Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 3457967, 14 pages
https://doi.org/10.1155/2021/3457967

Research Article

Blockchain Equity System Transaction Method and System Research Based on Machine Learning and Big Data Algorithm

Kanghua Peng

Guangdong Engineering Polytechnic, Information Engineering Institute, Guangzhou, 510520 Guangdong, China

Correspondence should be addressed to Kanghua Peng; pengkanghua@gdep.edu.cn

Received 23 April 2021; Revised 18 June 2021; Accepted 9 July 2021; Published 19 July 2021

Academic Editor: Wenqing Wu

With the development of machine learning and big data, traditional equity trading system methods can no longer meet the current trading needs, and there are still problems such as low operating efficiency and serious homogeneity. Blockchain technology has the characteristics of decentralization and can also complete transactions through smart contracts, innovating the way of equity system transactions. The purpose of this paper is to build an equity trading system in combination with blockchain in the context of machine learning and big data and provide innovative trading methods, so as to provide reference and reference significance for the construction of my country's equity market. This article uses literature data method, comparative analysis method, factor analysis method, and other methods to carry out research, in-depth study of machine learning and big data, blockchain-related concepts, system composition, application situation, etc., and discusses the allocation of equity trading market The functions of resources, risk diversification, risk transfer, price determination, etc., have built a blockchain equity trading system, designed a consensus mechanism, block generation protocol, block verification, decentralization, and smart contract platform, and finally conducted a national equity transaction the background of the market is analyzed, and the experimental results, simulation indicators, transaction time, transmission consumption, and other content of the system constructed in this article are analyzed. In the single-node test, the CPU usage of the PoW consensus mechanism algorithm reached 100%, but the improved PBFT consensus mechanism was only 16%, which saved a lot of computing power and improved computing performance.

1. Introduction

After the emergence of Bitcoin in 2009, blockchain technology has slowly entered people's field of vision. Blockchain technology has the characteristics of decentralization, transparency, reliability, etc., so it has attracted the attention of experts and researchers. As an application of blockchain technology, Ethereum provides a decentralized application platform. The application of blockchain technology initially developed from the financial field to other fields. In recent years, many fields have gradually tried to use blockchain technology, which has promoted the research and development of blockchain technology.

In recent years, the state has actively promoted the construction of a multilevel capital market and vigorously developed the over-the-counter market. Because the traditional equity trading system method can no longer meet the current trading needs, there are problems such as low operating efficiency and serious homogeneity, so there is a blockchain stock system trading method and system. In January 2013, the country established a share transfer system for SMEs, and the over-the-counter market entered a stage of rapid growth. At the same time, as an important part of the multilevel capital market system, the equity market promotes regional economic development, provides funds for small and medium-sized enterprises, and enhances the company's image. However, in the process of vigorous development of the equity market, some problems have been exposed, such as qualitative, redundant construction, and low efficiency. The equity trading market plays an important role in the country's economic development and provides strong economic support.

Boehm et al.'s growing demand for customized machine learning (ML) algorithms and the growing amount of data that need to utilize distributed data parallel frameworks (such as Map Reduce or Spark) pose a major challenge to the productivity of data scientists. They solved these challenges through declarative ML: (1) increased the productivity of data scientists because they were able to express custom algorithms in familiar domain-specific languages, including linear algebra primitives and statistical functions; (2) in distributed computers to run these ML algorithms transparently, the data parallel framework applies cost-based compilation techniques to generate efficient, low-level execution plans, which include single-node and large-scale distributed operations in memory. He described the end-to-end system ML on apache spark, including an in-depth understanding of various optimizers and runtime technologies and performance characteristics. They also shared the lessons learned from porting system ML to spark and declarative ML. Finally, system ML is open source, which allows the database community to use it as a testing platform for further research. However, his system contains too many mathematical algorithms and functions, so it does not have much practical performance [1]. Obermeyer and Emanuel proposed a deep learning-based traffic flow time series prediction model, deep TFP, which makes full use of the effectiveness of time series function analysis sequence data and deep learning to extract traffic flow features. Accurate and timely prediction of future traffic flow is an urgent need for personal travel, public transportation, and transportation planning. In recent years, with the explosive growth of traffic data, various traffic flow prediction methods based on big data analysis have been proposed. He proposed deep TFP, which uses a time series function that considers the temporal and spatial correlation of traffic data to predict and track changes in traffic flow. Deep TFP uses deep learning to extract the characteristics of traffic data as the basis of the time series function. The performance of the model is verified through comparative experiments. However, his experimental model is difficult to provide timely predictions for the processing of real-time traffic data [2]. The purpose of Hamzah et al.'s research is to study the motivation of risk transfer in debt and equity contracts based on criticizing the similarities between sukuk and bonds. He uses a theoretical and mathematical model to study whether there is a motivation for risk-taking: debt contracts and equity contracts. Based on this theoretical model, it believes that the risk transfer behavior only exists in the debt contract, because the debt will naturally produce risk transfer behavior when the transaction occurs. In contrast, stock contracts, by their very nature, involve the sharing of transaction risks and returns and are therefore considered undesirable risk transfer behavior. Nevertheless, previous researchers found that equity financing may also bring motivation for risk transfer. Even so, he believes that the amount of capital provided and the underlying assets must be considered, especially in the case of default. Through mathematical modeling, this element of equity financing can make risk transfer unattractive. However, his research model is only a theoretical model. There are many factors that need to be considered in actual operation, and the author did not take them into consideration [3].

The main innovations of this paper are (1) using literature analysis, comparative analysis, and other methods to describe and analyze the equity trading system, through the comparison of different system models, highlight the feasibility and effectiveness of the framework of this paper; (2) the combination of theoretical analysis and empirical research not only analyzes the shortcomings of the traditional equity trading system and explains possible countermeasures but also collects data and analyzes through data comparison.

2. Blockchain Equity System Transaction Methods and System Research Methods Based on Machine Learning and Big Data Algorithms

2.1. Machine Learning

2.1.1. Overview of Machine Learning. Machine learning is a field of computer science. There is no clear program design method for the computer to learn [4]. Its purpose is to allow the computer to simulate the human learning process, use existing knowledge (data) to form an independent knowledge system, and solve future problems. At the same time, the data is updated and improved through the knowledge system produced by machine learning [5, 6].

The mathematical description of the basic model of machine learning is as follows: the data generator H generates m samples $(x_1, y_1), (x_2, y_2), (x_3, y_3) \cdots (x_n, y_n)$, assuming that the probability distributions of x and y of the observation samples are consistent, and generates independent observation samples of the same distribution $G(x|y)$. Machine learning is to find the most suitable learning function $\{g(x, w)\}$ from a series of learning functions based on observation samples $g(x, w_0)$. Therefore, when predicting and evaluating the observed samples, the predicted risk $T($ $w)$ of x and y is the smallest:

$$T(w) = \int K(y, g(x, w))dG(x, y). \tag{1}$$

Among them, w is the general parameter of the learning function, $K(y, g(x, w)$ is the error function between the predicted value y and the actual value y when using the predictive learning machine $g(x, w)$ [7]. It is called the risk function. The form of the risk function varies with learning problems. General machine learning problems are divided into pattern recognition problems, regression evaluation problems, and probability evaluation problems. The risk functions corresponding to these three problems are as follows [8]:

Pattern recognition problem:

$$K(y, g(x, w)) = \begin{cases} 0, y = g(x, w), \\ 1, y \neq g(x, w). \end{cases} \tag{2}$$

