Design of Personal Credit Information Sharing Platform Based on Consortium Blockchain

Jing Zhang[1,2](✉), Rong Tan[1,2], and Yu-dong Li[1]

[1] Shanghai Business School, Shanghai 201400, People's Republic of China
zhangjing25@163.com, tanrong529@gmail.com, lydmm_2002@163.com
[2] Tehua Postdoctoral Programme, Beijing 100029, People's Republic of China

Abstract. The technical features of blockchain, including decentralization, data transparency, tamper-proofing, traceability, privacy protection and open-sourcing, make it a suitable technology for solving the information asymmetry problem in personal credit reporting transactions. Appling blockchain technology to credit reporting meets the needs of social credit system construction and may become an important technical direction in the future. This paper analyzed the problems faced by China's personal credit reporting market, designed the framework of personal credit information sharing platform based on blockchain 3.0 architecture, and studied the technical details of the platform and the technical advantages. The in-depth integration of blockchain technology and personal credit reporting helps to realize the safe sharing of credit data and reduce the cost of credit data collection, thereby helping the technological and efficiency transformation of the personal credit reporting industry and promoting the overall development of the social credit system.

Keywords: Consortium blockchain · Personal credit reporting · Credit information sharing

1 Introduction

With the continuous improvement of China's social credit system, personal credit has become an important credit indicator for natural persons in modern economic society. By the end of 2019, China's Central Bank's personal credit reporting system has collected credit information of 1.02 billion natural persons. While playing the leading role of the government, the marketization process of personal credit reporting in China is also accelerating. In May 2018, Baihang Credit was officially established, it mainly focuses on the credit business of Internet finance, which is an important complement to the Central Bank's personal credit reporting system.

The biggest development bottleneck of China's personal credit reporting industry is the integration and sharing of credit information. The Central Bank's personal credit reporting system is only open to enquiries from individuals, commercial banks, and financial institutions that conduct credit services. It is basically closed to the market, and is still far away from sharing data with Baihang Credit. The original intention of the

G. Xu et al. (Eds.): FCS 2020, CCIS 1286, pp. 166–177, 2020.
https://doi.org/10.1007/978-981-15-9739-8_14

establishment of Baihang Credit is to create a melting pot of market credit information. Its shareholders are all industry giants, covering massive Internet user data of group companies such as Ali, Tencent, and Ping An. However, the majority of shareholders did not share credit data with Baihang Credit.

1.1 Credit Data Ownership Confirmation

Data resource has the feature of being easy to copy, its confidentiality is difficult to maintain, and the credit data of the same information subject can always be collected by multiple financial institutions. Credit data is the core interest of credit reporting agencies. In the case where data ownership cannot be clearly defined, sharing data externally may lead to the leakage of commercial information and cause losses to the institution's own interests. In addition, there is currently no clear method for measuring the market value of credit data and the benefits brought by data exchanges, which also leads to the lack of incentives for institutions to share data externally.

At present, the ownership of personal information is still controversial in law. Theoretically speaking, personal information comes from individuals, and the information subject should enjoy priority data ownership. However, the information subject does not have the actual processing, use, sharing and other control capabilities of personal information [1]. The current unspoken rule in the credit reporting industry is that with the authorization of the information subject, the ownership of the personal credit information, the actual control rights, and the benefits of data transactions all belong to the information collector [2]. This mechanism obviously does not give personal information subjects the reasonable protection of the economic interest, but how to price fragmented personal information is a difficult point.

1.2 Credit Data Integration

Credit reporting is part of the financial infrastructure, and the quality of credit data has priority over the size of the data. With the deepening of Internet financial remediation, China has officially promoted the P2P online lending institutions to access the Central Bank's credit reporting system and Baihang Credit. The access of Internet financial data has brought greater challenges to the data quality management of credit bureaus. Unlike traditional financial institutions with a high degree of data normalization, Internet data sources are multi-source, diverse, and multi-domain [3], and the data integration is difficult with high cost. The data access institutions of Baihang Credit cover many fields such as online small loans, P2P platforms, consumer finance, and Internet bank. Their businesses, customer groups, and data formats are different, and their business levels are uneven [4]. Once the business operations of some institutions are not standardized, or the data is not strictly submitted in accordance with uniform standards, or even false data is submitted, the credibility of the entire credit reporting system will be reduced, which will have a great negative impact on credit management.

