

Research Article

Influencing Factors of e-Commerce Enterprise Development Based on Mobile Computing Big Data Analysis

Yixue Zhu¹ and Boyue Chai²

¹Yancheng Teachers University, Yancheng, 224007 Jiangsu, China

²Hebei University of Engineering, Handan, 056038 Hebei, China

Correspondence should be addressed to Boyue Chai; chaiboyue@hebeu.edu.cn

Received 15 April 2021; Revised 5 July 2021; Accepted 28 July 2021; Published 27 September 2021

Academic Editor: Wenqing Wu

Copyright © 2021 Yixue Zhu and Boyue Chai. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of increasingly advanced information technology and electronic technology, especially with regard to physical information systems, cloud computing systems, and social services, big data will be widely visible, creating benefits for people and at the same time facing huge challenges. In addition, with the advent of the era of big data, the scale of data sets is getting larger and larger. Traditional data analysis methods can no longer solve the problem of large-scale data sets, and the hidden information behind big data is digging out, especially in the field of e-commerce. We have become a key factor in competition among enterprises. We use a support vector machine method based on parallel computing to analyze the data. First, the training samples are divided into several working subsets through the SOM self-organizing neural network classification method. Compared with the ever-increasing progress of information technology and electronic equipment, especially the related physical information system finally merges the training results of each working set, so as to quickly deal with the problem of massive data prediction and analysis. This paper proposes that big data has the flexibility of expansion and quality assessment system, so it is meaningful to replace the double-sidedness of quality assessment with big data. Finally, considering the excellent performance of parallel support vector machines in data mining and analysis, we apply this method to the big data analysis of e-commerce. The research results show that parallel support vector machines can solve the problem of processing large-scale data sets. The emergence of data dirty problems has increased the effective rate by at least 70%.

1. Introduction

1.1. Background. With the continuous improvement of information technology, the data model has become highly complex, and the scale of data has also grown rapidly. The whole society has entered the era of big data. In recent years, my country's e-commerce has developed vigorously and on a large scale, and it has become the world's largest Internet market. At the same time, with the development of information technologies such as cloud computing and the Internet of Things, rapidly expanding data has brought mankind into the era of big data. It has become possible to use big data technology to analyze the development of the regional e-commerce industry chain and provide corresponding guidance. Due to differences in the economic development status of different regions, the construction of e-commerce infra-

structure, and the online shopping habits of local residents, the differences in the level of regional e-commerce development have been created. The rapid rise of big data has become a link to the ternary world. With the development of big data, the role of e-commerce in the network economy is becoming more and more important, and vigorously developing e-commerce has become the primary task of all countries and regions.

1.2. Significance. E-Commerce is one of the hottest topics nowadays, and it is also one of the fields with the most development potential. Countries around the world have listed e-commerce companies as strategic industries for development. At the same time, due to the rapid development of the network economy, big data will also become a strategic decision for countries to seize development, and the control

and use of data will become the core competitiveness of the future society. Regional e-commerce development research based on big data analysis will enhance the development of e-commerce in the region and better promote governments, enterprises, and other institutions to obtain benefits from big data, so as to lay the cornerstone of the big data industry and improve service capabilities and operation efficiency, etc. The development of e-commerce will inevitably have a huge impact on the organizational structure of traditional enterprises and cause changes in corporate management. The positive significance of e-commerce to the development of small- and medium-sized enterprises is very significant, and the development of e-commerce activities has become an urgent need for small- and medium-sized enterprises. e-Commerce breaks through time and geographical restrictions, creating good conditions for small- and medium-sized enterprises and large enterprises to compete on an equal basis. Especially when information technology is widely popularized and the transfer of technological achievements is accelerated, small- and medium-sized enterprises can use e-commerce to put new products on the market faster than large enterprises.

1.3. Related Work. With the vigorous development of e-commerce in China, more and more enterprises in China have begun to use the powerful capabilities of e-commerce to create new profit growth points, which has enabled many scholars and professionals to carry out in-depth theoretical and practical explorations. But referring to the current research, we found that the research on the factors affecting the profit of e-commerce companies is nonsystematic, unilateral, or inclined to management. Yang pointed out that this article integrates the e-commerce situation, corporate financial model, corporate performance, and other elements into the same overall framework; analyzes the internal connections between different elements; and then perfects related corporate financial management theories. With reference to relevant research results at home and abroad, the influencing factors of financial processing procedures and financial management models of e-commerce enterprises are deeply studied. In addition, a financial management model that is suitable for e-commerce companies and conforms to the characteristics of e-commerce companies has been constructed, but so far, my country's e-commerce economy still lags behind foreign countries in some aspects [1]. Yuen pointed out that this article is based on the Innovation Diffusion Theory (IDT) and analyzed the customer's intention to use self-pickup as the last mile delivery method. It is assumed that innovative characteristics are the key factors that affect customers' willingness to use self-service pickup services. Demographic characteristics were also tested. The survey data was collected from 164 consumers in Singapore and analyzed using hierarchical regression analysis. The results show that among the five key characteristics of innovation, relative advantages, compatibility, and trialability have a positive impact on customers' willingness to use self-service pickup services. The study also found that the first step to improve customer willingness is to integrate self-collection into consumers' lifestyles, values, and needs.

In addition, the marketing method of self-service pickup service should have obvious advantages over other last-mile delivery methods, but for some customers, door-to-door delivery is more satisfactory [2]. Tawalbeh pointed out that mobile devices are increasingly becoming an indispensable part of people's daily lives, facilitating various useful tasks. Mobile cloud computing combines mobile and cloud computing to expand its functions and advantages and overcome its limitations, such as limited memory, CPU power, and battery life. Big data analysis technology can extract value from data with four Vs: quantity, variety, speed, and authenticity. This article discusses the role of online healthcare and mobile cloud computing and big data analysis in its implementation. With the adoption of cloud computing in healthcare, the motivation and development of network healthcare applications and systems have also been revealed. Describe a cloudlet-based mobile cloud computing infrastructure for medical big data applications. The technology, tools, and applications of big data analysis are reviewed, but network delays still exist in the process of remote surgery [3]. In recent years, more and more IT industries have begun to pay attention to the impact of IT applications on the sustainable development of the environment, including companies such as IBM and HP. Simon Mingay defined green IT and believed that IT organizations should not only understand green IT in terms of energy efficiency. But his research does not consider the application of green industry in other areas, which has limitations [4]. Hegmin-Younger et al. made a comparative study on the overall mean imputation method, group mean imputation method, thermal card method, and regression imputation method when analyzing the relationship between college students' performance before and after enrollment. It embodies the content of big data to meet the green challenge. But his research did not introduce specific algorithms [5]. In the research, someone defined a knowledge base of preattack results to realize the causal association method of association analysis. Experiments show that this method can dig out the essential connections between independent alarm information, but the result of the scene construction is very dependent on the expert knowledge base. The completeness and correctness of the knowledge base construction greatly affect the results of the association analysis [6]. With the rapid development of the Internet, IP data services have been rapidly popularized. IP data services are uncertain and unpredictable, and WDM networks use fixed-size wavelengths as the smallest granularity to allocate bandwidth to services. However, due to the diversity of services, it is difficult for WDM networks to adapt to service requests of different granularities, resulting in low spectrum utilization [7].