Regression estimation problem:

$$K(y, g(x, w)) = (y - g(x, w))^2. \tag{3}$$

Probability density problem:

$$K(\sigma(x, w)) = -\ln \sigma(x, w). \quad (4)$$

2.1.2. Gemini Support Vector Machine. The traditional vector machine SVM realizes SRM by constructing a pair of parallel classification hyperplanes [9]. Different from traditional SVM, TWSVM looks for a pair of nonparallel classification hyperplanes, which are a positive hyperplane $\varpi_1^T \gamma(x_i) + b_1 = 0$ and a negative hyperplane $\varpi_2^T \gamma(x_i) + b_2 = 0$. The positive sample points are located near the positive hyperplane, while the negative sample points are far away from the positive hyperplane, and vice versa [10]. We assume that the positive sample with the label +1 is a matrix $X_1 \in i^{m_1 \times n}$, and the negative sample with the label -1 is a matrix $X_2 \in i^{m_2 \times n}$, so that the optimization problem can be expressed as a formula:

$$\begin{cases} \min\limits_{\varpi_1, b_1, \psi} \dfrac{1}{2} \|\gamma(X_1)\varpi_1 + e_1 b_1\|^2 + c e_2^T \psi_1, \\ s.t. -(\gamma(X_2)\varpi_1 + e_2 b_1) + \psi_1 \geq e_2, \psi_1 \geq 0, \end{cases} \quad (5)$$

and,

$$\begin{cases} \min\limits_{\varpi_1, b_1, \psi} \dfrac{1}{2} \|\gamma(X_2)\varpi_2 + e_2 b_2\|^2 + c e_1^T \psi_2, \\ s.t. -(\gamma(X_1)\varpi_2 + e_1 b_2) + \psi_2 \geq e_1, \psi_2 \geq 0, \end{cases} \quad (6)$$

where $c > 0$, ψ_1, ψ_2 is the relaxation vector. In order to obtain the corresponding dual problem, TWSVM assumes that the matrix $J^T J$ sum $H^T H$ is a nonsingular matrix [11], where $J = \gamma(X_1) + e_1$, $H = \gamma(X_2) + e_2$. Under the constraint of this condition, the corresponding dual problem can be expressed as a formula:

$$\begin{cases} \max\limits_{\beta} e_2^T \beta - \dfrac{1}{2} \beta^T H \left(J^T J\right)^{-1} H^T \beta, \\ \text{s.t.} 0 \leq \beta \leq c, \end{cases} \quad (7)$$

and,

$$\begin{cases} \max\limits_{\beta} e_2^T \chi - \dfrac{1}{2} \chi^T J \left(H^T H\right) J^T \chi, \\ \text{s.t.} 0 \leq \chi \leq c, \end{cases} \quad (8)$$

where $\beta \in R^{m_2}$, $\chi \in R^{m_1}$ is the Lagrange multiplier.

When $J^T J$ or $H^T H$ is a singular matrix, the corresponding dual problem cannot be expressed in the form of the above two formulas. In order to make the formula hold when $J^T J$ and $H^T H$ are singular matrixes, replace the $\left(J^T J\right)^{-1}$ sum $\left(H^T H\right)^{-1}$ with the $\left(J^T J + \varphi I\right)^{-1}$ sum$\left(H^T H + \varphi I\right)^{-1}$, respectively, where $\varphi > 0$ I is the identity matrix. After replacement, it can be expressed as the following formula:

$$\begin{cases} \max\limits_{\beta} e_2^T \beta - \dfrac{1}{2} \beta^T H \left(J^T J + \varphi I\right)^{-1} H^T \beta, \\ \text{s.t.} 0 \leq \beta \leq c, \end{cases} \quad (9)$$

And,

$$\begin{cases} \max\limits_{\beta} e_1^T \chi - \dfrac{1}{2} \chi^T J \left(H^T H + \varphi I\right)^{-1} J^T \chi, \\ \text{s.t.} 0 \leq \chi \leq c. \end{cases} \quad (10)$$

By solving the two dual problems of the above formula, a pair of nonparallel hyperplanes can finally be constructed.

$$\begin{cases} \begin{bmatrix} \varpi_1 \\ b_1 \end{bmatrix} = -\left(J^T J + \varphi I\right)^{-1} H^T \beta, \\ \begin{bmatrix} \varpi_2 \\ b_2 \end{bmatrix} = \left(H^T H + \varphi I\right)^{-1} J^T \chi. \end{cases} \quad (11)$$

It should be noted that, strictly speaking, the solution formula is only an approximate solution to the original problem. In classification prediction, the unknown sample is closer to the positive (negative) hyperplane than the negative (positive) hyperplane, and the sample is marked as the positive (negative) [12, 13].

In short, the difference between SVM and TWSVM is that SVM separates the positive and negative samples by a maximum interval classification hyperplane, while TWSVM separates the positive and negative samples by a pair of nonparallel hyperplanes [14]. In TWSVM, the decision planes of the two types of samples are based on the other type of samples as constraints, so that the two types of samples are closer to the type of decision plane, and at the same time far away from the nontype of decision plane. The samples are classified by judging the distance between the sample points and the positive and negative decision planes [15].

When dealing with nonlinear problems, the Gemini Support Vector Machine is the same as the ordinary support vector machine, which recognizes linear separation by mapping data samples from the low-dimensional input space to the high-dimensional space [16, 17]. The experiment in this chapter uses the RBF kernel function. The RBF kernel function parameter v and penalty coefficient c affect the prediction performance of the model. Therefore, it is very important to choose the right combination of parameters.

2.2. Blockchain Technology. With the advent of Bitcoin, a decentralized and highly reliable transaction system based on electronic virtual currency—a system associated with blockchain technology—has slowly emerged. The emergence of Bitcoin has received widespread attention in the financial and computer fields, and research scholars have also begun to further study blockchain technology [18, 19].

2.2.1. Related Concepts of Ethereum. Ethereum is a distributed computer platform based on the public chain, providing a distributed virtual machine that can run a complete scripting language. Ethereum is composed of smart contract layer, incentive layer, consensus layer, network level, and data level [20]. The data layer includes the most basic data structure and account encryption algorithm of the Ethernet core part of the Ethernet, especially the data transmission verification mechanism of each node of the Ethernet. Ethereum adopts a consensus mechanism based on consensus-level verification work [21, 22]. This is mainly used to induce nodes to independently mine and maintain Ethereum. The data layer, network layer, consensus level, and incentive level are also the basic content of the blockchain structure [23].

(1) Data Layer. The data layer merges the data structure of the block with the content related to data encryption. The block is divided into two parts: the block header and the block body. All transaction-related information is contained in the main body [24].

The network layer needs to include the implementation of the P2P network, message delivery mechanism, and data verification. The network layer is a network guarantee for the interaction between Ethernet nodes [25]. Ethernet uses P2P networks to solve the decentralization and other characteristics of Ethernet. Because the P2P network uses flags, the status of each node is the same, and there is no central server. Each node is equivalent to the routing, verification, and sending functions of the Ethernet system. P2P network architecture is the foundation of Ethernet distributed network.