1.3 Credit Data Security

Data security incidents have occurred frequently in the credit reporting industry. In September 2017, Equifax, one of the three largest personal credit bureaus in the United

States, suffered the most serious cyber-attack since its establishment, resulting in the disclosure of personal key information of approximately 150 million customers. According to the data released by the National Internet Emergency Center (CNCERT) in China, in the year of 2019, a total of more than 3,000 important data breach risks and incidents were discovered in China. Mainstream databases such as MongoDB, Elasticsearch, SQL Server, MySQL, and Redis have been exposed to have serious security vulnerabilities that may lead to the risk of data leakage.

While promoting the integration and sharing of credit data between institutions, it is necessary to technically guarantee the security of the system and data. The credit reporting system should implement strict access mechanism, identity verification and permission control for data access institutions, set up database access security settings, and minimize potential data security risks such as vulnerability utilization, DDoS attacks, and brute force cracking.

1.4 Personal Information Protection

The risk of personal information leakage in China's credit service industry is severe. In 2019, many big-data risk-control companies were investigated and prosecuted for using web crawlers to illegally crawl personal data. Some online loan Apps excessively claim user authorization, and collect personal information beyond the scope, resulting in rampant phenomena such as "routine loans" and violent collection. In the era of big data, personal information protection has become the focus of global attention. Both the EU's General Data Protection Regulation (GDPR) and the US California Consumer Privacy Act (CCPA) have been promulgated, giving detailed descriptions of the rights enjoyed by information subjects and the obligations of data controllers [5]. China's big data industry has entered a period of rectification, and the "Personal Information Protection Law", "Data Security Law" and other related laws will be introduced in the near future, placing stricter requirements on privacy protection in the credit reporting industry.

In order to maintain the standardized development of the credit reporting industry, the credit reporting agencies need to implement strict and complete compliance authorization management. When collecting and using personal information, they must ensure that they have obtained explicit authorization from the information subject and use it within the scope of authorization. At the same time, the credit reporting agencies should strengthen the privacy protection, anonymize personally identifiable information, eliminate the identity attributes and sensitive data in personal information, and fully protect the legitimate rights and interests of the information subject.

1.5 Data Acquisition Cost

The credit reporting agency's own data resources are relatively limited, and usually need to obtain external data from other data service providers through purchase, exchange and cooperation [6]. The construction cost of traditional centralized database is relatively high, the authorization mechanism is complex, and the data security is difficult to guarantee, which is not conducive to mutual trust and collaboration between institutions. There is often a competitive relationship between agencies, so it is inevitable that data service providers will stop providing data or raise the price of data. The fierce data source competition has caused credit reporting agencies to spend a lot of time and economic costs in the data collection process, resulting in waste of resources, which is not conducive to the improvement of the overall efficiency of the credit reporting industry.

2 Related Work

As a disruptive technology, blockchain is leading a new round of technological and industrial changes worldwide. In January 2016, the United Kingdom took the lead in listing blockchain as a national strategy, and believed that blockchain could construct an honest society. China's "Thirteenth Five-Year Plan for National Informatization" raises blockchain as a key frontier technology at the national strategic level. Blockchain has technical characteristics such as distributed storage, point-to-point transmission, quasi-anonymity, security, trustworthiness, and programmable. It is a natural fit for personal credit reporting, and can solve the problem of credit information sharing in a targeted manner.

Many Chinese well-known scholars have recognized the advantages of applying blockchain technology to credit reporting. Liao believed that blockchain technology could be applied to the collection, transmission, processing and inspection of information, bringing new opportunities for the credit reporting industry [7]. Ba pointed out that the characteristics of blockchains such as tamper-proofing, traceability, and privacy protection made it a good way to solve the problems in the field of credit data sharing [8]. Zhu pointed out that the functional attributes of the consortium blockchain such as access mechanism, distributed database, multi-center, smart contract and incentive mechanism are suitable for credit data collection, storage, transmission, verification and supervision [9]. Liu pointed out that blockchain could provide technical architecture support for traditional credit reporting systems and solve the current pain points of the credit reporting system [10].

