direct lender's investments. After 2010, there were about 200 platforms with a monthly investment of about 3 billion RMB. At

this stage, most lenders used P2P platforms as an information channel and conducted their own investigation on borrowers' creditability before committing to an investment. From 2010 to 2013, P2P platforms in China grew to number about 600 with a monthly investment of about 11 billion RMB. In this stage, the platforms showed certain characteristics of private lending where borrowers often set a high interest rate and there was a lack of careful credit assessment. As a result, the risk level of P2P lending was quite high. At the end of 2013, repayment became a common issue and many platforms closed. After 2014, with tightened regulations on P2P lending, the platforms stabilized at about 300, which operated at a relatively lower risk level (Hexun 2014).

The rise of internet credit services in China

One major problem in the capital market in China is the lack of sufficient credit services. Due to the lack of appropriate assessment of borrowers' credit risk, banks cannot provide service to them. In private lending or P2P lending, lenders need to devote much effort to the investigation of borrowers' credit. Otherwise, their investments are of high risk. The bad debt problem in private lending and P2P lending caused severe problems in China (Knowledge@Wharton 2011).

In China, there are only two public credit services bureaus: Credit Reference Center of the People's Bank of China (CRC), and its owned Shanghai Credit Information Services Co. Ltd. (SCIS). As of October 2014, they collected registrations for 19.63 million enterprises and identity information for 0.85 billion individuals in China. However, most of such information is not associated with credit-related activities, such as historical bank loan transactions, which are necessary for traditional credit rating models. In the CRC system, there are only about 0.3 billion people with historical financial transactions. Such limited coverage is one major reason hindering the development of China's capital market.

Noticing the lack of credit services in China, in early 2013, the People's Bank of China released the "Credit Industry Management Regulations" and allowed private credit service providers begin operating. Since then, over 70 private credit services were established, such as ZhiMa credit (Alibaba), Tencent credit, ZhongChengXin credit, and PengYuan credit. These companies can be classified into two types. The first type is established by e-commerce companies, represented by ZhiMa credit established by Alibaba (and 100Credit to be discussed in this paper). ZhiMa credit integrates data of hundreds of thousands of Taobao sellers' reputations and transactional activities on Taobao, Tmall, and Alipay platforms. The data was used to provide loans and account receivable mortgages (Abaogao 2014). The second type is established by financial institutions. For example, SCIS set up a credit system to collect company information, loan application information, loan repayment information, and special transaction information scattered on the Internet to build an integrated company profile for credit analysis (Huang 2014). The China PingAn Group also integrates online loan data, bank loan records, and insurance violations information to set up a financial intermediary (Huang 2014; Wang and Li 2013).

As we can observe, the new credit service providers generally take a similar approach of collecting information from the Internet to build new credit models. On particular reason for this choice is the rapid development of Internet industry in China. As compared with the financial industry, China's Internet industry is more advanced and

closer to the level of developed countries. Chinese people now have a high adoption of e-commerce, social networking, and other Internet services, which enables the collection of a large amount of online activity data about them. The use of Big Data is a unique characteristic of Internet credit services.

Literature review

Credit-related Information as economic goods

The major responsibility of credit service providers is to collect, analyze, and provide credit-related information. From a lender's perspective, credit-related information helps reduce their uncertainty on borrowers' future behavior. By differentiating lenders who may and may not repay their debts, lenders may choose their customers to increase expected utility. In previous research, using distance between lenders and borrowers to approximate the ability to get credit-related information on borrowers, Hauswald and Marquez (2006) and Agarwal and Hauswald (2010) found that credit-related information helped credit decisions and could help banks get customers from competitors. Credit-related information is of economic value and is a subject of studies on information economics (Stigler 1961).

In a lending market with multiple lenders and multiple borrowers, information asymmetry (Akerlof 1970) is a common problem. Borrowers typically have more accurate information than lenders about their willingness and ability to repay a loan. Information asymmetry will not only affect lenders' actions but also the entire market's efficiency. Since the expected gains from the loan contract are a function of both the pricing and the probability of repayment, when lenders can't distinguish good borrowers from bad borrowers, all borrowers would be charged the same interest rate that reflects their pooled experience. Such a scenario will cause the adverse selection problem (good borrowers drop out of the market due to the high interest rate) and the moral hazard problem (borrowers have incentive to default since there is no future consequence) (Stiglitz and Weiss 1981).

The need to reduce information asymmetry has significant consequences for the operation of credit markets. Credit services are established to serve the market with credit-related information (Barron and Staten 2003). Kohli and Grover (2008) argued that the ability to generate value from information was an important part of their business. The major purpose of credit services is to generate value from the information on borrowers in the credit market. This is a perfect illustration of the production of valuable information.

Information sharing in the credit market

Information exchange between lenders can occur voluntarily via private credit services or be enforced by regulation via public credit bureaus (Brown et al. 2009). At the individual borrower level, the direct impact of credit service is the improved prediction accuracy on defaults. Kallberg and Udell (2003) investigated the value added by private information exchanges and found exchange-generated information provided significant explanatory power in predicting repayment failure when controlling for other credit information available to lenders. Chandler (1992) found that using U.S. credit bureau data could outperform application data in predicting risk of credit card applications.

