Reflection on Big Data Technology: Problems and Countermeasures in "Big Data Credit Reporting" of Internet Finance in China

Cheng-yong LIU*
Beijing Institute of Technology, Zhuhai, NO.6 Jingfeng Road, Tangjiawan, Zhuhai, Guangdong, 519088 P.R. China
*Corresponding author

Keywords: Big data, Big data credit reporting, Internet finance, China.

Abstract. With the rapid development of Internet finance in China for the past few years, big data credit reporting agencies specifically for network credit information have been initially established. An analysis on the technology application, characteristics and operational difficulties of big data credit reporting shows that big data credit reporting in China still has many problems such as inefficiencies in privacy protection, illegal operation, credit reporting standard difference and credit rating model, etc. during its development. Therefore, the above problems can only be resolved if the future development of big data credit agencies in China will be standardized on a risk prevention basis without limiting the market creativity.

Introduction

With the development of economic globalization and information technology, financial markets are constantly generating mass data, and big data is needed by financial practitioners to support their financial decisions. Big data technology therefore is deeply integrated with the credit reporting industry to develop a new credit reporting channel for evaluating the credit reporting system through data acquisition, mining, analysis and other Internet technologies, and big data reporting agencies specifically for network credit information, such as “Sesame Credit”, are accordingly developed [1,2]. "Sesame Credit" objectively presents personal credit status through cloud computing and machine learning, etc. and provides credit services for users and merchants in hundreds of various industries including credit card, consumer finance, finance leasing, hotel, rental, travel, marriage, classified information, student service and public utility service, etc. Therefore, big data credit reporting has gradually become an important means to accelerate the development of Internet consumer financial industry in China.

Technology Application and Characteristics of Big Data Credit Reporting

Application of Big Data Technology

With the in-depth development of big data technology in China's Internet financial industry, the Internet consumer financial platform has started to use the big data technology in the credit reporting field [3,4], which involves the following aspects:

(1) Big data acquisition. In addition to deep mining of credit record data in the traditional credit reporting system, other data affecting the customer's credit status, such as occupation, marriage, social relations, purchase and payment habits, etc., is also taken into account to understand the real situation of the customer's environment and social network and find clues related to the lending behavior of the customer, providing a sufficient basis for credit decision.

(2) Big data storage. In order to store mass unstructured data generated based on customer behaviors, the following three storage solutions can be provided with the big data technology: MPP relational database, Hadoop non-relational database and big data appliance composed of hardware and software.
Big data extraction and cleaning. Data that can have an impact or a significant impact on the credit status of the customer should be extracted from mass structured and unstructured data, and then messy data is merged, categorized and sorted with the data cleaning technology for data analysis.

Big data mining. Data mining algorithm models are established through the mathematical and system science to analyze and explore the value contained in the data. Many algorithms used in the field of data mining, such as relational query, classification and neural network, etc., are iterative type.

Characteristics of Big Data Credit Reporting

The basic function of the traditional credit reporting is to process the data, that is, collecting, processing, analyzing, disseminating and using the data to evaluate and predict the behavior of the credit subject, and its roles are to relieve information asymmetry between the transacting parties, reduce credit risk and improve transaction efficiency. Although big data credit reporting does not subvert the basic functions of traditional credit reporting, the former is significantly different from the latter due to its technological innovation in terms of data collection, storage, processing, analysis and prediction, etc.

1. In addition to revealing the causal relationship between historical data and behaviors highlighted by the traditional credit reporting, big data is also applied to find the correlation between things and realistically “describe” behaviors, consumption habits, incomes, work and education of users, thus providing a new perspective to determine the personal credit status.

2. Only structured information can be collected and processed with the traditional credit reporting technology. It mainly comes from financial institutions, government departments and public enterprises that have a certain type of credit information, which has less information and few sources. Compared with that, the big data credit reporting technology can be used to obtain a wider range of data, process and analyze the fragmented information and incomplete data flow, especially text, audio, video, email and other unstructured data that cannot be collected and processed with the traditional credit reporting technology, including information based on network search and social behaviors; information on Internet transactions; real-time stock exchange data, etc.

3. The Internet-based big data application technology reduces not only the cost of collecting traditional credit reporting data, but also the cost of collecting new unstructured data. In addition, the development of distributed and cloud computing technologies in the era of big data changes the way that the computing power is improved solely by improving the computing performance of hardware devices, thereby improving the utilization efficiency of devices, reducing the data computing cost, and improving data processing and analysis efficiency.

Problems in China’s Big Data Credit Reporting

With the rapid development of social networks, Internet of Things, distributed storage and computing, and deep neural networks, etc., all data related to individuals can be collected by technical means and become the original data for evaluating personal credit. The application of new technology has brought great changes in the credit reporting industry, but also raised a lot of problems.

Privacy Protection

In addition to traditional credit data, personal Internet data, sensor data, behavior data and geographic location data, etc., can also be converted into personal credit rating by an algorithm model when big data credit reporting is used. In practice, some credit reporting agencies do not make a distinction among personal information, credit information and privacy information in the personal information collection process, resulting in the risk of disclosure and illegal use in personal privacy. Meanwhile, the current Chinese laws, regulations and related judicial interpretation are relatively lagging behind in terms of their protection on personal privacy in the era of big data. The laws involving personal information and privacy protection include Criminal Law, Tort Liability Law, Network Security Law, Consumer Protection Law, Regulation on the Administration of Credit Reporting Industry, Measures for the Administration of Credit Reporting Agencies, Provisions on
Protecting the Personal Information of Telecommunications and Internet Users, Interim Measures for the Administration of the Basic Data of Individual Credit Information, Guidelines for Regulating Credit Reporting Agencies, Information Security Technology – Guidelines for the Protection of Personal Information in Public and Commercial Service Information Systems and their judicial interpretations. Most of them are indirect, fragmented and framework with ambiguous duties and responsibilities of law enforcement agencies, unscientific and unreasonable information collection and processing rules, high law-abiding cost of enterprises, poor judicial remedy channels, and high cost and low efficiency in safeguarding the rights and interests, etc.