The application of blockchain technology to credit information sharing is a hot issue in academic research in recent years. Ju et al. designed a big-data credit reporting platform for multi-source heterogeneous data fusion based on blockchain technology and developed a prototype system [11]. Wang et al. proposed a blockchain-based information sharing and secure multi-party computing framework, and designed the detailed storage model, the consensus algorithm and the computing model [12]. Chen et al. designed and implemented a decentralized credit reporting system model, and the experimental results showed that the model could guarantee high data security [13].

Some research literature paid attention to the constraints of blockchain applied to credit information sharing. Lemieux used the land registration system of a developing country as the implementation environment of the blockchain, and studied the performance of the blockchain in the creation and preservation of trusted digital records [14]. Hofman et al. discussed the contradiction between the immutable nature of blockchain data and the de-identification of personal information specified in the EU's "General Data Protection Regulation" [15]. Franks analyzed the potential risks of blockchain's distributed ledger technology applied to information management [16].

In terms of industry applications, financial institutions have successively made preliminary attempts to combine blockchain technology with credit reporting, such as GXChain, CTRChain, LinkEye, Trust Union, JD Vientiane, etc. But in general, blockchain as a cutting-edge technology has some bottlenecks, and its application in credit information sharing is still in the exploratory stage. Existing research literature is mostly based on conceptual exploration, and its practical applications in the industry is far from being mature. In the future, it still needs a lot of technical investment and industry practice to prove its application effect.

3 Materials and Methods

3.1 The Platform Infrastructure Design

Depending on the scope, blockchain can be divided into three types: public blockchain, consortium blockchain and private blockchain. As an important part of the financial infrastructure, data security and data quality in the credit reporting business are extremely important. Therefore, when the blockchain is applied to the credit reporting field, the public blockchain mode in which nodes can freely enter and exit the network is inappropriate, and the consortium blockchain mode with strong controllability should be adopted. The consortium blockchain uses a real-name entry mechanism. Nodes can join and exit the network after being authorized. Access institutions include credit bureaus, regulatory authorities, Internet financial companies, and data service providers. The institutions form a stakeholder alliance to jointly maintain the healthy operation of the blockchain. The access mechanism of the consortium blockchain makes it more advantageous in terms of operating efficiency, cost and supervision, and it is also easier to implement applications.

As a cutting-edge technology, the basic architecture and application scenarios of the blockchain have been expanding. The blockchain 1.0 architecture is represented by bitcoin and uses blockchain as a support platform for digital currencies. The blockchain 2.0 architecture technically introduces Ethereum smart contracts, using the blockchain as a programmable distributed credit infrastructure. The application scenarios are extended to broader financial sectors such as payment, credit financing, financial transaction, securities, insurance and leasing. Blockchain 3.0 architecture has stricter access mechanism and permission control, and is oriented to enterprise-level application scenarios beyond the scope of currency and finance [17], represented by Hyperledger. The credit information service has high requirements for data security, and is more suitable for using a partially decentralized hybrid architecture. Strict access mechanisms should be implemented, and the alliance members read and write blockchain data based on authorization. Taking the blockchain 3.0 architecture as a framework, the basic architecture of the blockchain-based personal credit information sharing platform can be designed (see Fig. 1).

3.2 Detailed Design of the Platform and the Technical Advantages

Blockchain is a combination of multiple mature computer technologies including P2P network, encryption algorithm, timestamp, consensus mechanism, and smart contract. It has technical characteristics such as distributed storage, partial decentralization, quasi-anonymity, security, open source and programmable, which can solve the difficult problems in credit information sharing in a targeted manner.