At the market level, information sharing through credit services may overcome adverse selection and increases the volume of lending in the credit market (Pagano and Jappelli 1993). McIntosh and Wydick (2005) showed that the existence of a credit bureau might lower lender costs through lower default rates, which improved credit access for poor borrowers. As empirical evidence, Jappelli and Pagano (2002) found that bank lending was higher and credit risk was lower in countries where lenders shared information, regardless of the private or public nature of the information sharing mechanism. Djankov et al. (2007) confirmed that private sector credit was positively correlated with information sharing using 129 countries' data.

Furthermore, information sharing through credit services can alleviate the moral hazard problem. Klein (1992) showed that information sharing could motivate borrowers to repay loans. Brown and Zehnder (2007) conducted a lab experiment and found that information sharing increased repayment rates, as borrowers anticipate that a good credit record improves their access to credit. Padilla and Pagano (2000) argued that sharing default cases among lenders had a disciplinary effect on borrowers, who will consider a default's impact on their credit rating. However, sharing more detailed information about borrowers can reduce this disciplinary effect.

Traditionally, most information sharing between lenders contains negative information, which helps lenders to avoid risky borrowers. It is found that sharing positive information also helps mitigate borrower indebtedness, lower default rates, and reduce interest rates (Luoto et al. 2007). Vercammen (1995) argued that a certain level of adverse selection was beneficial to a credit market and limiting the length of borrower history might actually give borrowers reputation formation incentives.

From a lender perspective, Pagano and Jappelli (1993) argued that incentives for lenders to share information about borrowers are positively related to the mobility and heterogeneity of borrowers, the size of the credit market, and advances in information technology. The factor that discourages sharing of information about borrowers is the fear of competition from additional entrants.

From IT to finance: a brief history of 100Credit

100Credit (Beijing) Financial Information Service Co., Ltd. (100Credit in short), founded in March 2014, is an information service provider specializing in targeted marketing, credit risk analysis, and post-loan service for financial institutions. 100Credit uses Big Data technologies to solve financial institutions' problems in marketing and risk control, and aims to reduce their operation risk and improve their profitability.

100Credit is a spin-off of Beijing BaiFenDian Information Technology Inc. (BaiFenDian in short).¹ BaiFenDian is an IT company that focused on personalization and recommendation for e-commerce applications when it was established in 2009. BaiFenDian takes a platform strategy in providing recommendation as a service to e-commerce companies. By embedding APIs into their client companies' Websites, BaiFenDian can tailor the items to recommend to the clients' customers who are on their Websites. BaiFenDian's recommendation engine was based on users' multi-aspect information, including purchase, browsing, and click-through history, from multiple channels, including web pages, mobile apps, WeChat, and email. In 2012, BaiFenDian launched an "Analysis Engine" to facilitate enterprises' managerial decisions based on web access, product, and

transactional information. With the advanced data analysis technologies, BaiFenDian became a successful recommendation service provider. It has been adopted by companies in the industries of e-commerce, social networking, media, education, travel/transportation, etc., such as FANCL, 1HaoDian.com, JUMEI.com, YINTAI.com, and CBNweek.com. By partnering with these companies, BaiFenDian accumulates one of the largest e-commerce activity datasets in China.

Noticing the need for credit services in China, 100Credit was established to deliver Big Data-based credit rating. With the First Pot of “Data” from BaiFenDian, 100Credit further enriched and extended the data records significantly by patterning with financial institutions and other companies. They built business intelligence models to deal with the challenges of the financial industry. Within one year, 100Credit has partnered with many financial institutions, including China Merchants Bank, China EverBright Bank, and RenRenDai, in collecting financial data and building credit models. At the end of 2015, It is now being evaluated by more than 250 financial institutions (which include most of the banks, P2P lending companies, and personal loan companies in China) for potential collaborations. 100Credit has shown its initial potential in the Internet credit service market.

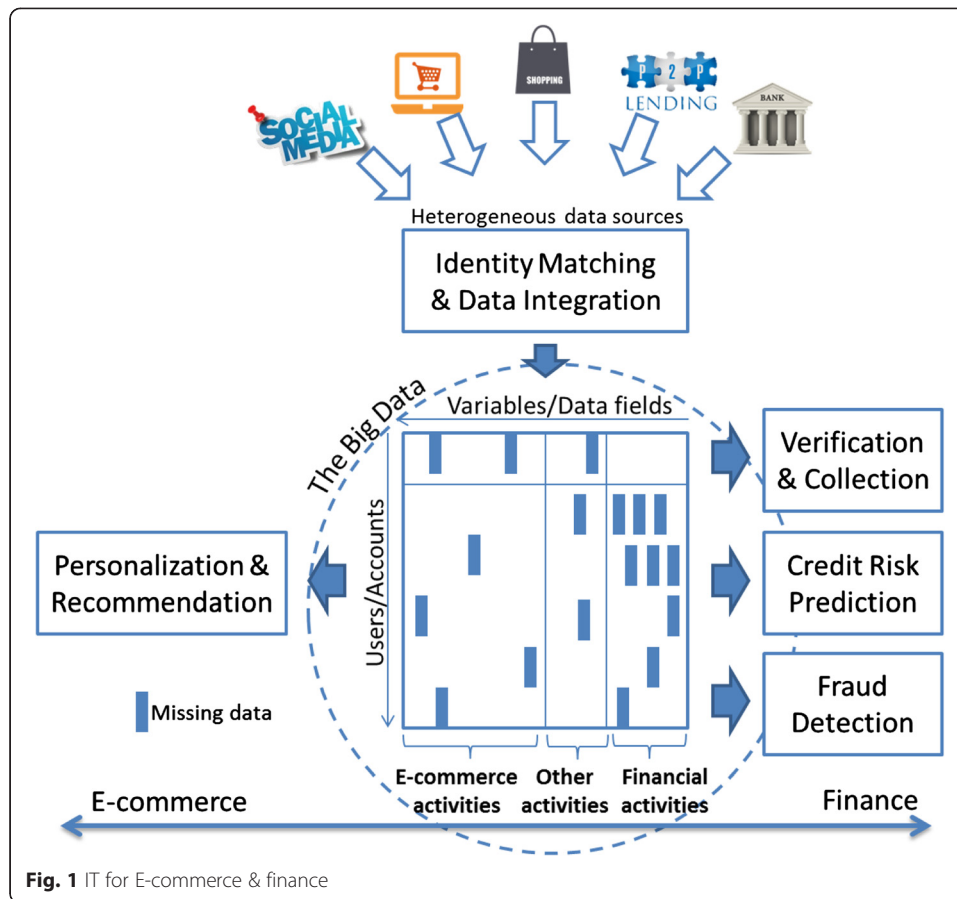
The value of big data to internet credit services

