

## Research Article

# Collocating Recommendation Method for E-Commerce Based on Fuzzy C-Means Clustering Algorithm

LiJia Wang<sup>1</sup> and Yanyan Jiang <sup>2</sup>

<sup>1</sup>Department of Economic Management, Xuchang Electrical Vocational College, Xuchang 461000, Henan, China

<sup>2</sup>Shanghai University of Finance & Economics Zhejiang College, Jinhua, Zhejiang 321013, China

Correspondence should be addressed to Yanyan Jiang; [z2011183@shufe-zj.edu.cn](mailto:z2011183@shufe-zj.edu.cn)

Received 10 January 2022; Revised 4 March 2022; Accepted 7 March 2022; Published 18 April 2022

Academic Editor: Miaochao Chen

Copyright © 2022 LiJia Wang and Yanyan Jiang. 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.

In order to reduce the time for customers to select commodities they are interested in, improve the purchase efficiency, improve the success rate of sales of merchants, and create greater economic benefits for enterprises and merchants, this project collects information and data of e-commerce users, using neural network model to analyze and mine data characteristics and shopping records of e-commerce users. According to the analysis results, a user commodity recommendation system based on e-commerce is implemented by using data mining technology. Through the combination of database technology, the transaction and browsing data generated in the process of e-commerce transactions are collected. The collected data is preformatted and used as the input of data mining. Then, it uses data mining technology to mine and analyze the commodities that users are interested in, makes matching according to the types of commodities, and recommends the commodities that users are interested in under a given scene according to the established prediction model. By combining fuzzy clustering with collaborative filtering algorithm, this paper recommends the products that users are interested in, which are mined from historical data and commodity information.

## 1. Introduction

The popularity of e-commerce has brought great convenience to people's life. By studying the behavioral characteristics of e-commerce users, we can provide users with high-quality services and greatly improve the success rate of e-commerce transactions. Therefore, analyzing the characteristics of e-commerce users' behavioral data is a key lesson in the current field of E-commerce [1]. E-commerce is a form of business with information network technology as the means to achieve the exchange of goods as the purpose [2]. It is a network based on the Internet, local area network, value-added network, and other network forms, not to the traditional monetary goods as a means, a kind of electronic transaction, to achieve the value exchange of the network, electronically. Because of the transformation of e-commerce transaction mode, the trading area is broader, showing a state of globalization. With the gradual expansion of e-commerce market scale, it not only provides users with a

wide range of goods, but also provides users with more abundant choices [3]. However, in the face of such a variety of commodity information, how users can quickly and accurately select the commodities they need has also become a topic of concern for users and e-commerce [4, 5]. In order to reduce the time to select interested in goods from the customers, improve the efficiency of purchasing, and improve business sales success rate for the enterprise and business to create greater economic benefits of this topic through e-commerce user information and data collection, analysis of mining electrical characteristics and shopping records, and user data, according to the analysis results using data mining technology for users to recommend commodities, a user product recommendation system based on e-commerce is implemented [6–9].

There are two main forms of product recommendation, conventional recommendation and personalized recommendation, as shown in Figure 1 [10]. Routine recommendation is relatively common, which means that