Illegal Operation

Firstly, the data, including online consumption, payment, network behavior, psychological test and payment, etc., collected by big data credit reporting agencies in practice has gone beyond the scope of collectable personal credit information specified in Article 2 and 14 of the Regulation on the Administration of Credit Reporting Industry (hereinafter to be referred as “ACRI”). The credit information refers to the information generated during social and economic activities by the data subject, obtained by the credit reporting agency from the credit information provider, and provided to the information user for determining the credit level of the data subject after analyzed and processed. Based on the absolute technical advantages of credit reporting agencies, personal information may even be collected without limit.

Secondly, as credit reporting agencies rarely provide a complete list of personal information collected and used, the information subject is unable to know whether the personal sensitive data collected and used contains any personal information prohibited by the law in terms of its collection, nor to confirm whether the personal information collected and used is true and accurate, and nor to exercise the right of informed consent granted under Article 13 of the ACRI.

Thirdly, big data credit agencies will output credit rating results to different scenarios, such as inquiry of credit scores based on travel and visa, etc. This violates Article 19 and 20 of the ACRI, which stipulate that personal information inquiry should not be used without the consent of the information subject.

Standard Difference

Due to large volume, low value density and fast update, etc. of the Internet big data, the standards of data type, data format and relevance with personal credit vary among data credit reporting agencies during data collection. A lack of uniform standards for collection of credit reporting data leads to high noise and disorganization of online data, and difficulty in effective verification of its authenticity and accuracy, which conflicts with the information accuracy guarantee required by laws, is detrimental to the consistency and authority of credit rating results and difficult to reflect the real credit status of the information subject.

Rating Model

The credit rating model and method are one of the most important core competencies of big data credit reporting agencies, and its accuracy, stability and interpretability are the main reference factors to measure the personal credit rating model of big data. However, the effectiveness of big data credit rating models in China remains to be verified for the following reasons: (1) the data of big data credit reporting agencies mainly comes from network platform transaction data, social data, telecommunication data and behavior data, etc. with high data pollution and data noise, and complex data screening and cleaning, etc., and the credit loan data that correlates highly with personal credit cannot be obtained by them; moreover, the role of social data in personal credit rating cannot be really determined [9]; (2) the credit rating model brings the challenge of complexity to credit risk, and the relationship between various types of data and personal credit still requires a long-term in-depth research and verification.
Countermeasures

The current big data credit reporting in China is still in its infancy compared with the European and American countries. In order to promote the healthy development of the credit reporting industry, strict risk prevention should be carried out on a risk prevention basis without limiting the market creativity.

Firstly, suggestions for privacy protection involved in China's big data credit reporting are as follows:

1. Accelerate the establishment of Administrative Measures on Credit Reporting Business for Enterprises and Individuals, Regulations on Protection of Personal Credit Reporting Information and other supporting systems, seek the balance between healthy development of the credit reporting industry and protection of legitimate rights and interests of the information subject, and standardize the collection and use of personal credit information as well as the rights and obligations of the information subject;

2. Grant consumers with the right to know the credit score through legislation, improve the collection and use standards as well as exchange and sharing mechanism of personal information, such as specifying the specific authorization in the authorization contract of Internet personal information, and protect the user's right to know and choose to avoid excessive collection of personal information, improper use of information or access by third parties without authorization;

3. Guide big data credit reporting agencies to promote secure authentication services such as real-name ID cards, digital certificates and electronic signatures, etc., and conduct identity authorization with the use of synchronous images, personal fingerprints and password security issues, etc. to safeguard data information security and personal privacy and prevent personal information disclosure;

4. Strengthen industry self-regulation, establish and improve the industry sharing mechanism and the disciplinary mechanism on violations, and strengthen risk control on personal information and privacy protection within enterprises.

Secondly, the supervision of big data credit reporting business should be strengthened on the basis of science, effectiveness and enforceability. This paper proposes to formulate the supervision guidelines for the Internet financial credit reporting model as soon as possible, improve supervision scopes, types and procedures of credit reporting products, conduct classified and differentiated regulation on different types of credit reporting products, clarify the responsibilities on information security, privacy protection, objection handling and relief, etc. of the Internet credit reporting agencies including big data credit reporting agencies, and guide big data credit reporting agencies to regulate their own business rules to improve business processes and internal control system.

Thirdly, a scientific and reasonable credit rating system should be established for big data credit reporting. This paper recommends:

1. Establishing sound personal information collection standards, credit report format, personal credit scoring standards, credit reporting database and credit management procedures, etc.;

2. Establishing the interests coordination and sharing mechanism among information subject, information intermediaries, credit reporting agencies and credit report users, etc. to ensure independent, transparent, fair and clear responsibilities and division of labor;

3. Reasonably define the boundary of credit reporting, clarify the scope of personal information that is clearly prohibited for collection and collected upon the permission agree, create a list of personal credit reporting information collected and used, and increase the transparency in terms of collection and use of personal information.

Conclusions

As an important part of the market economy system in China, the credit reporting market faces challenges in the contemporary era of data flow. China should follow the development trend of big data credit reporting, actively respond to new challenges, take the initiative to seize new opportunities and accelerate the healthy development of credit reporting market while seeking the balance between
healthy development of the credit reporting industry and protection of legitimate rights and interests of the information subject.

Reference


[4] Zhao Dawei: Research on Internet Consumer Finance Driven by Big Data Technology, Finance and Economics, issue 1, p. 44.