1.4. Main Content. The main purpose of this paper is to use big data information technology to analyze the influencing factors of the development of e-commerce enterprises, analyze the evaluation results, and combine the development of the domestic economic environment to propose targeted and construction to promote the development of e-commerce industry. Sexual opinions have contributed to the development of the domestic economy. First, the big

data preprocessing technology and methods are introduced in detail. Generally speaking, the data we use are often incomplete and inconsistent and are greatly affected by noise. Data cleaning is the use of various cleaning techniques to remove “dirty data” and obtain “clean” consistent collection of data. The next step is to conduct experiments on the research of big data quality evaluation standards and analyze it through big data prediction in the field of e-commerce. From the perspective of data availability, it can be evaluated from two aspects: accessibility and timeliness. Finally, the use of parallel support vector machines to analyze big data in the field of e-commerce has certain theoretical and practical significance for solving the problems of big data quality evaluation and predictive analysis in the field of e-commerce. From the point of view of the time to predict data, the parallel support vector machine only takes about one-seventh of the time of the traditional support vector machine, which shows the advantages of the parallel support vector machine in processing large-scale data sets. It is more effective than traditional support vector machines. The experimental results show that the parallel support vector machine has a good effect in predicting and analyzing data, both in accuracy and efficiency.

2. Big Data Preprocessing Technology Method

2.1. Data Cleaning. Generally, the data we use often has incomplete information and inconsistencies and is greatly affected by noise. Data cleaning is the use of various cleaning techniques to remove “dirty data” and obtain “clean” consistent collection data [6, 7]. Data cleaning is divided into the following categories:

(1) Cleaning of duplicate data

Delete the duplication in the data volume to ensure the speed and accuracy of data generation. There are two or more instances of the same object, and the data focus is returned as repeated data. Usually, in order to see the meaning of repetition, it is necessary to compare each example one by one, and then determine a common example. Statistical methods are usually used to test the numeric attributes of cells, and the attributes of different attributes and even the default values of different numeric attributes are based. In this way, information of anomalous nature can be determined, and the amount of duplicate data can be determined and avoided [8, 9].

(2) Missing data cleaning

In many fields, missing data is a complex and difficult to avoid problem. Due to improper operation, confidentiality of information, or unreliability of data sources, the contents of the data set are incomplete and incomplete. In data mining, the lack of data will cause many problems [10]. For example, if the wrong model is applied to the grassroots system, there will be serious differences in the evaluation and implementation of the effectiveness of resource management rules and the accuracy of export rules. Many modern methods can be used to clean fluids. Clothing can be divided into two categories:

- (1) *Ignore Incomplete Data.* Incomplete data is ignored, and its characteristics and examples can be deleted. Its advantage is that when the data is too small, too small to collect complete data, the data will be very effective. Its drawback is that when a large amount of incomplete data is in the subset, a smaller number can be obtained after deleting from the file, which can only increase the representativeness of the model and greatly reduce the reliability of the model [11, 12]
- (2) *Missing Value Interpolation Method.* Because the neglect method is likely to delete useful information, after being ignored, useful information may also be deleted to complete incomplete data. How will it make up for these missing values? Replace the value with the value closest to the data, and obtain the obtainable value of the quantity and quality of the data. The supplement will provide potentially valuable data, more information about the sample than the method used by the despised person, and will not lead to deviation of the data analysis results, making the designed model extremely unreliable [13]
- (3) *Noisy Data Processing.* Usually, we assume that the data set is data mining without interference. However, in real life, due to various reasons, in the process of data collection, a large amount of noise data is often generated, which we call “outliers.” If the data range is not reasonable, it may be difficult to ensure the quality of input and output in the data mining process, resulting in inaccurate results. So there are usually two ways to process voice control data. The first is fractal technology. The preliminary data is divided into different containers, and the data is smoothly displayed with corresponding values, including measuring two wide and deep containers. The second method is to use the group method for outlier analysis [14, 15]

2.2. Data Integration. In the context of big data, data sources are often diverse, and data integration is to store data in different databases into a unified data store. Obviously, it is not feasible to forcibly merge two databases of different designs. This often leads to redundancy and inconsistency of data sets. Therefore, how to match the patterns and objects of multiple data sources is the main problem solved by data integration [16]. The following problems are mainly solved in data integration:

- (1) *Mode Integration.* Integrate metadata from different data sources [17]
- (2) *Entity Recognition Problem.* Matching equivalent entities from different databases [18]
- (3) *Data Redundancy.* The same entity will have its corresponding attribute representation in different databases [19]

- (4) *Data Value Conflict*. In the real world, metadata is very different, because there are big differences between different ratios and different codes. For example, the size attribute is based on the metric system in one system and the British system in another system [20]

2.3. *Data Transformation*. Finding the characteristic representation of data is the core of data transformation. Through dimensional transformation or transformation, the number of effective variables is compressed, or the invariant of the data is found, and the data is transformed into a data suitable for various mining modes through methods such as normalization, switching, and projection form [21]. Commonly used transformation methods include the following:

- (1) *Function Transformation*. Use some mathematical functions to map each attribute value
- (2) *Normalize the Data*. Make the data scaled proportionally, and it is easy to reduce the scope [22]. In order to avoid excessive reliance on individual categories and algorithms and correct imbalances, some factors may need to be considered

2.4. *Data Protocol*. In today's information age, with the increasing scale of data sets, higher requirements are placed on data mining technology. However, the analysis and mining of massive data not only increase the complexity of the technology but also greatly extend the mining time, which is a huge challenge for data mining research. In response to this type of problem, the technology of data entry completion can be used to replace a large amount of original data; although its scale is much smaller than the data itself, it also retains the integrity and exclusivity of the original data. For complex and huge data sets, the data reduction step is essential. It can not only effectively reduce mining complexity and mining time but also has good and reliable mining quality [23]. Data reduction techniques mainly include dimension reduction, data compression, discretization, data cube technology, data block reduction, and concept hierarchy generation methods.