100Credit provides credit rating service for both individuals and small enterprises. For small enterprises, 100Credit collects and analyzes business data online, such as transactions, illegal activities, administrative penalties, and customer complaints, to build credit model. For individuals, 100Credit collects and analyzes individuals’ online behaviors in e-commerce, online reading, and social networking websites, and offline activities to build the model. For ease of discussion and without loss of generality, discussion in this paper is from an individual credit rating perspective.

As discussed, BaiFenDian was an e-commerce service. E-commerce and Finance are quite different applications. They have different target customer companies with different requirements. Nevertheless, from an IT perspective, the two applications share great methodological commonalities, which made it possible for 100Credit to get into the financial service field.

Figure 1 illustrates the general approach of 100Credit/BaiFenDian’s operations. 100Credit’s data is collected and accumulated from multiple sources over a long time. The first step of operating such heterogeneous data is to conduct identity matching (Wang et al. 2011) and compile a (as-comprehensive-as-possible) user profile dataset. Since the data sources have different coverage of consumers, the combined dataset has significant missing values. Based on the combine dataset, from an e-commerce perspective, one can predict the preference of each consumer given his/her previous activities. From a finance perspective, one can predict the level of credit risk for each user (Yu et al. 2015), on which financial institutions can make loan or credit card approval decisions.

Although the two applications share methodological commonalities, the unique context of financial service shows the unique values of Big Data. In this paper, we focus on four aspects: data volume, prediction correctness, service variety, and information protection.



Value comes from data volume

The first and direct impact of Big Data to credit service is the ability to assess more borrowers, especially those without sound financial backgrounds. As elaborated in the background part, the two public credit bureaus in China only have 0.3 billion peoples' financial records. For other people, they at most have identity and demographic information (such as ID, name, age, marriage status, and education level), and it is not plausible to get reliable credit risk predictions using traditional models. This situation significantly limits financial institutions from approaching new consumers.

The advantage of 100Credit and other Internet companies is that they can accumulate a large volume of data on borrowers. For example, BaiFenDian occupies 90 % of the market share in the third party recommendation field. It has about 0.3 billion page views and 82 million unique views every day. After 2013, another spin-off of BaiFenDian, XinBai, get into offline recommendation area, such as in offline retailing, airlines, and telecom industries. Such commerce-related data could be related to people's financial activities and have prediction power on their credit rating.

Starting from BaiFenDian's data records, 100Credit teams with its business partners and expands their data collection to 0.6 billion real-name and 1.1 billion anonymous users, including their online and offline browsing, consumption, reading, social networking, travel, location/real estate information, and blacklist data. Such information primarily comes from the following seven types of sources.

1. Self-reported information by individuals and companies on themselves.
2. 1500 websites' online user browsing/consumption activities. The websites are BaiFenDian's customers, including 700 major brands and retailers and 800 Internet media and web forums.
3. The offline transactional activities of brand partners who have online to offline services.
4. Financial data provided by financial institutions, such as banks, insurance companies, security brokers, mutual funds, small loan companies, and Internet finance companies.
5. Data from third party strategic partners.
6. Public data from government agencies, such as law enforcement records.
7. Public data crawled from the Internet.

Table 1 provides some examples of 100Credit's partners on data. Such data records are a gold mine for financial institutions. Even the information on anonymous users is useful. Anonymous users can be identified by the device they use. If they later request financial service online, the device identity can be matched with the dataset and provide a certain level of credit assessment.

In fact, partially due to the economic value of 100Credit's large dataset, 100Credit was "invited" to the financial industry. In 2013, BaiFenDian was approached by the credit center of China CITIC Bank for targeted marketing of credit card consumers. The purpose was to improve the quality of Internet-based promotion of credit cards to improve credit line assessment accuracy and reduce bad debt. The two companies employed data on e-commerce consumption, high-end product consumption, online reading, house relocation, and online social networking activities for the task. They successfully expanded the promotion to consumers interested in gaming, sports, fashion, and travel, etc., with their specifically designed credit card products. The project was a good example of the high value of Big Data in approaching potential consumers of an Internet credit service.