(2) Consensus Layer and Incentive Layer. In a centralized decentralized system, the method for each node to reach a consensus is the main content, and Ethereum also faces the inevitable problems of a distributed system. Ethereum uses a consensus mechanism based on work evidence to enable systems with decentralized decision-making power to reach consensus. Each node solves mathematical problems through the competition of computer power and rewards the first node who solves the problem, thereby promoting the competition of the overall computing power of the network.

Based on the PoW negotiation mechanism, the total computing power of the nodes in the network must be large enough to avoid the concentration of computing power. Therefore, in order to make full use of the total computing power of the entire network, it is necessary to rely on an incentive mechanism that promotes the utilization of the entire network to confirm that the decentralization of Ethernet is safe and reliable.

2.2.2. Smart Contract. Blockchain technology simplifies the value transfer process through its network architecture and the virtual electronic currency it realizes. In the blockchain, a transaction is divided into the following three steps: (1) A sends a message to the blockchain network, and a transaction is defined in the message; (2) B accepts the transaction through broadcast, indicating that B accepts the transaction; (3) participants in the blockchain network verify the legality of the transaction and complete the transaction. To be sure, this transaction model defines a process in which the owner of a certain currency in the blockchain has changed. But when sending or accepting transactions, the value in the message can also be conveyed, just like the electronic envelope, the electronic currency may be contained in the electronic envelope and transmitted through the network, but it can also transmit some other information to achieve additional utility.

A typical transaction can be regarded as a simple "script," this script is equivalent to a set of instructions. The instruction sets allow nodes in the blockchain to execute transactions. However, when the script changes and contains information other than transactions, then users can use this script to implement complex "transactions." A smart contract is a convenient, verifiable, and executable computer agreement with the terms of a commercial agreement. This concept is not a new concept nor is it proposed by the blockchain. What the blockchain realizes is the decentralization of smart contracts. In other words, smart contracts are built in public databases without central supervision and execution. Participants can automatically execute the content of the contract without the supervision of a third party. Smart procurement is also the basis of blockchain decentralized applications.

In Ethereum, a smart contract is a collection of code and data and is stored in the Ethereum blockchain with the address of the smart contract account. Smart contracts are transmitted on the network in the form of messages, and smart contracts in Ethereum are stored in transactions. EVM judges whether it is a contract type transaction by whether a transaction contains a code. EVM implements the execution of the contract by executing the binary bytecode in the contract. In the Ethereum smart contract, it is defined that the application binary interface (ABI) is strongly typed, and the ABI will be formed and persisted during the compilation of the smart contract. Here, ABI is similar to the regulations in the actual contract, which must be obtained every time the contract is called. The first four bytes of a function call data specify the function to be called. The account uses the smart contract ABI to call the method in the contract.

2.3. Equity Trading Market

2.3.1. Overview of the Equity Market. The positioning of my country's equity trading market has become increasingly clear with the practice of various regional equity markets and the continuous release of relevant policies. From the perspective of policy analysis, the positioning of my country's equity trading market is mainly divided into three aspects: first, the equity trading market is an extension of our multilevel capital market, an important part of my country's capital market system, and it belongs to the private equity market; second, it serves as a platform for small, medium, and microenterprises in provincial administrative regions to cultivate and regulate small, medium, and microenterprises; third, it is a comprehensive platform for local governments to support and promote the development of small, medium, and microenterprises in the region.

2.3.2. Functions of the Equity Trading Market. The equity trading market does not exist independently. It is an important part of my country's capital market. The main functions of the capital market are basic functions such as the optimal allocation of resources, the dispersion of risks, the transfer of risks, and the determination of prices. It also has a unique function as the foundation of the capital market: to cultivate small, medium, and microenterprises and regulate their development as an incubator; to strengthen the function of local governments to guide and support the development of small, medium, and microenterprises in the region.

The equity trading market is an extension of the capital market's function, which is manifested in the following aspects: (1) optimal allocation of resources: The equity trading market is an important part of the multilevel capital market. The basic function is to adjust the function of funds and transfer the surplus of funds in the market. The fund is transferred to the party lacking funds, but it is biased to serve small, medium, and microenterprises and provide more financing channels for small, medium, and microenterprises; on the other hand, it also provides a new investment platform for institutional investors to make it from small, medium, and microenterprises. (2) Risk diversification and risk transfer: when institutional investors invest in small, medium, and microenterprises, they often face greater investment risks because they face market risks and the moral hazard of major shareholders. The equity trading market provides institutional investors with more investment opportunities, so that it can effectively reduce the risks faced by institutional investors. In the early stages of development, small and medium-sized and microenterprises, especially those in the industry in the introduction and growth stages, have a large demand for funds and often face higher financial risks. The regional equity trading market provides them with equity and convertible bonds. The way in which financing transfers its risk. (3) Determine the price: SMEs can increase their credit by putting stocks into the market by SMEs, thereby expanding financing channels, providing a new evaluation platform, and reducing financing costs.

The unique functions of the equity trading market are (1) the incubator function for cultivating small, medium, and microenterprises and standardizing their development: The equity trading market can carry out professional corporate governance and investment and financing training courses through its own capital market's professional attributes, thereby improving the performance of listed companies. Professional management capabilities and a sound internal control system have laid a solid foundation for its rapid development. (2) Strengthen the ability of local governments to guide and support the development of small, medium, and microenterprises in the region. Since the establishment of China's capital market is to serve state-owned enterprises, and the capital market's characteristics of "dislike the poor and love the rich", a huge number of small, medium, and microenterprises obtain a very small proportion of resources from the capital market. The guidance and support of local governments can effectively reduce the difficulty of financing small and medium-sized enterprises in the capital market and also help local governments develop business clusters with local characteristics, thereby driving the development of local economy and social people's livelihood.

In addition to the above functions of the equity trading market, the equity trading market is also a comprehensive functional market for local governments to promote the development of small, medium, and microenterprises in the region. In this market, local governments can directly introduce targeted preferential policies to attract banks, brokerages, insurance, and VC/PE markets. Participants actively participate in the development of local enterprises. At the same time, the market can also guide equity investment institutions and bond investment institutions to cooperate for high-growth enterprises to introduce comprehensive financing plans for the long-term development of enterprises.

3. Blockchain Equity Trading System Design Model Based on Machine Learning and Big Data Algorithms

Based on the background of machine learning and big data algorithms, this paper designs a blockchain equity trading system model, adopts the Ethereum PBFT consensus mechanism, and through the decentralized smart contract design, it improves computing time and efficiency and reduces the communication overhead. In addition, the implementation process is given in this article, and the experiment is verified.

3.1. Problem Assumptions and Definitions

3.1.1. Problem Assumption. The mechanism proposed in this paper is based on the alliance chain scenario and has the following assumptions: (1) there are n nodes,$\{M_1, M_2, \cdots, M_n\}$, each node is independent of each other, and there is no centralized platform. During the operation of the system, nodes will not be dynamically added or removed. (2) To establish a link between nodes, the verification method provided by Ethernet is required. Verification is considered safe, and malicious nodes cannot pretend to be connected to other nodes in the network. (3) After the node is disconnected, the system tries to reconnect to the node, but the connection fails after the limit time is exceeded, which is an error node.

The generation of a block of the blockchain is jointly decided by all preselected nodes (the preselected nodes participate in the consensus process), and other access nodes can participate in the transaction, but the process is not involved.