3. Big Data Quality Evaluation Standard Research Experiment

First, let us briefly introduce the self-organizing competitive neural network (SOM). The SOM artificial neural network will simulate the function of the human brain neural network. The network consists of an input layer and an output layer. The input neurons are a one-dimensional matrix that receives the input signals of the network. The output layer is arranged by neurons in a certain way into a two-dimensional node matrix. The number of neurons is from the corresponding number. The vector based on the signals received in the network and the neurons from the network stand in a specific way in an X-ray matrix X-ray node target. When the network receives a signal, the neurons in their input layer are connected to the neurons in the output layer.

The network initialization assigns small random numbers to the weight vectors of the output layer and normalizes them to obtain $w(j=1, 2, \dots, m)$, establish the initial winning field $N(0)$, and the initial learning rate η value. m is the number of neurons in the output layer [24].

$$\frac{1}{2}\eta = \sqrt{\frac{N + w_i(a)}{m}}. \quad (1)$$

An input pattern is randomly selected from the training set and normalized to obtain $X(i=1, 2, \dots, n)$, where n is the number of neurons in the input layer.

$$\eta = N_j [x_m - w_{ij}(t)]. \quad (2)$$

Calculate the dot product of X and w , and find the winning node j with the largest dot product. If the input mode is not normalized, the Euclidean distance should be calculated according to equation (2), and the winning node with the smallest distance should be found from it.

$$d_j = \|X_t - W_j\| = \sqrt{\sum_{j=1}^m [X_i - W_j]^2}. \quad (3)$$

Let j be the center to determine the weight adjustment domain at time t . Generally, the initial domain $N(0)$ is large, and $N(t)$ shrinks with the training time during the training process.

$$j \in N_j^*(t). \quad (4)$$

Adjust the weights of all nodes in the winning field $N(t)$

$$w_{ij}(t+1) = w_{ij}(t) + a(t, N). \quad (5)$$

Before introducing structural risk minimization, first introduce the following concept: Suppose VC-dimension h is finite, then the following formula holds with probability $\geq 1 - \eta$:

$$R(a) \leq R_{\text{emp}}(a) + \forall a \in \Lambda, \quad (6)$$

where η is a small integer:

$$\varepsilon = 4 \left[\left(\ln \frac{2}{h} + 1 \right) - \ln \left(\ln \left(\frac{\eta}{4} \right) \right) \right] / \int. \quad (7)$$

For the linearly separable training set, its samples all meet the following conditions:

$$y_i((w \cdot x_i) + b) - 1 \geq 0. \quad (8)$$

Then, looking for the optimal classification surface of the training set Ω is to satisfy the condition of equation (3) to solve the hyperplane equation that minimizes $\|w\|$ [25]. This is a typical constrained optimization problem, so after

TABLE 1: The basic information of 7 numerical variables in the data set.

Statistical magnitude	Collection	Original cost	Price	Discount	Has evaluation
Least value	0	1	1	0.05	0
Lower quartile	186	22	15	0.42	52
Median	524	57	29	0.7	124
Mean	2095	141	71	0.6914	686
Upper quartile	1538	180	94	1.00	318.00
Maximum	900	2380	1850	11.00	5459.7
Missing value	0	8	8.00	0.00	8

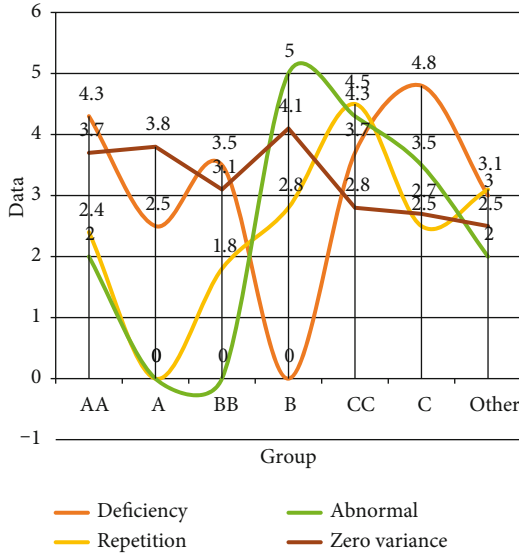


FIGURE 1: The frequency of missing values for only four variables.

transforming the problem of minimization into a problem of minimizing equivalently, equation (9) can be obtained:

$$\min \frac{1}{2} \|w\|^2. \quad (9)$$

Generally, to solve the convex quadratic optimization problem of equation (9), it can be solved by converting it into its dual problem by Lagrange optimization method, as shown in equation (10).

$$\max \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j a_i a_j (x_i \cdot x_j). \quad (10)$$

i in equation (10) is the Lagrange multiplier corresponding to the constraint condition [26, 27]. We can know that the edge hyperplane is only related to a few samples on its plane. These samples are what we call the support vector; that is to say, the Lagrange multiplier k corresponding to the support vector Xk must be greater than 0, and X must be the following conditions:

$$1 - y_i((w \cdot x_i) + b) = 0. \quad (11)$$

From equation (11), we find b in the classification function:

$$b = y_k - W \cdot X_k = y_k - \sum_{a_j > 0} a_j y_j (X_j, X_k). \quad (12)$$

In this way, we can get the expression of the classification function $f(x)$:

$$f(x) = \text{sgn} \left[(W \cdot X) + y_k - \sum_{a_j > 0} a_j y_j (X_j, X_k) \right]. \quad (13)$$

Since it is not necessary to ensure that the training samples are classified correctly in all cases, sometimes we need to weigh the empirical risk and generalization ability. This is the approximate linear separability problem [28]. This kind of problem can be obtained by adding a relaxation factor ξ to the original problem, and formula (14) can be obtained:

$$\min_{w, b, \xi} = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i. \quad (14)$$