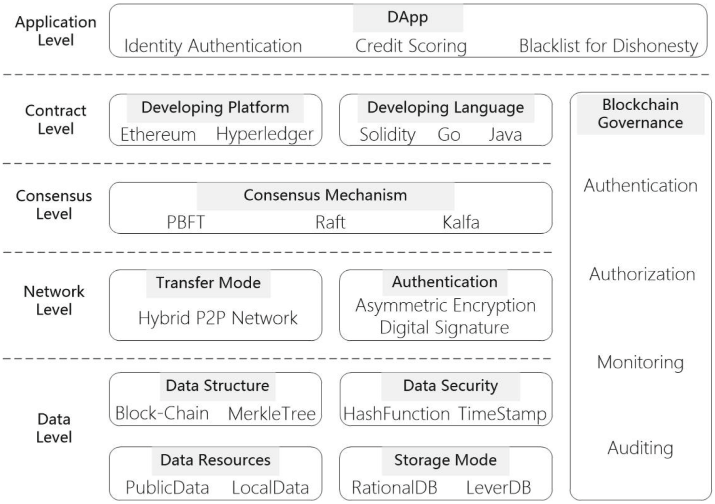

Fig. 1. Infrastructure diagram of the blockchain-based personal credit information sharing platform.

Hybrid P2P Network to Achieve Distributed Storage of Credit Data. The blockchain uses a peer-to-peer (P2P) network structure to organize all network nodes. It does not have a centralized node, but uses distributed storage technology. Each node stores a copy of the complete data. In the field of personal credit reporting, a centralized credit bureau is indispensable, so a "partially decentralized" hybrid P2P network (see Fig. 2) can be used to ensure the controllability of the system. Credit bureaus, regulatory authorities, and data service providers serve as super nodes to form a distributed network. Each super node and several ordinary nodes (users) form a local centralized network. Data does not need to be shared globally, only the super nodes have the authorities to read and write data.

The P2P structure can effectively use the large and scattered storage resources in the network to achieve distributed storage of credit data. Each node maintains a complete database, which means that all business data is open, transparent, and completely consistent. There is no need for data integration later, which reduces the cost of mutual trust between nodes. Data can be directly transmitted between nodes without going through a third party, reducing the risk of data leakage. And the reliability of the system is better. The damage of any node database will not affect the normal operation of the entire blockchain system.

Hash Function to Achieve Anonymization of Personal Information. Each piece of business data in the blockchain can be mapped into a series of hash values similar to garbled characters composed of numbers and letters through a hash encryption function, thereby hiding specific information. For example, the SHA256 hash function can convert business data of any length into a string of hash values composed of 64 numbers or letters. The hash function is unidirectional, and there is no way to reverse and decrypt. It can

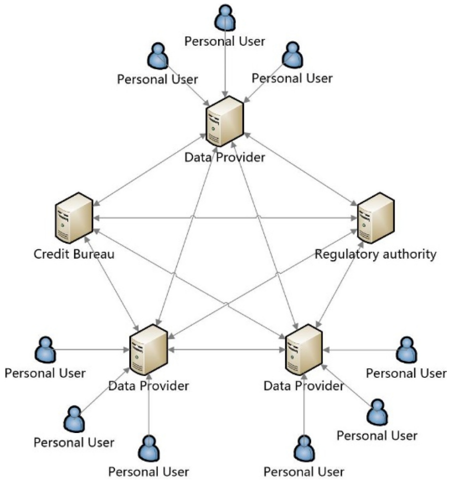

Fig. 2. Hybrid P2P network structure diagram

be used to encrypt personal identity data and sensitive data, and strengthen the privacy protection of the information subject.

Blockchain Structure to Ensure the Credit Data Cannot Be Tampered With. The blockchain adopts a block-chain data structure. The system creates a block at regular intervals. All credit business data is stored in the block, and each block is connected into a chain in the order of creation time. The block identifier in the block header is used to uniquely identify the hash value of the block. The blocks are linked by the hash value of the previous block (also known as the parent hash) and can be traced back to the first block (see Fig. 3). Credit business data is stored in the Merkle tree structure in the block body, and the leaf nodes are paired up to perform a hash operation up to the root of the Merkle tree in the block header.