Value comes from prediction correctness

A major purpose of bringing credit service to the lending market is to improve credit assessment accuracy. In traditional credit services, Kallberg and Udell (2003) and

Table 1 Examples of 100Credit's partners on data collaboration related to individual credit rating

Category		Note	Example sources
Demographics		Age, Gender, etc.	
Activities/Behaviors	Online Consumption	PC client	1HaoDian; MaiBaoBao; HaoLeMai;
		Mobile client	ZhiWo; JiuXian
	Offline Consumption		360 GroupBuy; WangFuJing; InTime;
	Online Reading	Web Forum	SMTH; XICI
		Media	YICAI; CNBeta
	Finance	Banking	China Merchants Bank, China Everbright Bank
		Insurance	China Pacific Insurance; Xinhua Insurance
		Security Broker	China Asset Management; Essence Securities
		P2P	RenRenDai; JieYue; HengChang
	Online Gaming		QQ Game
Social Networking		Weibo; QQ	

Chandler (1992) showed that credit reports provided additional explanation power, compared to data in application forms, for predicting borrowers' credit risk. For 100Credit, it is necessary to address the same problem and provide accurate credit rating service. Otherwise, it cannot survive in the credit rating market.

As aforementioned, 100Credit has a large dataset from multiple sources. The heterogeneous data sources lead to the high dimension (thousands) of variables. Modeling such a dataset is different from traditional financial models, which generally have a handful of variables. Furthermore, borrowers' online activities generally only have a weak correlation with credit risk. Without appropriate modeling techniques, the high volume and high dimensional data is a burden rather than an advantage. Nevertheless, the advances in information technology are an important factor in credit services (Pagano and Jappelli 1993). With advances in data mining (Yu et al. 2015), parallel computing, and Big Data techniques (Goes 2014; Agarwal and Dhar 2014), it is possible to generate valuable information from the raw data from heterogeneous sources. In fact, the ability to analyze data is considered a critical capability for contemporary organizations (Davenport 2006), and the ability to generate value from information is one important function of IT (Kohli and Grover 2008).

100Credit's credit scoring model is trained based on default cases and creditable borrowers provided from multiple different financial institutions. These training data are for multiple different market segments and together cover a large and varied customer base. 100Credit's credit model is mainly based on borrowers' activities and is built using machine learning and Big Data technologies. They first employ a variety of data mining models to preprocess the original features (that have a low prediction power) to create strong features. Then they combine the strong features into the final credit model. They employ algorithms such as Naive Bayes, Nearest Neighbor, Support Vector Machines (SVM), Stochastic Gradient Descent (SGD), Decision Tree, Random Forest, Neural Network (NN), and Logistic Regression to classify the purpose of e-commerce transactions, the interests of online readers, the targets of travelers, etc. They employ K-Means, MiniBatch K-Means, Spectral Clustering, and Gaussian Mixture Models (GMM) to cluster borrowers and their items of interest into groups for ease of modeling. They employ Principal Component Analysis (PCA) and Factor Analysis for dimension reduction. They also employ LASSO (Least Absolute Shrinkage and Selection Operator) regression, ElasticNet, and ARMA to predict numerical values. Through these operations, they address the problem of handling a large number of borrowers and deliver the credit ratings.

The technique approach 100Credit took allowed them to deliver high-quality credit rating predictions. 100Credit tested their model with three commercial banks and one P2P financial company in China. Testing on 1 million users of the three banks showed that 100Credit's model can reduce the Nonperforming Loan (NPL) ratio by about 30 % to 50 % as compared with consumers solicited from offline channels. (The customers solicited from offline channels are generally of higher quality, since banks carefully choose the location to conduct offline promotions. As compared with customers attracted through online channels, 100Credit's model leads to about 70 % NPL ratio reduction, as shown in testing by one of the banks.) Testing on a leading P2P lending company (who generally targets less creditable consumers) shows that the NPL ratio reduction is about 50 % to 70 %.

In general, the ability to improving credit rating credibility comes from the variety and velocity of Big Data. The different dimensions of the dataset allow identification of subtle correlations (rather than causality) between borrowers' daily behaviors and their financial behaviors. For example, examining credit card spending, 100Credit found that the following customers would spend over 30 % more than ordinary customers:

- Consumers who read real estate information or look for apartments online
- Consumers who read fashion information or purchase fashion products online
- Consumers who read food information or purchase food online
- Consumers who read travel information or purchase travel products online

In terms of credit risk, they found that for individual users in small cities:

- The more spending on online gaming and entertainment activities, the higher the credit risk;
- The more spending on education and/or scientific activities, the lower the credit risk.

Such interesting correlations cannot be identified using traditional financial data.

Prediction correctness also comes from the velocity of Big Data. Since Internet companies can capture the change of borrowers' online activities in an efficient manner, they are able to identify changing behavior that can best describe "current" rather than "past" credit risk. For example, one of 100Credit's customer companies works on small loans through mobile clients, mainly for borrowers in small cities. Knowing the recent activities of borrowers allows for identification of multiple parallel loans or loan attempts in multiple financial institutions and helps rule out such high-risk applicants.

Value comes from service variety

The richness of the dataset (large volume of users and large dimensions of variables) also makes it possible to provide a variety of credit products to fit the needs of different market segments. It is a unique advantage of Internet credit service providers as compared with traditional credit bureaus.

The core of a credit service is credit assessment. Due to the availability of data from multiple sources, 100Credit is able to provide multiple dimensions of assessments of credit rating as suggested by (Lin et al. 2015). In 100Credit, such assessments include aspects on daily consumption, online reading, credit card expense, social relationship, flight travel, and location/contact update. The credit score can be selected from different dimensions according to the needs of different financial products.