The nonlinear separability situation means that there is no hyperplane for a training set that can directly separate the two types of samples. To solve this kind of problem, a nonlinear transformation function, that is, a kernel function can generally be used to map the training sample data to a high-dimensional space to make it linearly separable [29]. The optimization problem after introducing the kernel function can be expressed as:

$$\sum_{i=1}^N y_i a_i = 0. \quad (15)$$

4. Big Data Predictive Analysis in the Field of e-Commerce

4.1. e-Commerce Mobile Computing. The data used in the experiment in this article comes from the “searching glasses sales ranking 100 pages of baby data” provided by a data provider. The purpose of this experiment is to use the support vector machine parallel method to predict the sales of glasses. Although before doing the experiment, we hope that the larger the amount of experimental data, the better, but

because the data acquisition is really difficult, in the end, we only obtained about 5,000 data. Some of these 5000 data have certain problems. We need to analyze the entire data set through the following data processing and quality evaluation. First of all, let us take a look at the situation of the collected data. There are 5,000 pieces of data collected and 7 variables in total. Two of the seven variables are text variables: credit and commercial name; the other five variables are numerical variables: collection, original price, price, discount, and postage; therefore, we first analyze the seven numerical variables in the data set that makes a summary, as shown in Table 1:

We know that some of the product names in the product name are repeated, and the highest frequency reached 5, indicating that the data we have taken may be repeated, and we need to delete the repeated values; from the perspective of credit, the main points are there are six types of “AA, A, BB, B, CC, C,” so we need to convert the data to facilitate our next analysis. In summary, we know that the data set contains at least missing, duplicate, outlier, and zero variance phenomena, so we need to clean the data to get a high-quality data set that can be used for analysis.

As shown in Figure 1, when we know that there are missing values among the 7 variables, the frequency of missing values for only four variables is 0, especially for the description score, service score, and shipping score as high as 24 missing values. The maximum and minimum values of the variable of the main credit ratio are both 8, which means that all the data of the variable is 8, which means that the variable has zero variance and is invalid information and the description points, service points, and shipping points. The highest value of is greater than 5, and the minimum is basically above 4.4, but according to the rules, we know that the score is between 0 and 5, indicating that the sampling information is not accurate, and some data has abnormal values. Therefore, data cleaning must be necessary for the data, and the “dirty data” can be filtered out before the following analysis can be carried out. In the next step, we will perform a summary analysis on the two text variables.

This section is to provide data preprocessing and quality evaluation for the following data analysis work. First, we preprocess the 5000 “glasses” data we have collected and finally get 3953 data that can be used for quality evaluation. Next, the quality of this part of the data is evaluated. Each quality dimension has its corresponding quality element, and the quality element corresponds to its index. The purpose of establishing an evaluation baseline is to compare the data set with the quality dimension after data preprocessing to see if it meets the evaluation baseline. If it meets the evaluation baseline, it can be used for data analysis. Otherwise, data collection must be performed again. The specific process is shown in Figure 2.

In the process of data preprocessing, each step is to process the data set into a data set that can be used for parallel SVM predictive analysis. For example, in the data cleaning process, we cleaned the “Taobao” data set for missing values, duplicate values, outliers, and zero variance; in the data transformation process, we processed the data in the “credit” dimension numerically; data specification in the link, we

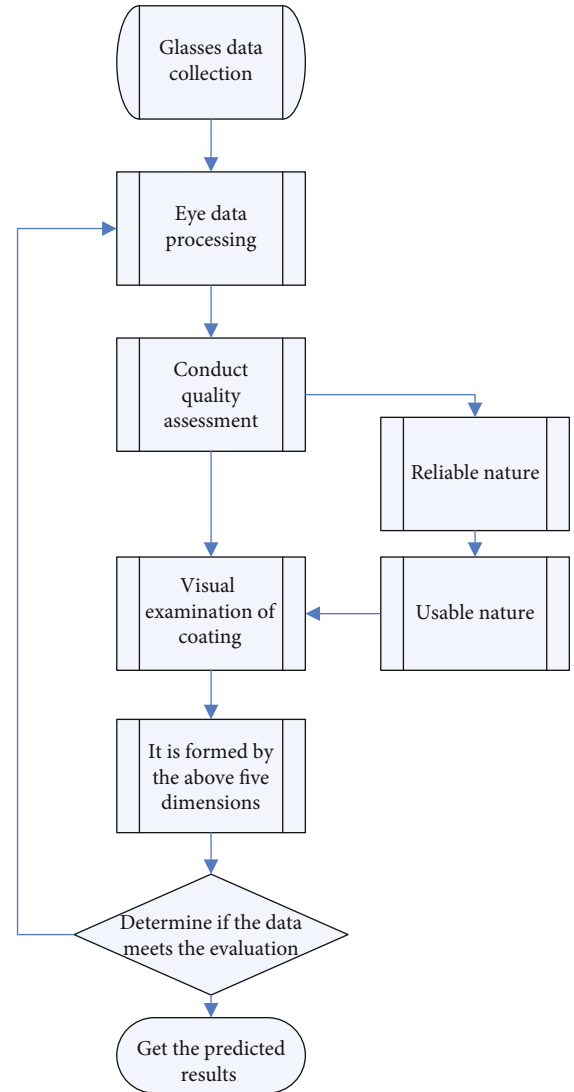


FIGURE 2: Glasses data quality evaluation and predictive analysis process.

TABLE 2: The basic situation of two textual variables in the data set.

Product name	Frequency	Credit	Frequency
Bow	24	Star	1372
Sun glasses	21	4 drill	519
Spectacle frame	20	3 blue crown	438
Polariscope	18	5 drill	418
Buy frames	17	1 blue crown	322
Authentic sunglasses	17	2 blue crown	328
(Other)	20	(Other)	1598

made a correlation analysis and deleted the dimension that has a lower correlation with the “30 glasses volume” dimension. In the end, we get 1598 data with a dimension size of 7. This part of the data can be used in the following data quality evaluation, as shown in Table 2.