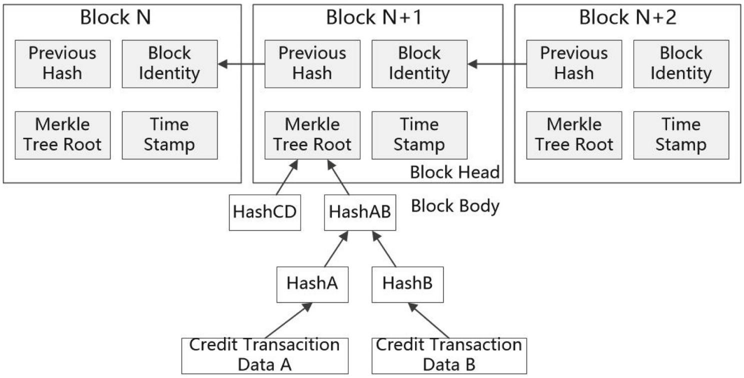

Fig. 3. Schematic diagram of blockchain data structure

The blockchain data structure can ensure that the credit business data cannot be tampered with, because once a certain piece of business data in the block is modified, the Merkle tree of the block needs to be recalculated, so that the Merkle tree root and block identification in the block header change, and no longer match the parent hash saved in the next block. The immutable nature of blockchain makes it a natural "trust machine".

Time Stamp to Ensure the Traceability of Credit Data. The timestamp is a valid proof of the order in the blockchain. It records the generation time of each block and the entry time of each piece of business data to the nearest millisecond. The timestamp adds a time dimension to the data in the block-chain, making the data easier to trace. At the same time, it also provides proof of existence for the data, ensuring the authenticity and unforgeability of the data, further increasing the difficulty of tampering with the data, and improving the credibility of the credit data.

Asymmetric Encryption Algorithm to Strengthen Credit Data Security.
The asymmetric encryption algorithm means that each node in the blockchain has a unique pair of keys, where the public key is public and indicates the identity of the node, and the private key is not public, indicating the control of information. Information encrypted using one of the keys can only be decrypted by the corresponding other key. Elliptic curve cryptography (ECC) is a classic asymmetric encryption algorithm, the equation can be written as:

$$y^2 = x^3 + ax + b(mod\, p) \tag{1}$$

In the above formula, a and b are coefficients, p is a prime number greater than 3, and $G(x, y)$ is a discrete point on the finite field F_p.

The elliptic curve has the following properties: given a certain point G of the elliptic curve, it is easy to find the points $2G$, $3G$, …, kG; on the contrary, if the point G and the point kG are known, it is very difficult to find k. Using this characteristic of the elliptic curve, using k as the private key and kG as the public key, asymmetric encryption can be achieved.

The information transmission process under asymmetric encryption mechanism is shown in Fig. 4. Suppose the receiver's private key is k, and the public key is $K = kG$. The sender selects a random number r, encrypts the information M with the public key K, and generates the ciphertext $C = \{rG, M + rK\}$. The information is transmitted on the network in the form of ciphertext C. After receiving the information, the receiver can decrypt the information with the private key k, that is, $M + rK - k(rG) = M$, and other nodes that receive the information cannot decrypt it. During information transmission, the private key k is not exposed on the network, which can reduce the risk of data leakage.

Digital Signature for User Identity Verification. Hash functions and asymmetric encryption algorithms can be used for digital signatures (see Fig. 5). Suppose the sender's private key is k, and the public key is $K = kG$. The specific process of digital signature is:

a. The sender uses the hash function to map the information M into a hash value h;
b. Select a random number r and calculate the point $rG(x, y)$;

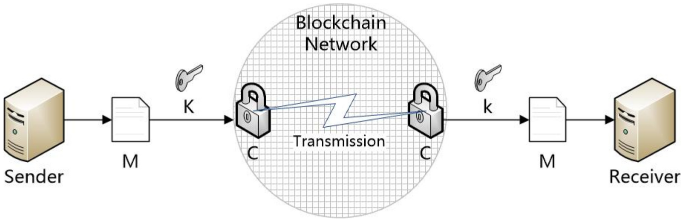

Fig. 4. Schematic diagram of asymmetric encryption mechanism

c. Encrypt the hash value h with the private key k, calculate $s = (h + kx)/r$, and get the digital signature $\{rG, s\}$;
d. Send the information M and the signature $\{rG, s\}$ together to the receiver;
e. After receiving the information and signature, the receiver first obtains the hash value h according to the information M;
f. Use the public key K to decrypt the signature, calculate $hG/s + xK/s$, and compare to rG, thereby verifying whether the information comes from the sender.