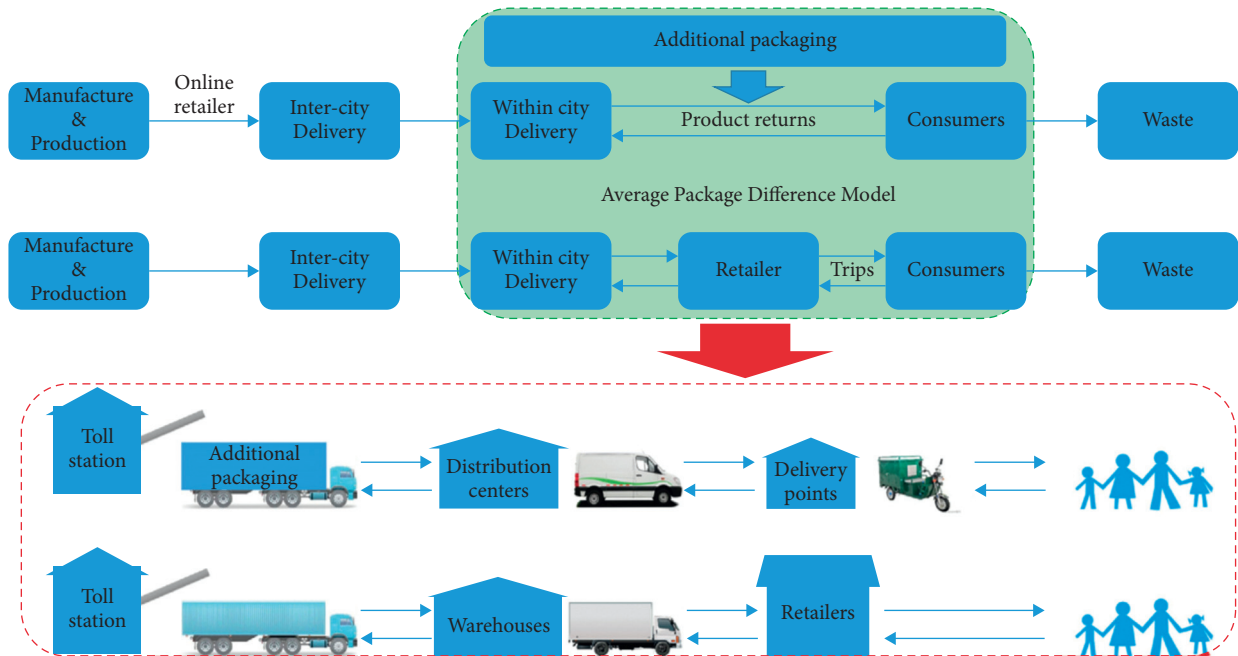


FIGURE 1: Schematic diagram of the theoretical framework of electronic commerce preparation.

merchants display certain mainstream specific products in the recommendation position and then associate similar products [11]. For example, if consumers buy air conditioners on the platform, the system will automatically recommend refrigerators for them. Personalized recommendation is different from this. It mainly analyzes consumers' shopping habits and takes the characteristics of products as the main recommendation indicators. For example, consumers can choose to recommend options for other products that they have purchased, so as to realize the correlation between human nature and physical nature [12]. The most common product recommendations are as follows: guess you like at the bottom of the home page, like the shopping cart, and recommend other products you will be interested in after shopping. It is not easy to analyze the user behavior in view of the huge e-commerce user group, so it is very necessary to analyze the user behavior characteristics accurately by using efficient data analysis method in the e-commerce user behavior data [13]. As an effective method of data mining, cluster analysis has the characteristics of unsupervised learning, which can find hidden behavior patterns in a large number of irregular and noisy e-commerce user behavior data, presenting the characteristics of hidden e-commerce user behavior data, and providing theoretical support for people's decision and judgment. However, the traditional *K*-means clustering algorithm has the defects of large amount of calculation and long time to obtain the optimal solution [14].

## 2. Importance of the Applied Clustering Algorithm

With the development of Internet finance, it has brought more convenience to people's life and provided more finance

for the development of the e-commerce [15]. In e-commerce trade, because it takes the Internet as the necessary medium of exchange, buyers and sellers can realize transtemporal and transregional transactions, instead of face-to-face transactions. The main forms of e-commerce are online shopping, online electronic payment, online transaction communication among merchants, etc., and the related business finance and comprehensive service activities have gradually become a new operation mode [16]. At present, with the rapid rise of e-commerce, e-commerce activities around the world increase; relatively large Internet companies have built a transaction system with e-commerce as the core platform. The establishment of e-commerce platform will attract more consumers, improve the platform database, and lay a foundation for cross-border integration and cooperation of large enterprises. Therefore, with the rapid development of e-commerce today, the traditional ERP system can only meet the most basic and simple needs of e-commerce in the transaction process. With the increase of various user needs and various product information, it has become more and more important to recommend goods on the basis of traditional e-commerce mode [17].

The continuous expansion of e-commerce market has brought abundant and diversified commodities to consumers and also increased the choice of commodities in the electronic field [18]. However, the ensuing problem is that the diversified products impact the eyes of consumers, making them unable to use effective time to quickly choose the products they need at present. At present, the most popular recommendation algorithms mainly include the following: (1) recommendation algorithms based on product content. The recommendation method mainly analyzes the products properties to produce the newly listed products. However, this method has some limited needs for

personalized development. (2) Algorithms based mainly on commodity knowledge: this algorithm system needs to analyze the knowledge level of consumers and other relevant information, and at the same time, it also needs to make use of the related characteristics of knowledgeable and specialized products, which mainly depends on the information exchange state between consumers and system products. (3) The recommendation algorithm of coordinated filtering is mainly used [19]. As the most classical type of recommendation algorithm, the improved traditional collaborative filtering recommendation algorithm includes online coordination and offline filtering method. At present, the collaborative filtering is still the most widely used recommendation algorithm and plays an extremely important role in the recommendation field. In order to reduce the time for customers to select goods they are interested in, improve the purchase efficiency, and improve the success rate of business sales and create greater economic benefits for enterprises and businesses, this topic designs and realizes the recommendation system of user goods based on e-commerce [20]. This system not only realizes the transaction behavior of e-commerce. In addition, users' preferences can be analyzed and mined from consumers' behaviors such as retrieval, browsing and shopping, so as to realize the recommendation of products to users.