3.1.2. Algorithm Definition. The PBFT mechanism has the following definitions:

Definition 1. We set Quorum as a set group, and there is an intersection between the two. The set of system nodes is set to $\nabla = \{W_1, W_2, W_3, \cdots, W_n\}$, and $W_i \subseteq I$ if the formula is satisfied, W is called a quorum.

$$\forall W_i, W_j \in \nabla \quad \text{and} \quad W_i \cap W_j \neq \varphi. \tag{12}$$

Quorum has the following properties: (1) any two quorums have at least one common and correct node. (2) There must be a quorum without errors. Among them, this article defines a $2g+1$ node as a quorum, which can ensure that there is at least $g+1$ node in a quorum without errors, where g is the maximum number of error nodes in the system.

Definition 2. The PBFT organization will perform segmentation calculations for the execution process and mark it as b. There are master nodes in the view, the node set is N, the number of master nodes is recorded as q, and the other nodes are called backup nodes. When copying, each node is represented by an integer in order $\{0, 1,\cdots,N-1\}$, which satisfies

$$q = b \bmod N. \tag{13}$$

Here, b is the number of projections. When the master node in the group fails, the next numbered node will become the master node, and the view number will be modified accordingly.

Definition 3. Certificate (certificate), this article specifies the type of message sent in the system matching process. This message is called a certificate. The request is an information type $\{\text{REPLY}, m, b, i\}$. According to the certificate type, m means block or block fragment, b means project number, and i means node number.

3.2. System Design

3.2.1. Description of Consensus Mechanism. In a system with M nodes, the master node q is responsible for generating a block, and after all nodes negotiate and verify the block, they reach a consensus on the block. In order to tolerate Byzantine errors, there must be at least four nodes in the system designed in this paper. Each node adopts the same sequence during initialization. Each node establishes internode communication according to the IP address in the configuration file. After the communication is established, it will send the corresponding public key to other nodes, and the other nodes will complete the verification through the private key signature, and the verification will indicate that a valid connection is established between nodes. During the negotiation process, for a message node, there are three states: preprepare, prepare, and commit, which are represented by $PP_i^{\{v,m,n\}}$, $P_i^{\{v,m,n\}}$,and $C_i^{\{v,m,n\}}$, respectively, where v represents the number of the view, m is the certificate message, and n is the number of the node that sends the certificate message. i is the number of the current node.

In the consensus algorithm, when a certificate is generated, the state of the node is first judged, and the method of the current state will be executed, and at the same time, it will enter the next state, where the preprepare state needs to judge the master node, and only the master node can execute it. When a certificate has passed the three-state condition judgment, it is considered that the content of the certificate has reached a consensus among nodes.

As can be seen from the above, the experiment uses role-based access control (RBAC) and discretionary access control (DAC).

3.2.2. Block Generation Protocol. In the blockchain, transactions are sent in blocks and are permanently recorded in the blockchain. This article uses the Ethereum transaction verification protocol. After the same node confirms the validity of the transaction, the transaction is recorded in batches in the block, and the block is sent. The transaction on this block will be verified by other nodes on the network, and added to the blockchain, the transaction is considered to be executed. When a transaction occurs, the master node writes the transaction into the block and transmits the block. If all the network nodes agree on the block, try to add the block to the blockchain. The whole process is inconsistent. The order of the blocks is guaranteed by the block number and partition value of the block record. If the transaction list is empty, the node monitors the timing and system time of the best block in the blockchain. When the time exceeds t, an empty block will be created and added to the blockchain. There may be network delays in the sending process. The longest time for a block to be added to the blockchain after reaching consensus is Δt. Here, t must satisfy $t > \Delta t$. In this way, when a blank block is created, all transactions will be completed through the network. If an empty block is added to the blockchain, the master node will stop the creation of the block and reset the block after waiting for the transaction to arrive.

3.2.3. Block Verification. The verification process of the block is mainly based on the verification of the block header information. This article combines the block structure in Ethereum to perform the block verification. First, check ParentHash to determine whether the newly generated block points to the BestBlock of the blockchain in this node. When a block does not show the BestBlock of the current blockchain, the block will be added to the list first and another block will be added to the list. The thread surrounds to determine if there is a block that shows the BestBlock of the current blockchain in the list. If there is, it executes the block verification process and tries to add it to the blockchain. When the block points to the BestBlock of the current blockchain, the transactions in the block are first verified. The transactions in the block are sorted according to the time sequence of the transaction execution. According to this order, the Merkle tree can be generated, and the rootHash of the Merkle tree is compared with the root hash of the transaction tree of the new block is verified. The verification will verify each transaction and then generate a copy of the Repository object, where the repository contains the local database data of the node, and the next operation is performed in the copy. Because the state tree verification needs to modify the repository state, the operation in the copy can be rolled back. When the number of transactions in the block, the state tree, and the receipt tree roothash verification are all passed, the copy is submitted, and the block is added to the blockchain. Transaction verification is the core content of the entire blockchain verification. Here, we define the status transaction function as ω, the transaction is Y, and the execution of the transaction will change the local state database. Here, if the database status is defined as U, then, there is

$$U' = \omega(U, Y). \tag{14}$$

Among them, U' is the changed state, a transaction valid judgment needs to meet the following conditions:

$$u(Y) \neq \omega \wedge U[u(Y)] \neq \omega \wedge Y_n = U[u(Y)]_{\text{nonce}} \wedge v_0 \leq U[u(Y)]_{\text{balance}}. \tag{15}$$

Among them, $u(Y)$ is the initiator of a transaction, which Y_n and v_0 are the nonce and transaction amount in the transaction, nonce is the number of transactions in each account, used to ensure the orderliness of the transaction, and recorded in the account information. When a transaction fails, the local warehouse must implement the rollback function. Here, we define the state $U0$ as the rollback state. Then, there is

$$U_0 = U, \text{except} : \begin{cases} U_0[u(Y)]_{\text{balance}} = U[u(Y)]_{\text{balance}} - c, \\ U_0[u(Y)]_{\text{nonce}} = U[u(Y)]_{\text{nonce}} + 1. \end{cases} \tag{16}$$

Among them, c is the cost of executing this transaction. Next, the operation of adding the block to the blockchain will be executed, and after the successful addition, the state of the block being generated will be set to empty. Changes in these two states will trigger the next block generation process.

3.3. Implementation of Decentralized System Platform

3.3.1. Ethereum Interface Development. Since Ethereum only provides the Json-RPC call interface, it is necessary to write a Web Service interface for Ethereum to realize the interaction process with the web server. The interface management information is shown in Table 1.

In the smart contract module, it is also necessary to open the Web Service interface and open the web server to call the relevant interface of the smart contract.

3.4. Implementation of Smart Contract. The global variable taskMapping is defined in the smart contract to store all transactions. The structure describes the parameter information of the contract, taskId is used to distinguish tasks, and taskSender and taskReceiver are the account addresses of the task initiator and recipient, respectively. Used for bounty transactions, the amount is the value of the bounty, and state is the current state of the character. The event changeTask is to perform corresponding operations according to the triggered parameters each time it is triggered, including the modification of the task status and the bounty transaction process. Through this smart contract, the decentralization of the task release platform is realized.