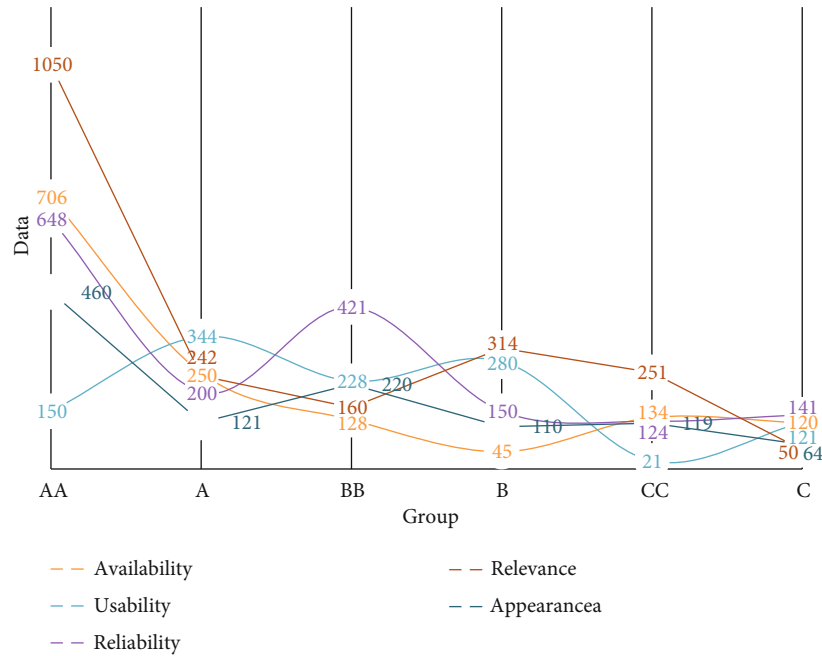


FIGURE 3: Further numerization is carried out according to the level of credit.

TABLE 3: The shows that the data has good timeliness.

Construction stage	Cost management content	Calculation basis	
Investment decision stage	Project proposal, feasibility study, and preparation design task book	Prepare investment estimates	Investment estimation indicators, previous similar project cost data
Design phase	Initial design	Preparation of the overall design budget	Preliminary design drawings, relevant budgetary quotas, or budgetary indicators
Bidding stage	Bidding	Compile the block price (mark bottom) to determine the contract price	Construction drawings, consumption quota, and relevant funding standards
Construction stage	Contract/project implementation	Control cost, stage settlement	Control according to the contract price
Completion acceptance stage	Completion acceptance	Completion settlement/final accounts	Completion settlement (final accounts) and other documents

Here is a point to explain, because the data we collect comes from this e-commerce platform; the method of data integration is not used in the data processing process. However, this method is also indispensable when analyzing other e-commerce big data. For example, when we need to analyze the sales of “iPhone X” mobile phones on these two e-commerce platforms in a certain quarter, we must use data integration technology. Integrate data from two different platforms for analysis.

As shown in Figure 3, the frequency of “AA” credits is the largest, namely, 1050, followed by the frequency of “BB” credits reaching 421, followed by “A” and “B”; the lowest frequency is “CC” and “C.” From the distribution point of view, the credit distribution of the sampled data is relatively uniform. Therefore, the digitization is further carried out according to the level of credit, the value of high credit is high, and the value of low credit is low, and the final digitization result is as above.

4.2. Forecast and Analysis of the Development of e-Commerce Enterprises. In this section, we evaluate the quality of the “clean” data obtained after data preprocessing. The method of quality evaluation is to compare the data to be evaluated with the evaluation baseline formed by the five dimensions of availability, usability, reliability, relevance, and appearance quality. From the perspective of data availability, it can be evaluated from two aspects: accessibility and timeliness. Accessibility usually refers to the difficulty for users to obtain data. This data is the sales data of a certain product. It is an open that the data can generally be obtained through Internet search or data providers, so the data accessibility is relatively high; timeliness usually refers to whether the data can be quickly obtained during the collection process, and the data is for the current task. Whether it is the latest data, after communicating with the “glasses” data supplier, it shows that the data has good timeliness, as shown in Table 3.

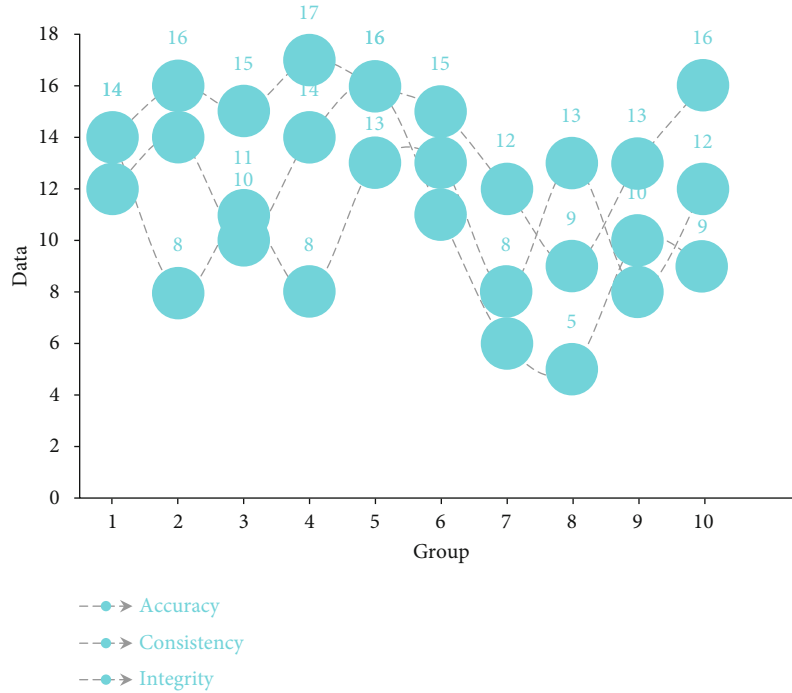


FIGURE 4: From the reliability point of view generally data.

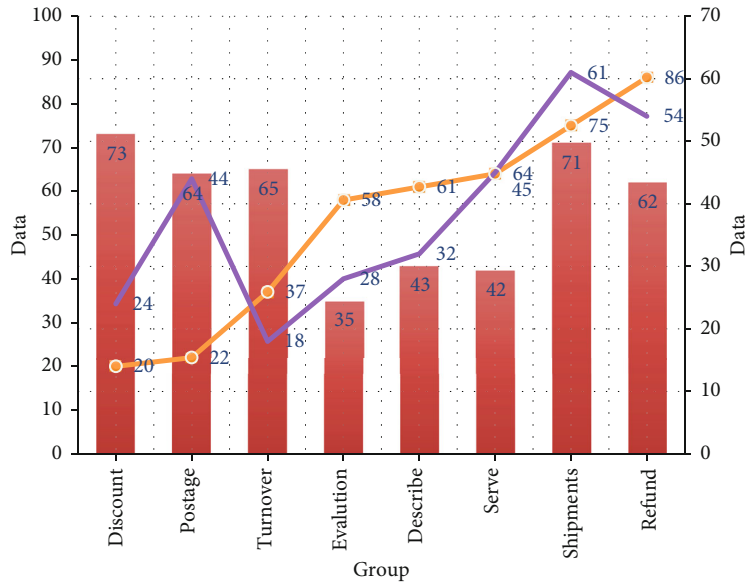


FIGURE 5: Relevance refers to whether the data is relevant to the task at hand.