The ability to connect data from multiple sources into one dataset, i.e., identity matching, also makes it possible to provide two value-added services: information verification and collection support service. Information verification is detecting intentional and unintentional mistakes when users input information, such as address or phone number, in loan applications. It can reduce mistakes in the financial institution's records. Collection support service can help lenders reach borrowers if they lose contact, such as when a borrower changes his/her phone number. With the multiple sources of data, it is possible to identify the borrowers' latest address or contact information for collection.

Another service related to risk assessment is fraud detection. The illegal use and abuse of borrowers' information to get financial products is certainly different from the regular credit application scenario. By using heterogeneous data, an individual's behaviors on different platforms can be jointly considered in fraud detection. Note that it is possible for individuals to behave differently in one/some dimensions of their life, such as in financial activities, but it is unlikely and difficult for a person to fake all aspects of his/her online life. So using multiple dimensions of data may lead to a more reliable assessment. For example, many financial institutions conduct credit risk assessment using point-of-sales (POS) terminal transactions, which infer revenue. Therefore, some dishonest enterprises make fake credit card transactions through POS to achieve a higher credit rating. However, if POS data and store owners' personal online activities are combined, 100Credit found that it was easier to identify fraud activities and reduce credit risk. 100Credit's fraud detection approach includes items such as blacklist checking, credit application analysis (on the pattern of making credit applications), and offline activity assessment (such as whether the device, location, etc. are changed frequently).

Another service provided by 100Credit is financial product recommendation. Promoting financial products is common in the financial industry. At 100Credit, financial product recommendation not only considers user interest (as in traditional recommendation tasks), but also considers the risk level of users. As aforementioned, China CITIC Bank's collaboration with BaiFenDian started with credit card recommendation. 100Credit's financial product recommendation service is now used by equity brokers to recommend stocks to customers.

Table 2 provides a summary of the services provided by 100Credit. Particularly, we align such services with the lifecycle of the lender-borrower relationship. When a user first accesses a channel control by a financial institution, the financial institution can recommend a product to the user. When the user initializes the application, the information verification functions can help reduce mistakes in application forms. When the application is submitted, 100Credit can judge the risk of the user and also detect financial fraud, such as multiple trials of applications. After credit approval. If bad debt appears, the collection support service can improve the chance for financial institutions to reach users. All these functions in the credit service lifecycle are supported and enabled by the heterogeneous data collected from different sources on people's online life.

Like many Internet companies, 100Credit's service are delivered through both API and online systems. For big companies, which often have their own management systems, the requested credit assessment information can be delivered through API. For small enterprises, such services can be delivered in a Software-as-a-Service (SaaS) form. The online platform allows enterprise users to adapt and define their own rules

Table 2 100Credit's services during the lender-borrower relationship lifecycle

Time	Service	Related technology
Before Application	Financial product recommendation	Recommendation
During Application	Information verification	Identity matching
During Approval	Fraud detection	Anomaly detection
During Approval	Credit assessment	Credit modeling
After Approval	Collection support service	Identity matching

for credit assessment and fraud detection. Figure 2 provides a screen shot of the platform.

Value comes from information protection

One problem related to information goods is that they are difficult to produce but easy to reproduce. Theoretically, the equilibrium selling price of credit-related information is zero, since the recipients could secretly resell the information at a lower price (Birchler and Butler 2007). In the case of Internet credit service, as shown in Fig. 1, there are two types of information produced: the identity matching of data between multiple sources and the credit rating generated from the credit models. Internet credit services certainly shall protect their credit rating information, but such information will eventually be sold to the lenders. One cannot stop the lenders from saving such information for future use. Nevertheless, in the short term, lenders may not sell credit ratings directly to their competitors, who may steal borrowers from them. In the long term, since the borrowers' ratings can change with the enriching and updating of users' online activities, to keep and resell outdated credit ratings is of less value and won't be a major concern.²

Protecting data that are matched with identities is more challenging. Such data (either accumulated from the credit service's operation or acquired from other companies) is of value to other Internet companies. There is a strong incentive for people getting such data to improve their own models (for crediting rating, recommendation, etc.). Besides, Internet credit service company's data comes from multiple sources, including user online behaviors. Using such data raises privacy concerns (Belanger and Crossler 2011).³ After identity matching, the data in Internet credit services can connect multiple aspects of a person, which may reveal deep secrets and lead to strong opposition from users.

Information protection is also of particular value to Internet credit services. First, identity matching results have commercial value. For example, 100Credit provides