4. Blockchain Equity System Transaction Method and System Research Analysis Based on Machine Learning and Big Data Algorithm

4.1. Background Research on the National Equity Exchange Market. Through a certain webpage, you can get the data of the national equity trading centers and show in the form of tables that the development level and overall scale of the Chinese equity trading centers have large regional differences, but the data obtained is still only the data of some provinces in the country. As shown in Table 2, the regional equity trading centers in eastern China are relatively active, especially in Shanghai, Shenzhen (Qianhai), and Zhejiang. These three equity trading centers have relatively more indicators such as the number of listed companies, total shares, and total assets than the central and western regions. This regional difference is directly proportional to the level of economic development and financial development to a certain extent. In particular, it can be seen from the distribution of the number of listed companies in Figure 1 that this trend is more obvious. For example, Shanghai, Shenzhen (Qianhai), Zhejiang, Wuhan, Beijing, and Liaoning equity trading centers in the east are relatively active.

It can be seen from Figure 1 that in terms of the number of listed equity trading platforms, the activity of regional equity trading centers is roughly proportional to the level of economic development and financial development, especially the top transactions. Most of the centers are located in the eastern coastal areas, but there are still equity trading centers such as Guangxi, Xinjiang, Chengdu (Sichuan-Tibet), and Chongqing, and the number of listed companies is relatively large.

The regional differences in the performance of regional equity trading centers are mainly due to the existence of thousands of small, medium, and microenterprises in the eastern coastal area. The huge base number has given birth to a large number of listed companies, and because the financial industry in this region is relatively developed, it can provide enterprises with financial services such as good corporate planning, listing training, and listing counseling, so the number of listings is relatively large. However, due to the limited number of enterprises and imperfect financial supporting services in the central and western regions, the number of listed companies in the equity trading center is limited and the transaction volume is relatively insufficient.

We divided the industries into three categories, namely, industry, agriculture, and service industries, and made specific subdivisions. We selected Y city as the data collection point and collected the number and proportion of the industry distribution. The results are shown in Table 3. From the perspective of industry distribution, at this stage, Guizhou stock exchange center listed companies basically cover industry, agriculture, and service industries, accounting for 80% of the total and emerging industries such as information technology and financial services account for a relatively low proportion.

Table 1: Blockchain management interface information.

Name	Parameters and description
AccountInfo	The parameter is the account address, used to query account related information
BlockNum	Used to query the number of current blockchain blocks
BlockInfo	The parameter is the block number, which is used to query the information of the block
ActiveNode	Return the reliable node considered by the current node
Balance	The parameter is the account address. Return the account balance
BatchTransactions	The parameters are the account address and the corresponding account amount. Used for system initialization account amount
BatchAccounts	The parameter is the account address, used to initialize the account

Table 2: National equity exchange center data.

Exchange	Number of listed companies	Total number of shares (ten thousand shares)	Total assets (ten thousand yuan)	Total net assets (ten thousand yuan)	Total operating income (ten thousand yuan)
Shanghai equity custody center	8353	77185	2318728.96	1006714.32	845103.31
Qianhai equity exchange center	7588	0.00	8944099.28	3959800.11	1028475.09
Zhejiang equity exchange center	3093	16353	417248.66	164218.83	152012.29
Gansu equity exchange center	2244	16000	464167.32	302259.12	32529.56
Wuhan equity custody center	2154	0.00	43316.71	19912.26	18383.01
Guangzhou equity exchange center	1870	3300	62396.12	47492.84	6138.67
Beijing equity exchange center	1765	0.00	404534.95	259568.38	128411.28
Strait equity exchange center	1465	0.00	130682.83	55944.22	487825.15
Liaoning equity exchange center	1128	0.00	1874387.19	307181.09	301953.34
Guangxi Beibu gulf equity center	735	0.00	2051062.88	975735.54	20322.86
Qilu equity exchange center	630	6636.15	792448.28	197742.58	268369.11
Tianjin equity exchange center	528	5463	2902140.41	1425672.28	761876.04
Xinjiang equity exchange center	510	0.00	794447.95	624536.68	24811.05
Shaanxi equity exchange center	407	0.00	2874714.24	548469.71	2580821.85
Chengdu (Sichuan-Tibet) center	338	0.00	7091.92	4893.71	56.08
Jiangsu equity exchange center	281	0.00	1449275.06	754236.32	71365.69
Chongqing share transfer center	255	14450	1317880.58	300509.70	561124.37

4.2. Performance Analysis of Blockchain Equity Trading System. Figure 2 shows the number of features selected by each method, the time it takes to select features, and the accuracy and accuracy of the selected feature subsets in the prediction of the test set and the time-limited training set. It can be seen from the experimental results in Figure 2 that the accuracy and accuracy of the support vector machine model without feature selection are the worst among many

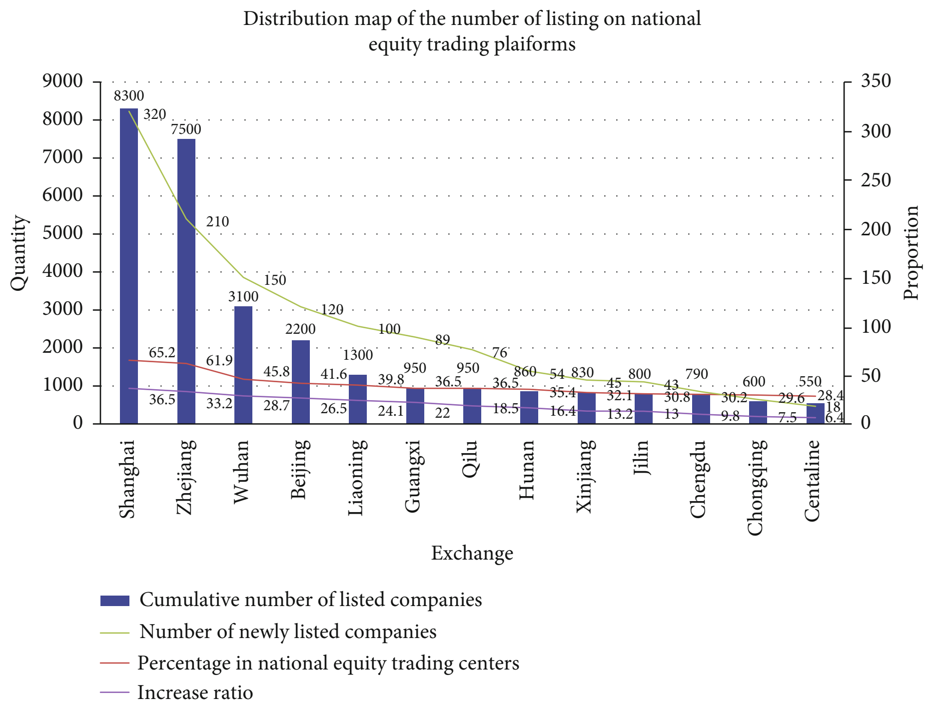

Figure 1: Distribution map of the number of listings on national equity trading platforms.

Table 3: Industry distribution of listed companies.

—	Industry category	Quantity	Proportion
Industry	Manufacturing	274	23%
	Construction and real estate development	188	16%
	Mining industry	38	6%
	Transportation, storage, and post	32	5.8%
Agriculture	Agriculture, forestry, animal husbandry, and fishery	201	17%
Service industry	Information transmission, software, and information technology services	39	3%
	Business and tourism services	132	11.5%
	Wholesale and retail	157	13%
	Science technology and technical services	31	3.8%
	Financial industry	24	2.8%
	Accommodation and meals	14	2.2%
	Culture and entertainment industry	17	2.5%
	Other	10	2.0%
—	Total	1184	100%

methods, and the training time is also the highest. The feature selection process of the PCA-SVM model is very fast, but its accuracy and accuracy are not as good as the GA-SIM model and DFS-BPSO-SVM model.