As shown in Figure 4, from a reliability perspective, reliability generally refers to how much we can trust this part of the data. We searched the top 10 eyewear shops in the data set on the Internet and found that by comparing some of the data, the data is basically consistent with the real data, which shows that the data set is very reliable. From a usability perspective, we can see progress in three areas: accuracy, consistency, and completeness. Accuracy means that the real data and complete information next to the data are free of incompleteness and omissions. For the study of these two

indicators, we use data cleaning in data preprocessing to eliminate duplicate data, missing data, and abnormal data, and irrelevant data is processed, and the “clean” data obtained meets these two indicators; consistency usually indicates whether the logical relationship between data and data is clear. Our research found that the data dimensions in the “glasses” data set are in addition to “credit.” The data of the dimensions are all expressed in numerical values. In order to facilitate the data prediction and analysis in this section, we use the data transformation in the data

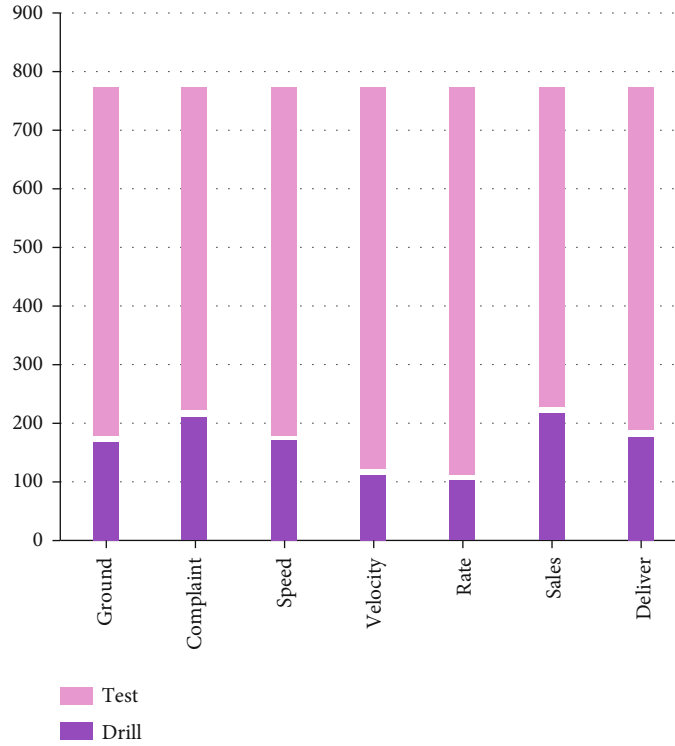


FIGURE 6: Glasses data quality evaluation and predictive analysis process.

TABLE 4: Kernel parameter P and penalty parameter C are used as training.

	Parameter	Search	Punishment	Optimal	Scope	Comprehensive
Parallel	1.5	2.5	4.5	-1.5	6.0	4.5
SVM	2.5	3.0	2.0	7.5	4.0	2.5
LIBSVM	7.5	6.0	4.5	5.0	5.0	10.0

preprocessing to numerically process the data expressed in Chinese characters in the dimension “credit” to ensure the data consistency.

As shown in Figure 5, from the perspective of relevance, relevance refers to whether the data is suitable for the current task. The purpose of this experiment is to predict the “30-day sales of search glasses,” so in order to make the prediction results more accurately, we conducted a correlation analysis on the data set through the data specification in the data preprocessing. Through correlation analysis, we selected the dimensions related to sales volume for analysis, while other dimensions were deleted. The remaining dimensions are “evaluated, discount rate, postage, 30-day transaction volume, service points, delivery Goods points, description points, and the refund rate in the past 30 days”; from the analysis of appearance quality, it is used as an additional attribute to improve user satisfaction. It usually indicates whether the data can be clearly understood by people. From the data we get from the description point of view, it includes “collection, reputation, page views,” etc., which are all easy-to-understand terms, and the content of the data is also very clear, all expressed by numerical values, so the data has a high appearance quality. After data quality

evaluation, we finally got a total of 3953 useful data with a dimension of 8.

As shown in Figure 6, the sales volume of glasses can be predicted by analyzing the dimensions related to the “30-day trading volume of search glasses.” We divide the experimental process into two steps. The first step is to use the training algorithm to get the training model, that is, to find the hyperplane that classifies the data. The second step is to judge whether the training model is accurate by predicting the test set in the data set, so as to verify whether the training algorithm is effective. In the data set of the experiment, the data is divided into two parts: the training set and the test set. The first 70% (2767) data are used as the training set data, and the last 30% (1186) data are used as the test set data.

This experiment was carried out on a PC (2.40 GHz CPU, 16GB memory), the experiment platform was MATLAB 2017b, and the experiment environment was the LIBSVM toolbox. All experiments use Gaussian kernels, and the kernel parameters and penalty parameters are determined by grid search. The grid search range of the kernel parameter P and the penalty parameter C is between -10 and 10, and the step size is 0.5. Take each job, the kernel