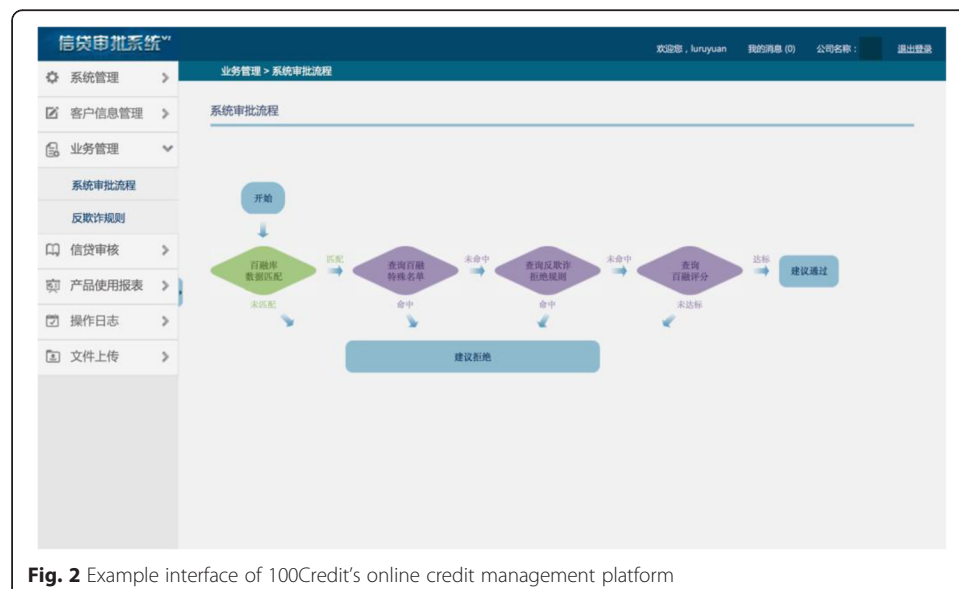


Fig. 2 Example interface of 100Credit's online credit management platform

collection support service to help financial institutions contact borrowers. Financial institutions may have outdated contact information, which can be updated through identity matching. If the borrowers contact information is sold directly to lenders, they can resell the information even to their competitors (and contact information is not as dynamic as rating information). To avoid such deeds reducing the value of the identity matching information produced by 100Credit, they provide collection support as a service rather than selling borrowers' recent contact information.

Second, as a credit service provider, information protection affects the reliability of the credit model. If the features used for credit ratings are revealed, one would expect users to change their online behaviors to fool the rating system. For example, in the traditional model of using POS activities to determine a merchant's credit level, some merchants deliberately fake credit card transactions to achieve a higher credit level. Such an operation is more difficult in the Big Data era, since the credit model deals with many dimensions of variables. Nevertheless, revealing enough details on the model-related data will cause the old model to be void and require more efforts to develop new models.

Due to such concerns, information protection is related to the value that Big Data can bring to Internet credit services. 100Credit expends a great deal of effort to protect its data. From a network security perspective, it follows Chinese government regulations to obtain security certifications from the Ministry of Public Security. It also follows the international standard of ISO27000. 100Credit's data is stored in servers that are disconnected from the Internet and cannot be physically accessed by employees. The data processing program has to be executed on a computer terminal that does not have data storage capabilities. Thus, employees cannot export the company's data from the servers.

At the application level, to further protect the identity matching results and individual privacy, 100Credit anonymizes their data. When receiving any data from external sources, the first step is to convert Personal Identifiable Information, such as ID, phone number, email, etc., which can be used to identify individuals in the real world, to a random string. The random string is used internally as the user's identifier. This step is conducted automatically, so that no employees have access to the identity information. Necessary access to personal identifiable information needs internal approval from the company's data committee. After approval, the access process is monitored by three (rotating) members of the committee with three authentication codes for accessing the data.

Even after removing the personal identifiable information, there is a possibility that one can figure out the identity of a person by connecting different evidence. Therefore, 100Credit segments their data and employees only see data attributes related to their work.

Through such operations, 100Credit enforces the protection of its data, which ensures the sustainability of its business from a value of information perspective.

Overall, as summarized in Table 3, the value of Big Data is shown in four aspects in Internet credit services. First, the large data volume makes it possible to assess more users' credit levels, who can be potential borrowers. Second, the improved prediction correctness caused by the variety and velocity of Big Data makes it possible for financial institutions to make a profit on consumers who were considered less profitable. Thirds, the variety of services enabled by Big Data allows catering to different needs of financial

Table 3 A summary of the values of big data to internet credit service

Dimensions	Effects
Data Volume	Reaching more consumers for financial institutions
Prediction Correctness	Identify the profitable customers for financial institutions
Service Variety	Catering to different types of needs of financial institutions
Information Protection	Ensuring the sustainability of the credit service

institutions and is one competitive advantage of Internet credit services. Fourth, it is necessary for Internet credit services to protect their data as an asset and protect the users' privacy to sustain their services.

Discussion and implications for future research

Past: the unique opportunity of China

Aligned with the four values of Big Data we argued above, China provides a unique opportunity for Internet credit services. As compared with developed countries, the credit service in China is relatively amateur. At the same time, China has a rapid developing economy that desperately needs credit services. There is a strong need to reach a larger group of consumers in a short period of time, which makes the "value of volume" meaningful and significant.

In the past 20 years, China has produced a generation of well-educated IT workers in the Internet industry. This makes it possible to develop and implement advanced models to take advantage of the value of prediction correction and provide a variety of services.

Moreover, China's privacy protection laws are not as strict as in Western countries. This is a challenge but also provides opportunities for Internet credit service companies. It allows them to collect the data needed for their service. In addition, their work can contribute to society by suggesting suitable privacy protection guidelines.

Future: the first mile of credit history

According to our argued four values of Big Data, we would project that Internet credit services will play a supporting role to regular credit services in future. With the development of the economy and the financial industry in China, traditional credit services will get an enlarged set of consumer records. The value of data volume will be reduced and be mainly reflected in "new" consumers to the financial industry, such as borrowers who are young, have a short time on the job or at their residence, and lower income teenagers. In fact, such financially more vulnerable (higher risk categories) borrowers were considered to benefit the most from credit services in previous literature (Barron and Staten 2003). For such consumers, Internet credit services will continue to play a very important role, since it is easier to build a record on the Internet (such as in online gaming, social networking, online reviews) than with financial services. Finance institutes will highly rely on Big Data to build the first mile of their consumers' credit history.