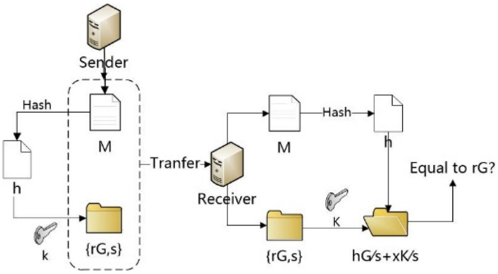

Fig. 5. Schematic diagram of digital signature principle

Consensus Mechanism to Ensure the Consistency of Credit Data. The consensus mechanism is a mechanism that uses mathematical algorithms to create trust between nodes without central control. The consortium blockchain has higher requirements for consistency, and the consensus mechanism usually uses the Practical Byzantine Fault Tolerance (PBFT) algorithm. The PBFT algorithm divides nodes into two categories: a master node and several slave nodes. The master node is responsible for sorting the requests issued by each node, and the slave nodes execute the requests in this order. After a certain node broadcasts the information to the net-work, each node uses the pairwise interaction to make the consistency judgment. When two-thirds of the nodes in the network reach a consensus on storing this information, the information can be stored in the block, thus ensuring the consistency of data storage.

The basic process of applying the PBFT algorithm in the platform is as follows:

a. Select R_0 as the master node, and the other 3n alliance members as slave nodes, and the new block is generated by the master node;

b. The alliance member R_1 sends a credit data write REQUEST to the blockchain network with private key signature and time stamp;
c. After receiving the REQUEST, the master node R_0 determines its order in the multiple credit data to be written into the new block, synthesizes the sequence number M and the REQUEST into a PRE-PREPARE message, and broadcasts it to all alliance members;
d. Each alliance member receives the PRE-PREPARE message, generates a PREPARE message, and broadcasts it to other alliance members;
e. Each alliance member receives the PREPARE message from other alliance members. When it receives 2n identical messages, it will confirm the PREPARE message, generate a confirmation message COMMIT, and broadcast to other alliance members;
f. After any alliance members in the network receives $(2n + 1)$ COMMIT messages, it means that a consensus can be reached and the credit data can be written into a new block.

Token and Smart Contract to Achieve Credit Data Pricing. Token is a value transmission carrier in the blockchain network. By issuing tokens, the blockchain can quantify the data contribution of each node and give corresponding token rewards to realize the increase of value in the information sharing platform. A smart contract is a computer program deployed on the blockchain, which triggers automatic execution once the conditions are met. In order to clarify the benefits of data transactions and ensure a more orderly operation of the platform, tokens and smart contracts can be combined to issue a certain number of tokens on the plat-form. In the smart contract, economic factors and token incentive rules are written in the form of computer programs to realize automated and rational incentives to ensure the compensation and fairness of data transactions.

Unlike the public blockchain, the purpose of the consortium blockchain platform to issue tokens is not to seek the appreciation of tokens through secondary market transactions, but to replace the circulation of currencies and improve the efficiency of data transactions. When querying external data, the alliance agency needs to pay tokens to the data provider and access the data only after being authorized, thereby ensuring the economic interests of the data provider and encouraging institutions to actively share data.

4 Conclusion

Blockchain has many technical advantages such as distributed storage, security, reliable, quasi-anonymity, tamper proofing, open source and programmable, by applying it to the field of credit reporting, credit agencies can achieve higher efficiency to share personal credit information at a lower cost. For example, the P2P network structure of the blockchain can realize the distributed storage of massive credit information, and the system has better security, reliability and scalability; the blockchain data structure, hash algorithm and timestamp can ensure the integrity, the immutability and traceability of the credit data, improving the quality of credit data; asymmetric encryption algorithm helps to strengthen user identity verification and personal information protection in the credit reporting transaction, and improve data security; the consensus mechanism uses mathematical algorithms instead of third-party intermediaries to create trust, which can solve the problem of data ownership and ensure the consistency of data storage; smart contract can build an automated and fair incentive mechanism to guide data service providers to actively participate in data sharing.