It can be seen from Figure 3 that under the basic transaction input vector, the Gaussian radial kernel function performs best, and all indicators perform the highest. The polynomial kernel function has the highest accuracy, but its accuracy, recall rate, and F-measure performance are far inferior to the Gaussian radial kernel western number and linear kernel function. The performance of the sigmoid kernel function is the worst, and the performance of all four evaluation indicators is far inferior to other types of kernel functions. Therefore, at the end of this article, we choose to use the Gaussian radial kernel function to build a dynamic model under the basic market transaction input vector system.

Figure 4 shows the usage rate of the five types of consensus mechanism algorithms on the CPU. It can be seen from Figure 4 that in the single-node test, the CPU usage rate of the PoW consensus mechanism algorithm reached 100%, but the improved PBFT consensus mechanism was only

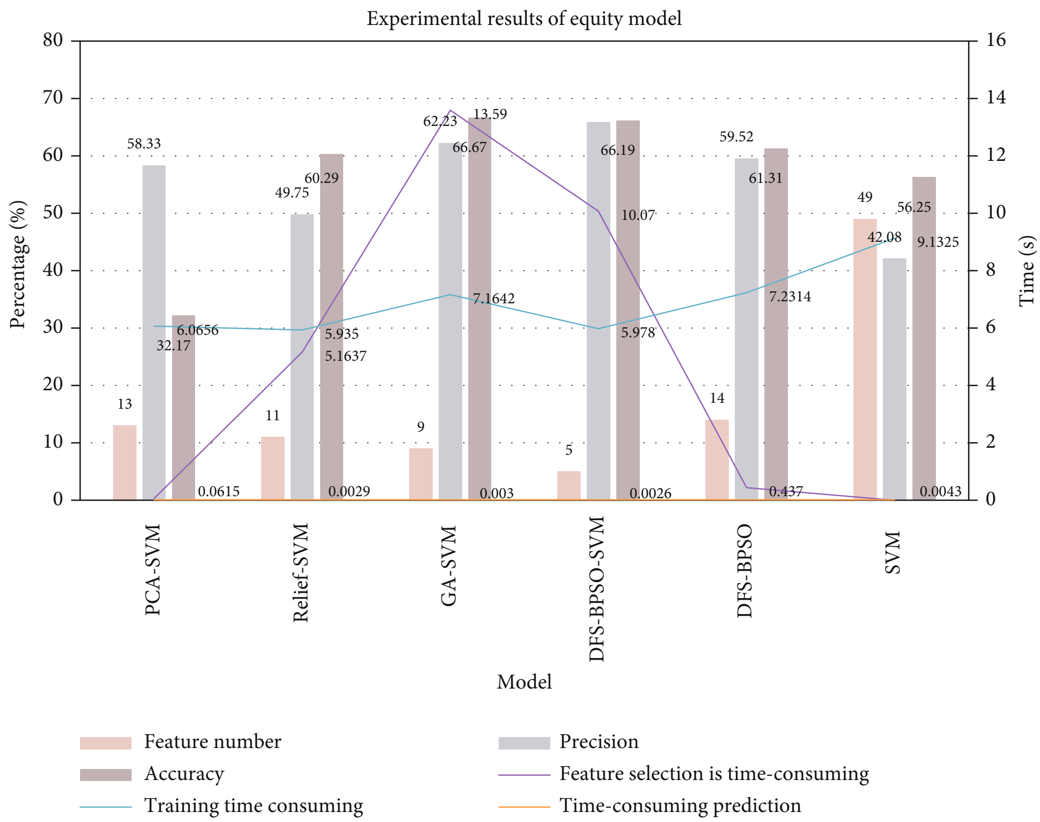

FIGURE 2: Experimental results of equity trading model.

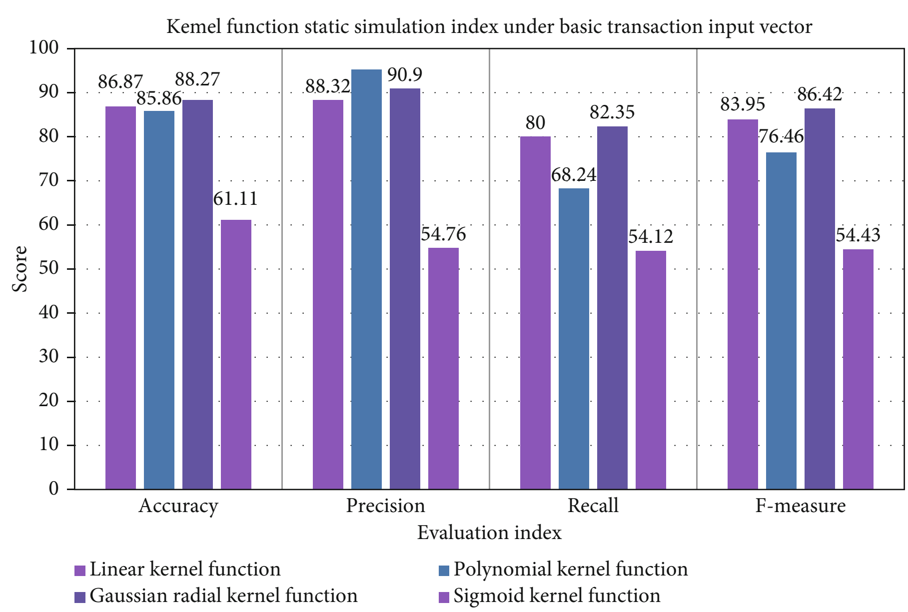

FIGURE 3: Kernel function static simulation index under basic transaction input vector.

16%, which resulted in a large energy saving and reduced computing power, improved computing performance, and the CPU can also run more stable.

Finally, it is experimentally verified that the improved PBFT mechanism can reduce the data transmission when the node has an error, and the result is shown in Figure 5. This paper uses the PBFT mechanism and the improved PBFT mechanism to test, respectively. In a complete process of deleting the certificate, the test result of the certificate transmission network overhead is shown in Figure 5. Among

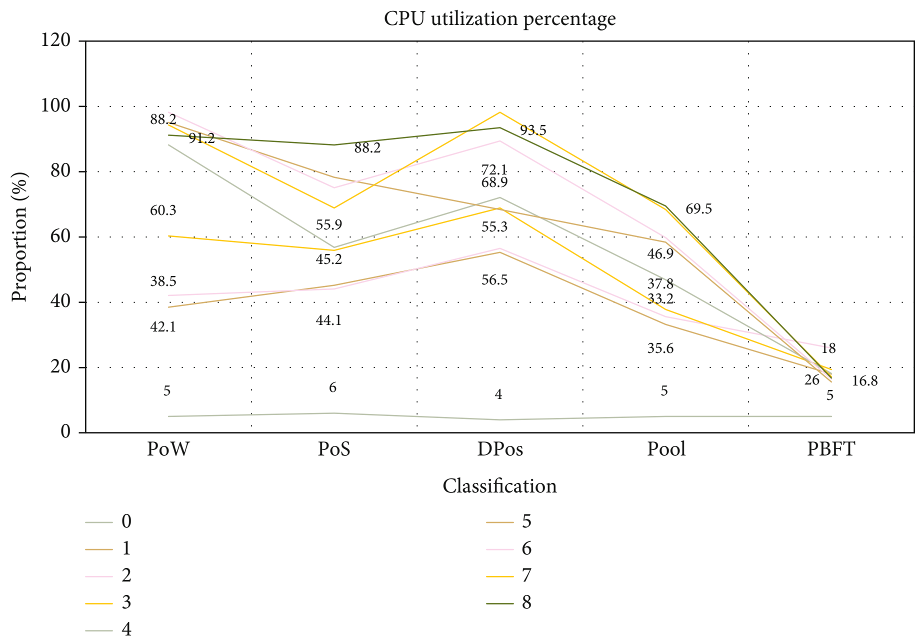

Figure 4: CPU usage.