We do not think Internet credit services will eventually replace traditional credit services. As we know, the most important data in credit services is still financial

information (including both cases of bad debts and financial activities that can predict such bad debts). Such data will be protected by the financial industry due to its high value and privacy concerns. Thus, holding such an important data source provides value and a reason for traditional credit services' existence, mainly to serve consumers with a longer financial history.

Due to privacy concerns, we do not think Internet credit services will eventually merge with traditional credit services. Note that online activities and financial histories represent different types of confidential information. Merging them into one organization will create major public concern. Furthermore, according to (Padilla and Pagano 2000), sharing default cases among lenders has a disciplinary effect on borrowers, which will be reduced if other information is also shared. Allowing credit bureaus to hold some powerful financial information will help maintain the disciplinary effect in the finance domain. We do not want to extend the disciplinary effect to all dimensions of online activities. Two types of credit services serving relatively different market segments may better benefit society.

Future research directions

To improve the data analytics framework

As an important future direction, everyone would expect that improving the data analytical framework will increase credit prediction correctness. The significant amount of missing data creates a major challenge (for this task). Since the data are collected from different websites/applications, which have different target users, it is impossible to have all data fields for all users. It is possible to address some problems using classic methods, such as replacing the missing value with zero or with the mean value. Nevertheless, we think a new framework is the best way to address this significant problem. Other aspects of the data analysis model can also be improved.

To quantify the value from each dimension

In addition to improving the credit model, it is also possible to improve on other aspects of an Internet credit service. From the management perspective, it is worthwhile to understand how improving service level at different dimensions may affect the profit (or utility) of the company. With this knowledge, the Internet credit service company can better allocate their resources on data collection, model development, service improvement, and information protection. More research on this aspect is needed to direct their strategic plan.

To examine regulations on internet credit services

At a country level, it is necessary to study how policies on Internet credit services may affect their operations and the entire economy. For example, it may be subtle to balance the pros and cons of Internet credit services due to the privacy concern. Making regulations too strict may kill Internet credit service companies. Making regulations too loose may affect the individual benefits. Furthermore, it is necessary to closely monitor the credit markets to avoid illegal competitions. To answer such questions needs further studies from the policy perspective.

Conclusion

In this paper, we inspect the business operations of 100Credit, an Internet credit service company in China. Internet credit service providers make use of heterogeneous data collected from multiple sources on people's online and offline activities to predict their credit rating. We summarize four values Big Data has brought to Internet credit service companies through 100Credit's practice, namely, the value from volume, prediction correctness, service variety, and privacy protection. We discuss the unique opportunity that makes Internet credit service possible in China and the future of this service.

Internet finance is a promising field, of which credit service is one important pillar. Understanding the issues related to Internet credit services not only enriches our knowledge to this new phenomenon but also holds the potential of addressing practical problems. Big Data plays an essential role in many Internet credit service companies' business. We expect the case study provides some preliminary evidence and opinions that can lead to future research on Internet credit services.

This study is not immune from its limitations. First, Internet finance and Internet credit service are under rapid development. Their mechanisms may change over time. Our study only provides assessment for the current time point and the finding may change from a longitudinal perspective. Second, our study is based on practices of one company. Although the company is active and playing an important role in the market, it is possible that other companies experienced a different path can also provide interesting insights to this field. Third, our study is based on the information centered around one company. Information from other parties, including employees of partner companies, may help us better understand the value of Big Data to related companies in Internet finance.

In the future, we plan to extend the current study from two aspects. First, we will collect richer data from the company, including business operation data, interviews, and surveys, to study the value of Big Data to people with different roles in an Internet credit service. Second, we will conduct network analysis of 100Credit's business partners to understand the value of Big Data to the external related parties of an Internet credit service.

Endnotes

¹The two companies are now entirely independent in management, finance, and operations.

²There is a concern about stabilized credit rating, which will be discussed in Future: the first mile of credit history.

³The focus of this paper is not on privacy. Nevertheless, to alleviate the privacy concern, 100Credit pursues explicit user authorization when collecting data from users' devices and ensures providing authorized data in response to queries.

Competing interests

The first two authors work for 100Credit. But 100Credit does not provide any financial support to this paper. It is not part of 100Credit's projects or operations.

Authors' contributions

SZ and XL developed the design and the major idea of the paper and wrote key sections of the paper. WX and WN wrote the majority of the paper. All authors read and approved the paper.

Authors' information

SZ is the cofounder and CEO of 100Credit. WX is the strategy director of 100Credit. WN is an associate professor at the institute of automation, Chinese Academy of Sciences. XL is an assistant professor in the Department of Information Systems, at the City University of Hong Kong.

Author details

¹100Credit Financial Information Service Co., Ltd., Beijing 100089, China. ²Institute of Automation, Chinese Academy of Sciences, 95 Zhongguancun E Rd, Beijing 100190, China. ³Department of Information Systems, City University of Hong Kong, Kowloon Tong, Hong Kong.