The decentralized, tamper proofing, and self-incentive features of the blockchain pose new problems for current laws and regulations, and profoundly affect the economy, finance, society, organizational form and governance. There are still inconsistencies between the technical characteristics of the blockchain and the existing regulatory systems and methods of the credit reporting industry, which are mainly manifested in: the decentralized nature leads to the dispersion of supervision and weakens the control of the regulatory authority; the tamper-proof feature of blockchain conflicts with the "Credit Management Regulations" in China which stipulates that "credit information over 5 years should be deleted", and the archiving function can be considered to delete part of the cold data to reduce storage burden; the legal validity and compliance of smart contracts and token issuance are subject to discussion, etc. These need to be followed and explored in theoretical research and industry practice.

Blockchain research is not only a technical issue, but more importantly, it should effectively play the role of blockchain technology in reducing costs, improving efficiency, and optimizing the integrity environment in specific applications. The research on blockchain-based credit reporting should focus on the precise needs and real application scenarios of the personal credit industry, reasonably take advantage of the technical features of blockchain, rationally study the value of blockchain in traditional industries, and develop practical solutions with broad application prospects and potential value, thus giving full play to the role and positive influence of blockchain on personal credit reporting and effectively promoting the coordinated development of the social credit system.

References

1. An, X., et al.: Electron. Libr. **33**(6), 1047–1064 (2015). https://doi.org/10.1108/EL-04-2014-0059
2. Onay, C., Öztürk, E.: J. Finan. Regul. Compliance **26**(3), 382–405 (2018). https://doi.org/10.1108/JFRC-06-2017-0054
3. Zhang, J.: New trend for personal credit scoring in big data era. Credit Ref. **35**(12), 7–12 (2017)
4. Zhang, J., Li, Y.: Study on the market-oriented process of china's personal credit reporting industry from the perspective of Baihang Credit Co. Ltd. Credit Ref. **37**(12), 54–60 (2019)
5. Poritskiy, N., Oliveira, F., Almeida, F.: The benefits and challenges of general data protection regulation for the information technology sector. Digit. Policy Regul. Gov. **21**(5), 510–524 (2019). https://doi.org/10.1108/DPRG-05-2019-0039
6. Lee, C.S.: Datafication, dataveillance, and the social credit system as China's new normal. Online Inf. Rev. **43**(6), 952–970 (2019). https://doi.org/10.1108/OIR-08-2018-0231
7. Liao, L.: The developing situation and thinking of personal credit industry in China. People's Tribune **20**, 76–77 (2019)
8. Ba, S.S.: Blockchain is a good solution to solve the problem of credit reporting market. China Security News 2019–06–22(A07)

9. Zhu, H.Q.: The application prospect of financial technology in the field of market-oriented personal credit reporting. Finan. Comput. **12**, 41–44 (2018)
10. Liu, X.H., Jia, H.Y., Han, X.L.: Blockchain: a new perspective and technical architecture for credit reporting. Credit Ref. **38**(04), 13–21 (2020)
11. Ju, C.H., Zou, J.B., Fu, X.K.: Design and application of big data credit reporting platform integrating blockchain technology. Comput. Sci. **45**(S2), 522–526+552 (2018)
12. Wang, T., Ma, W.P., Luo, W.: Information sharing and secure multi-party computing model based on blockchain. Comput. Sci. **46**(09), 162–168 (2019)
13. Chen, C.L., Shen, Y., Yu, H.: Research on decentralized model for credit information system. Comput. Technol. Dev. **29**(03), 122–126 (2019)
14. Lemieux, V.L.: Trusting records: is Blockchain technology the answer. Rec. Manag. J. **26**(2), 110–139 (2016)
15. Hofman, D., Lemieux, V.L., Joo, A., Batista, D.A.: The margin between the edge of the world and infinite possibility-Blockchain, GDPR and information governance. Rec. Manag. J. **29**(1), 240–257 (2019)
16. Franks, P.C.: Implications of blockchain distributed ledger technology for records management and information governance programs. Rec. Manag. J. (2020). https://doi.org/10.1108/RMJ-08-2019-0047
17. Maesa, D.D.F., Mori, P.: Blockchain 3.0 applications survey. J. Parallel Distrib. Comput. **138**, 99–114 (2020)