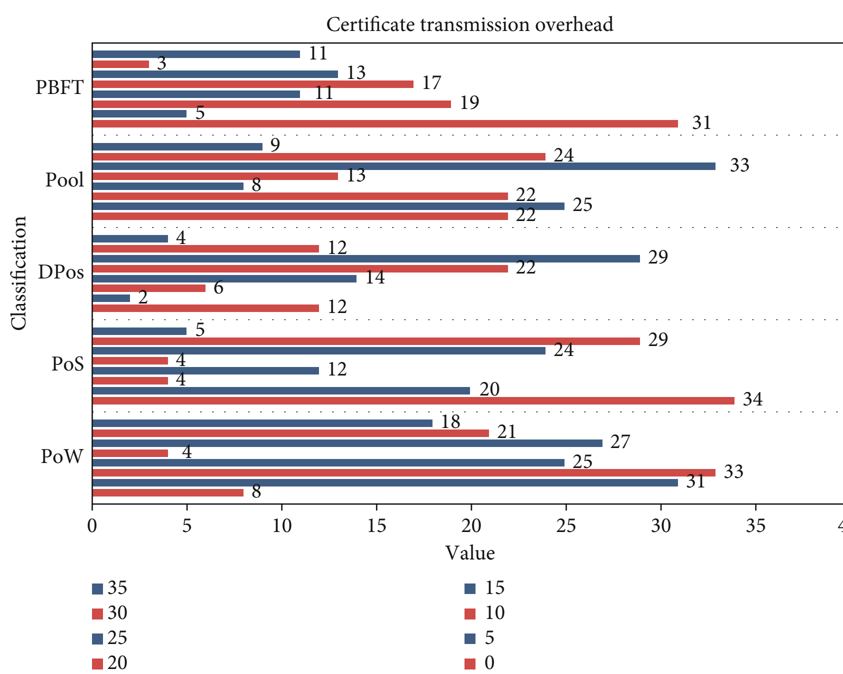

Figure 5: Certificate transmission overhead.

them, the abscissa in the figure is the number of certificates that need to be cleared for a checkpoint. In the figure, it is represented by blockcount, and the ordinate represents the transmission consumption, that is, the number of block hashes that need to be transmitted for each checkpoint execution. As can be seen from the figure, the Ethereum certificate delivery overhead using the PBFT consensus mechanism will increase in proportion to the number of blocks contained in the certificate. This is because each cleanup request needs to transmit the hash value of the block, so the more blocks, the more hash values passed, and the larger the certificate size, which causes a certain network overhead. However, the improved PBFT consensus mechanism is adopted, and there is no certificate transmission, so it is zero.

Figure 6 shows the transmission consumption and overhead required in view switching when an error occurs in the master node. It can be seen from Figure 6 that in the PBFT negotiation mechanism, increasing the maximum number

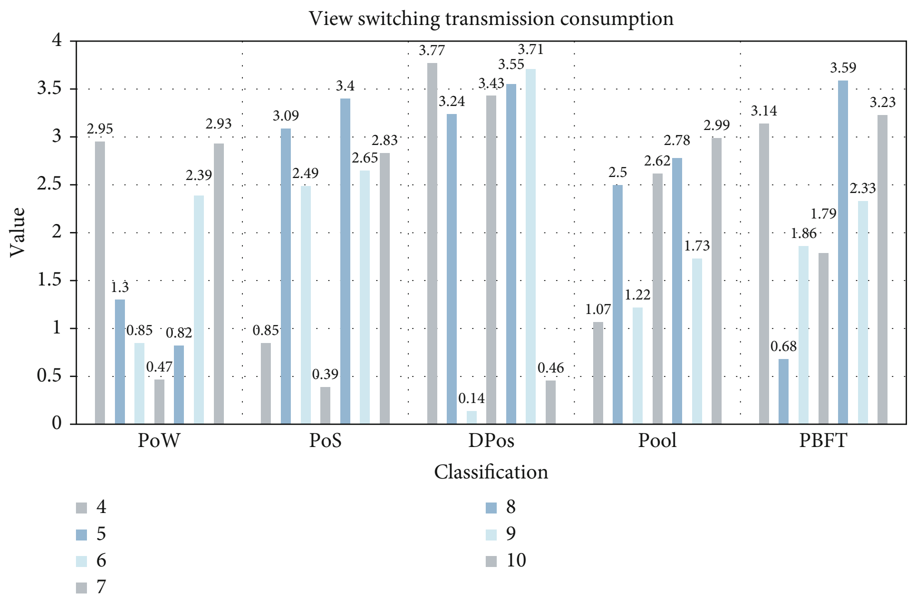

Figure 6: View switching transmission consumption.

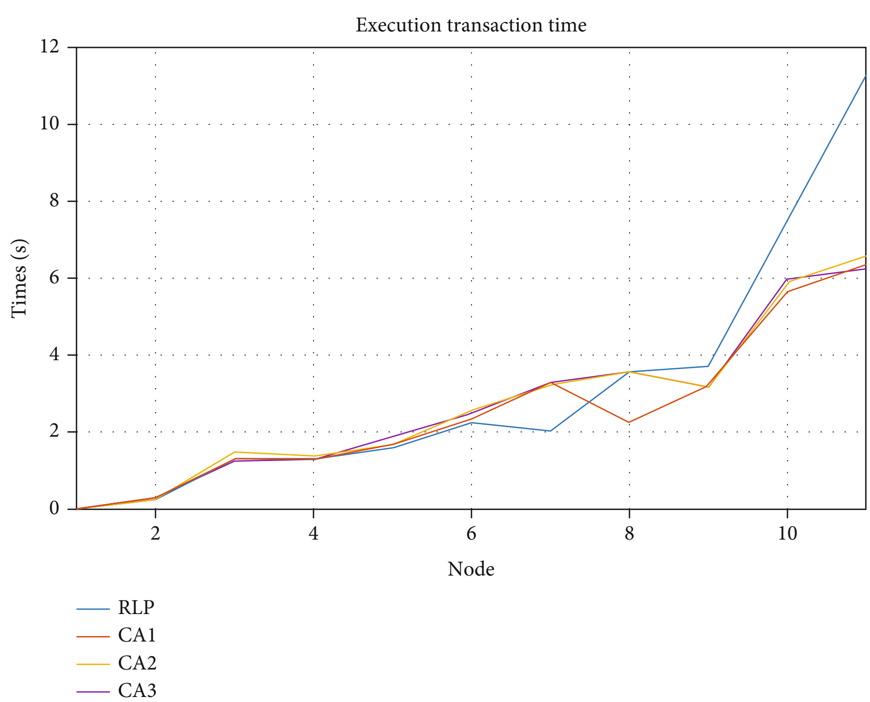

Figure 7: Execution transaction time.

of resistances for node errors will increase the cost, while in the improved PBFT consensus mechanism, this part of the cost is zero. This also shows that compared with the PBFT consensus mechanism, the improved PBFT consensus mechanism will increase the overall cost to a certain extent.