Received: 26 November 2015 Accepted: 8 December 2015

Published online: 24 December 2015

References

- Abaogao (2014) The Development of Chinese Internet Financial Industry in 2014. <http://www.abaogao.com/c/jinrong/006189RP0Y.html>. Accessed date: 1 Nov 2015
- Agarwal R, Dhar V (2014) Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Inform Syst Res* 25(3):443–448. doi:10.1287/isre.2014.0546
- Agarwal S, Hauswald R (2010) Distance and Private Information in Lending. *Rev Financ Stud* 23(7):2757–2788. doi:10.1093/rfs/hhq001
- Akerlof G (1970) The Market for Lemons. *Quarterly Journal of Economics* 84:488–500
- Barron JM, Staten M (2003) The value of comprehensive credit reports: Lessons from the US experience. In: *Credit Reporting Systems and the International Economy*, MIT Press; pp. 273–310.
- Belanger F, Crossler RE (2011) Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. *Mis Quart* 35(4):1017–1041
- Birchler U, Butler M (2007) *Information Economics*. UK: Routledge.
- Brown M, Zehnder C (2007) Credit reporting, relationship banking, and loan repayment. *J Money Credit Bank* 39(8):1883–1918. doi:10.1111/j.1538-4616.2007.00092.x
- Brown M, Jappelli T, Pagano M (2009) Information sharing and credit: Firm-level evidence from transition countries. *J Financ Intermed* 18(2):151–172. doi:10.1016/j.jfi.2008.04.002
- Chandler GGRWJ (1992) The Benefit to Consumers From Generic Scoring Models Based on Credit Reports. *IMA Journal of Mathematics Applied in Business and Industry* 4:61–72
- China Securities Regulatory Commission (2011) Statistics on December 2010. http://www.csrc.gov.cn/pub/zjhpublic/G00306204/zqscyb/201101/t20110126_191194.htm. Accessed date: 1 Nov 2015
- Davenport TH (2006) Competing on analytics. *Harvard Bus Rev* 84 (1):98
- Djankov S, McLiesh C, Shleifer A (2007) Private credit in 129 countries. *J Financ Econ* 84(2):299–329. doi:10.1016/j.jfineco.2006.03.004
- Dong W (2012) There is a high financial risk on private lending in China. *China Youth Daily*, China
- Goes PB (2014) Big Data and IS Research. *Mis Quart* 38(3):lii–Viii
- Hauswald R, Marquez R (2006) Competition and strategic information acquisition in credit markets. *Rev Financ Stud* 19(3):967–1000. doi:10.1093/rfs/hhj021
- Hexun (2014) The four development stages of P2P lending in China. <http://iof.hexun.com/2014-05-16/164842883.html>. Accessed date: 1 Nov 2015
- Huang X (2007) Opinions on the development of personal loan service in Chinese commercial banking. *China Water Transport* 5(7):164–166
- Huang X (2014) Thinkings on the development of Chinese credit rating business in Internet finance era. *Credit Reference* 5:50–53
- Jappelli T, Pagano M (2002) Information sharing, lending and defaults: Cross-country evidence. *Journal of Banking & Finance* 26(10):2017–2045
- Kallberg JG, Udell GF (2003) The value of private sector business credit information sharing: The US case. *Journal of Banking & Finance* 27(3):449–469. doi:10.1016/S0378-4266(02)00387-4
- Klein DB (1992) Promise Keeping in the Great Society: A Model of Credit Information Sharing. *Economics and Politics* 4(2):117–136
- Knowledge@Wharton (2011) The cause and crisis for the rapid development of underground banks in China. <http://www.knowledgeatwharton.com.cn/article/2911/>. Accessed date: 1 Nov 2015
- Kohli R, Grover V (2008) Business value of IT: An essay on expanding research directions to keep up with the times. *J Assoc Inf Syst* 9(1):23–39
- Lin Z, Whinston AB, Fan S (2015) Harnessing Internet finance with innovative cyber credit management. *Financial Innovation* 1:5
- Luoto J, McIntosh C, Wydick B (2007) Credit information systems in less developed countries: A test with microfinance in Guatemala. *Econ Dev Cult Change* 55 (2):313–334. doi:10.1086/508714
- McIntosh C, Wydick B (2005) Competition and microfinance. *Journal of Development Economics* 78(2):271–298
- Padilla AJ, Pagano M (2000) Sharing default information as a borrower discipline device. *European Economic Review* 44(10):1951–1980
- Pagano M, Jappelli T (1993) Information Sharing in Credit Markets. *Journal of Finance* 48(5):1693–1718
- People's Bank of China (2002) China's financial status is healthy and stable in 2001. <http://www.pbc.gov.cn/diaochatongjisi/116219/116225/2821519/index.html>. Accessed date: 1 Nov 2015
- People's Bank of China (2011) Statistics Report on Financial Institution Investments in 2010. <http://www.pbc.gov.cn/diaochatongjisi/116219/116225/768654/index.html>. Accessed date: 1 Nov 2015
- People's Bank of China (2014) Statistics Report on Financial Institution Investments in 2013. <http://www.pbc.gov.cn/diaochatongjisi/116219/116225/745257/index.html>. Accessed date: 1 Nov 2015

- Stigler J (1961) Economics of Information. *Journal of Political Economics* 69:213–225
- Stiglitz JE, Weiss A (1981) Credit Rationing in Markets with Imperfect Information. *Am Econ Rev* 71(3):393–410
- Vercammen JA (1995) Credit Bureau Policy and Sustainable Reputation Effects in Credit Markets. *Economica* 62(248):461–478
- Wang Y (2009) The adjustments of loan structure on personal business in China commercial banks. *Commercial Times* 3:64–66
- Wang X, Li S (2013) Internet Finance Promotes the Development of Credit Rating Business. *China Finance* 24:60–62
- Wang GA, Atabakhsh H, Chen HC (2011) A hierarchical Naive Bayes model for approximate identity matching. *Decis Support Syst* 51(3):413–423. doi:10.1016/j.dss.2011.01.007
- Yu L, Li X, Tang L, Zhang Z, Kou G (2015) Social credit: a comprehensive literature review. *Financial Innovation* 1:6

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