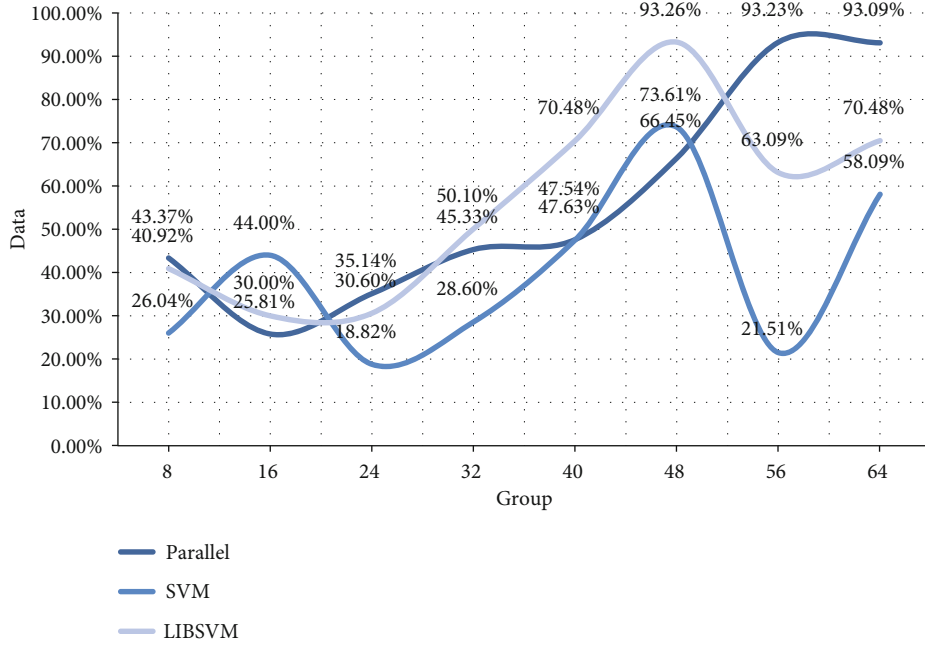


FIGURE 7: Computers will first go through the SOM algorithm.

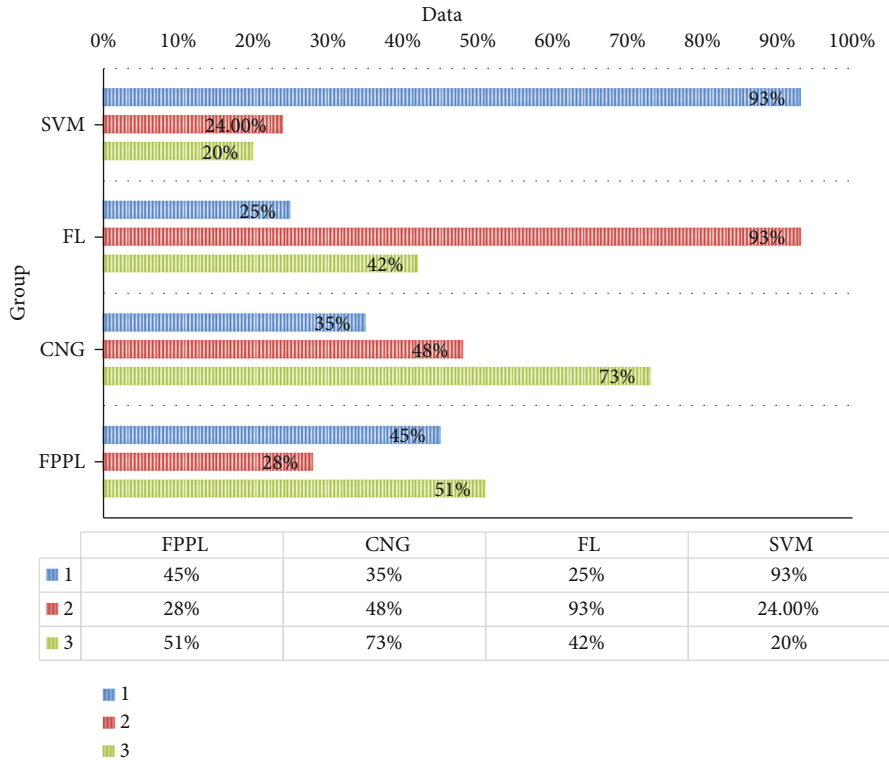


FIGURE 8: Through the comparison of the above two groups of experiments.

function parameter P and the penalty parameter C with the highest corresponding accuracy on the subset are used as the optimal parameters for training, as shown in Table 4.

Before doing the parallel SVM prediction model, we first use the SOM self-organizing neural classification algorithm to group the 2767 data in the training set and set the maxi-

num number of cycles T to 64. According to the SOM self-organizing neural classification algorithm, we get 64 sets of data, according to parallel algorithm, assign these 64 groups of data to a computing core of parallel operation and run the SVM fitting algorithm to find the support vector, and get each support vector subset SV. The support vector subsets

are combined to obtain the support vector set SV, and the SVM fitting algorithm is used for it, and the grid search method is used to select the optimal kernel parameter P and the penalty parameter C . After that, we use the obtained optimal kernel parameter P and penalty parameter C to train the entire support vector set SV to obtain the prediction model of the parallel SVM. Finally, the model is used to predict 1186 data on the test set, and the prediction result is obtained.

In order to verify the efficiency and accuracy of the support vector machine parallel algorithm, we also used traditional SVM to perform prediction experiments on this data set. The experimental configuration all adopts the default configuration of LIBSVM, and the experimental results are as follows:

As shown in Figure 7, when performing parallel prediction, we found that the computer's memory usage has increased significantly. This is because when the parallel algorithm is executed, the computer first allocates 64 sets of data obtained through the SOM algorithm to the computer's 64 operation core, and fitting at the same time, the whole running time is 106 seconds. We first used the obtained prediction model to predict 2767 data on the training set with a prediction accuracy of 93.26%; then, we predicted 1186 data on the test set with a prediction accuracy of 70.48%.

When using traditional SVM for predictive analysis, the computer's memory occupancy has almost no change. This is because the serial algorithm used by traditional SVM does not occupy too much computer cores, and the entire system runs for 779 seconds. Similarly, we first use the prediction model obtained by traditional SVM to predict 2767 data on the training set with a prediction accuracy of 93.09%; then, we predict 1187 data on the test set with a prediction accuracy of 58.09%.

As shown in Figure 8, through the comparison of the above two sets of experiments, it is found that the prediction accuracy of the two sets of experiments on the training set is about 93%, which can show that the prediction models of parallel SVM and traditional SVM are accurate. From the comparison of prediction accuracy, the accuracy of the parallel SVM is much higher than that of the traditional SVM. This may be because the parallel SVM extracts a large number of useful support vectors during the execution process and deletes the support that has no influence or side effects on the classification results. Vector thus ensures that the classifier has good generalization performance, and from the time of predicting data, parallel SVM only takes about one-seventh of the time of traditional SVM, which shows that parallel SVM is in processing large-scale data sets. It is more efficient than traditional SVM. The experimental results show that the parallel SVM plays an excellent role in predicting and analyzing the data, both in terms of accuracy and efficiency.