This article separately stores 400 accounts in Ethereum's existing RLP encoding method and adds the clustered account classification to the Merkle Patricia tree for testing. This article performs batch transaction execution according to the transaction information in the sample and repeats the experiment 50 times. Take the average value, and the experimental results are shown in Figure 7. In the figure, CA1 is the K-means clustering algorithm that uses the degree of relevance to select the initial value, and CA2 is the K -means clustering algorithm that uses random sampling to select the initial value. From the figure, it can be seen that the batch transaction execution time is based on the two clustering results. Similarly, so the two clustering results are considered to be similar. For accounts inserted in RLP codes, using the algorithm of two account modifications in one transaction in Ethereum, the time consumption is significantly higher than the execution time of the related account processing, and the execution time of the related account processing is 70% of the RLP encoding processing. Therefore, by optimizing the account storage structure, the performance of the state tree in Ethereum is improved.

5. Conclusion

This article mainly researches the blockchain equity system trading methods and systems based on machine learning and big data algorithms. In the context of machine learning and big data algorithms, the construction of the blockchain equity trading system has innovated the way of equity trading, improved transaction efficiency, and increased operating efficiency. Taking into account the heterogeneity of nodes and access control strategies, when nodes are added or deleted, data distribution may be uneven, and data accuracy may also be affected. Therefore, there is room for further improvement in the scheme proposed in this article. The innovations of this article are the combination of qualitative analysis and quantitative analysis and the combination of theoretical analysis and empirical analysis, which fully and accurately illustrate the comprehensiveness of the system constructed in this article, overcome the shortcomings of traditional trading methods, and provide new trading methods. The methods have broken through the shackles of the equity trading market and promoted the construction of a new equity market. The disadvantage of this article is that the amount of data in the experimental model is less, and more in-depth research and more detailed analysis are needed. It is hoped that the research in this article can provide effective reference and theoretical support for the construction of the equity market.

Data Availability

Data Availability. No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by natural platforms and projects of universities in Guangdong Province: Key field (Rural Revitalization) project of colleges and universities in Guangdong Province (Number: 2020ZDZX1084) and Natural Science Foundation of Guangdong basic and applied research foundation (2114050003056).

References

[1] M. Boehm, M. W. Dusenberry, D. Eriksson et al., "SystemML," *Proceedings of the VLDB Endowment*, vol. 9, no. 13, pp. 1425–1436, 2016.

[2] Z. Obermeyer and E. J. Emanuel, "Predicting the future-big data, machine learning, and clinical medicine," *The New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, 2016.

[3] S. R. Hamzah, N. Ishak, and A. F. N. Rasedee, "Risk shifting elimination and risk sharing exposure in equity-based financing – a theoretical exposition," *Managerial Finance*, vol. 44, no. 10, pp. 1210–1226, 2018.

[4] A. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2017.

[5] C. Helma, T. Cramer, S. Kramer, and L. De Raedt, "Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds," *J Chem Inf Comput*, vol. 35, no. 4, pp. 1402–1411, 2018.

[6] S. Mullainathan and Z. Obermeyer, "Does machine learning automate moral hazard and error?," *The American Economic Review*, vol. 107, no. 5, pp. 476–480, 2017.

[7] M.-D. Yu, M. Hiller, J. Delvaux, R. Sowell, S. Devadas, and I. Verbauwhede, "A lockdown technique to prevent machine learning on PUFs for lightweight authentication," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 2, no. 3, pp. 146–159, 2017.

[8] Y. Chen, T. Chen, Z. Xu, N. Sun, and O. Temam, "DianNao family," *Communications of the ACM*, vol. 59, no. 11, pp. 105–112, 2016.

[9] E. Hesamifard, H. Takabi, M. Ghasemi, and R. N Wright, "Privacy-preserving machine learning as a service," *Proceedings on Privacy Enhancing Technologies*, vol. 2018, no. 3, pp. 123–142, 2018.

[10] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: a new perspective," *NEUROCOMPUTING*, vol. 300, pp. 70–79, 2018.

[11] D. Assouline, N. Mohajeri, and J. L. Scartezzini, "Quantifying rooftop photovoltaic solar energy potential: a machine learning approach," *Solar Energy*, vol. 141, pp. 278–296, 2017.

[12] D. Sacha, M. Sedlmair, L. Zhang et al., "What you see is what you can change: human-centered machine learning by interactive visualization," *Neuro computing*, vol. 268, pp. 164–175, 2017.

[13] M. Zaharia, R. S. Xin, P. Wendell et al., "Apache spark: a unified engine for big data processing," *Communications of the ACM*, vol. 59, no. 11, pp. 56–65, 2016.

[14] E. D. Siew, R. K. Basu, H. Wunsch et al., "Optimizing administrative datasets to examine acute kidney injury in the era of big data: workgroup statement from the 15thADQI consensus conference," *Canadian Journal of Kidney Health and Disease*, vol. 3, no. 1, p. 98, 2016.

[15] S. Athey, "Beyond prediction: using big data for policy problems," *Science*, vol. 355, no. 6324, pp. 483–485, 2017.

[16] M. Chen, Y. Ma, J. Song, C.-F. Lai, and B. Hu, "Smart clothing: connecting human with clouds and big data for sustainable health monitoring," *Mobile Networks & Applications*, vol. 21, no. 5, pp. 825–845, 2016.

[17] S. Underwood, "Blockchain beyond bitcoin," *Communications of the ACM*, vol. 59, no. 11, pp. 15–17, 2016.

[18] D. Kraft, "Difficulty control for blockchain-based consensus systems," *Peer-to-Peer Networking and Applications*, vol. 9, no. 2, pp. 397–413, 2016.

[19] J. Kang, R. Yu, X. Huang, S. Maharjan, Y. Zhang, and E. Hossain, "Enabling localized peer-to-peer electricity trading

among plug-in hybrid electric vehicles using consortium blockchains," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3154–3164, 2017.

[20] H. J. Kim, "Study on equity of taxation for non-residential property by analysis of actual transaction price," *journal of the korean data & information science sociaty*, vol. 27, no. 3, pp. 639–651, 2016.

[21] L. Xu and X. Yu, "Thorny roses: the motivations and economic consequences of holding equity stakes in financial institutions for China's listed nonfinancial firms," *China Journal of Accounting Research*, vol. 10, no. 2, pp. 105–125, 2017.

[22] B. Grdoi, E. Skreli, and D. Imami, "Determinants of sustainable relationships in the Albanian apple production sector," *International Journal on Food System Dynamics*, vol. 7, no. 1, pp. 50–65, 2016.

[23] B. E. James and J. B. Mcguire, "Transactional-institutional fit: corporate governance of R&D investment in different institutional contexts," *Journal of Business Research*, vol. 69, no. 9, pp. 3478–3486, 2016.

[24] F. Baetje and L. Menkhoff, "Equity premium prediction: are economic and technical indicators unstable?," *International Journal of Forecasting*, vol. 32, no. 4, pp. 1193–1207, 2016.

[25] V. Jha, "Timing equity quant positions with short-horizon alphas," *The Journal of Trading*, vol. 11, no. 3, pp. 53–59, 2016.