5. Conclusions

With the development of information technology, big data has become popular in all industries, and the great value of

big data has been widely recognized. Bad data leads to inefficiency and often leads to serious wrong decisions, and the quality of existing big data does not even have a standard. High-quality data is important for analyzing and using big data and ensuring its value. How to ensure the quality of big data and how to ensure the information and knowledge behind the data have become an important issue in the industry and the scientific community. In addition, in order to analyze the big data in the field of e-commerce, this topic uses support vector machines for research. Although there are many documents on big data research at present, there are still few studies on the quality evaluation of big data, and the research level of this article is limited. There are many shortcomings. In terms of quality standards and evaluation processes, they should be measured according to the standards of authoritative experts. Especially people like big data are still in the cutting-edge technology of the continuous cognitive process. As time progresses, the amount of data will inevitably become larger and more complex, the dimensions and indicators will inevitably have errors, and the evaluation system will not be able to adapt to the future development of data, but I believe that these issues will continue to be studied in the future has been solved.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] J. Xu, Z. Luo, H. He, and Y. Zhou, "Research on the spatial characteristics and influence factors of the development of e-commerce in China's county areas," *Shanghai Urban Planning*, no. 2, pp. 90–97, 2017.
- [2] N. Wang and X. Wang, "The mobile online shopping behavior execution intent prediction model based on situational big data," *Computer Age*, vol. 322, no. 4, pp. 30–33, 2019.
- [3] H. Yuhua, "Distributed big data's group behavior pattern mining algorithm," *Yulin Normal College Journal*, vol. 40, no. 2, pp. 151–157, 2019.
- [4] J. Wu, S. Guo, H. Huang, W. Liu, and Y. Xiang, "Information and communications technologies for sustainable development goals: state-of-the-art, needs and perspectives," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2389–2406, 2018.
- [5] J. Wu, S. Guo, J. Li, and D. Zeng, "Big data meet green challenges: big data toward green applications," *IEEE Systems Journal*, vol. 10, no. 3, pp. 888–900, 2016.
- [6] R. Atat, L. Liu, J. Wu, G. Li, C. Ye, and Y. Yang, "Big data meet cyber-physical systems: a panoramic survey," *IEEE Access*, vol. 6, pp. 73603–73636, 2018.
- [7] H. Song, J. Bai, Y. Yi, J. Wu, and L. Liu, "Artificial intelligence enabled Internet of Things: network architecture and spectrum access," *IEEE Computational Intelligence Magazine*, vol. 15, no. 1, pp. 44–51, 2020.

- [8] Z. Miao, "The design and implementation of the mobile terminal precision marketing system based on big data analysis," *Electronic Design Engineering*, vol. 26, no. 21, pp. 71–76, 2018.
- [9] D. Xiaomei, J. Wang, and Z. Lihong, "Mobile big data analysis algorithms based on deep learning and spark computing," *Journal of Changsha University*, vol. 34, pp. 13–18, 2020.
- [10] C. Haixuan, "Heterogeneous real-time computing architecture analysis based on mobile Internet big data," *Automation Technology and Applications*, vol. 39, pp. 48–51, 2020.
- [11] L. He, "Health data management and analysis system research based on edge computing," *Electronic Design Engineering*, vol. 28, pp. 60–64, 2020.
- [12] W. Wu, Z. Long, and S. Chao, "MEC-based indoor positioning of big data applications," *Telecommunications Science*, vol. 35, no. S2, pp. 207–211, 2019.
- [13] Z. Tianxuan, "Spatial distribution statistics of urban population based on mobile communication big data," *Computer and Modernization*, vol. 273, pp. 49–53, 2018.
- [14] J. Peng, Y. Wang, and C. Fudong, "Power mobile Internet of Things information security terminal architecture based on big data analytics," *Electrical*, no. 1, pp. 63–66, 2017.
- [15] L. Guofeng, "Research based on mobile Internet big data analytics platform," *Consumer Guide*, no. 19, pp. 43–44, 2017.
- [16] J. Wang and S. Sai, "Build a big data analytics system based on the whole process of purchasing," *Digital Design*, vol. 6, no. 3, pp. 12–14, 2017.
- [17] X. Wei, "Cloud-based big data mining system building analysis," *Chinese and foreign entrepreneurs*, vol. 673, no. 11, pp. 116–116, 2020.
- [18] H. Cong, "Study on the route tracking of urban rail trains based on big data analysis," *Modern Electronics*, vol. 41, no. 508, pp. 110–115, 2018.
- [19] S. Lin, T. Jin, Y. Cheng, and Y. Jin, "Wireless big data analytics user QoE perceived improvement strategy," *Communications Technology*, vol. 51, no. 10, pp. 2418–2422, 2018.
- [20] Y. Cheng and G. Wu, "Empirical study of intelligent bracelet big data analysis system based on "T-in-One" ecological chain," *Software*, vol. 39, pp. 30–33, 2018.
- [21] L. Zexin, "Smart city systems based on big data," *Communications World*, vol. 26, no. 1, pp. 79–80, 2019.
- [22] S. Wen, "To analyze the application of big data analysis in the optimization of mobile communication networks," *Global Markets*, no. 17, pp. 180–180, 2017.
- [23] Z. X. Lu, "Power mobile Internet of Things (IoT) information security terminal architecture analysis based on big data," *Electronic Design Engineering*, vol. 28, pp. 159–163, 2020.
- [24] Y. Smelting and L. Yangwu, "Research based on the ocean aquatic big data price analysis and prediction system," *Wireless Connected Technology*, vol. 17, pp. 44–45, 2020.
- [25] C. Long and H. Danxuan, "Big data processing and analytics systems based on the hadoop framework," *Fujian Computer*, vol. 34, no. 5, pp. 120–121, 2018.
- [26] O'Jianlin, "Hadoop-based commercial bank big data platform research and implementation," *China Financial Computer*, no. 1, pp. 50–53, 2019.
- [27] S. Wang, Z. Bao, and L. Chenxuan, "Spark-based big data analytics technology," *Computer Systems Network and Telecommunications*, vol. 1, no. 2, pp. 254–258, 2019.
- [28] W. Yang, S. Xiaoxing, H. Luo, and Y. Xuan, "Based on operator big data public opinion innovation application research," *Jiangsu Communications*, vol. 33, no. 1, pp. 70–72, 2017.
- [29] Y. Wang, "Based on the innovative analysis of enterprise financial management in the era of big data," *Tax*, vol. 12, 2018.