In this study, the commodities that users are interested in are mined and analyzed by data mining technology and matched according to the types of commodities. Furthermore, the improved traditional collaborative filtering recommendation algorithm is studied, and fuzzy  $c$ -means (FCM) clustering and collaborative filtering algorithm are combined into the system. The recommendation algorithm selects the most interesting products from the set of alternative products according to certain rules to recommend. In this paper, the fuzzy  $C$ -means clustering algorithm is constructed for data processing. Besides, the collocating recommendation method is performed as a tool to conduct consumer behavior analysis. Finally, the consuming behavior in e-commerce is identified and classified through big data.

### 3. Establishment of Prediction Model

**3.1.  $C$ -Means Clustering Algorithm.** The working principle diagram of fuzzy  $C$ -means clustering algorithm for collocating recommendation method is shown in Figure 2, which is a core studying object of this proposed method. A special recognition process of recommend products for users in e-commerce based on calculations by computer is used for analyzing the discrete model generated by big data. Besides, the parallel clustering and distributed clustering algorithms were applied to implement in the computing clusters [21]. The data analysis process has gone through three steps: data collection, data analysis, and data integration. In order to recommend products for users in e-commerce, the most appropriate analysis method is found by analyzing various data methods [22]. At present, cluster analysis, analysis based on the existing associated components of data, and analysis based on the sequential relationship between data

are used. FCM analysis, a branch of multivariate statistical analysis and an important branch of unsupervised pattern recognition, is a method of grouping similar samples into one class. The FCM clustering algorithm is one of the methods of cluster analysis and also has the advantage of unsupervised learning; especially when the number of clusters is known, this method has a good effect on pattern recognition. When the values of data set  $X$ , clustering category number  $C$ , and weight  $r$  are known (generally adopted), FCM algorithm can be used to determine the optimal fuzzy classification [23]. This paper will be effective large data volume clustering techniques. The core idea of big data clustering technology is to deal with the relationship between computational complexity and computational cost, as well as scalability and speed. Therefore, the focus of big data clustering algorithm is to improve the scalability and execution speed of the algorithm at the cost of minimizing the clustering quality. The parallel clustering and distributed clustering algorithms applied in this study need to be implemented in computer clusters. The hardware architecture of multimachine clustering is shown in Figure 3. This study introduces functional requirements and nonfunctional requirements. Among them, functional requirements mainly include the overall framework of the system, the requirement analysis of each part of the function, and the role of users, while nonfunctional requirements include the requirements of network system performance, application system performance, and data performance. Compared with neural network algorithm, clustering algorithm is a kind of unsupervised algorithm. Because it does not need training set, the algorithm is simple and fast. In addition, adaptive clustering does not need to set a specific value in advance and can make clustering results adaptively [24, 25].

The fuzzy  $C$ -means clustering algorithm divides the data set into  $C$  subsets, and the corresponding fuzzy partition matrix  $U$  is generated.  $c_j$  is the center of each cluster and can be recorded as  $c$ ,  $\mu_j$  is the membership function of the  $i$ th sample corresponding to the  $j$ -th class. The clustering loss function based on the membership function is shown in the following equation [25]:

$$J_m = \sum_{i=1}^c \sum_{k=1}^n (\mu_{ik})^m (d_{ik})^2. \quad (1)$$

When the number of iterations is  $c$ , the sum of center vector of clustering is calculated as

$$\sum_{i=1}^c \mu_{ik} = 1, \quad \forall k = 1, 2, \dots, n. \quad (2)$$

The input of the cluster center vector of class  $I_k$  in iteration:

$$\begin{cases} \mu_{ik}^b = \frac{1}{\sum_{j=1}^c (d_{ik}/d_{jk})^{2/(m-1)}} & I_k = \varphi \\ \mu_{ik}^b = 0, \quad \forall i \in I_k, \sum_{i \in I_k} \mu_{ik}^b \neq \varphi. \end{cases} \quad (3)$$

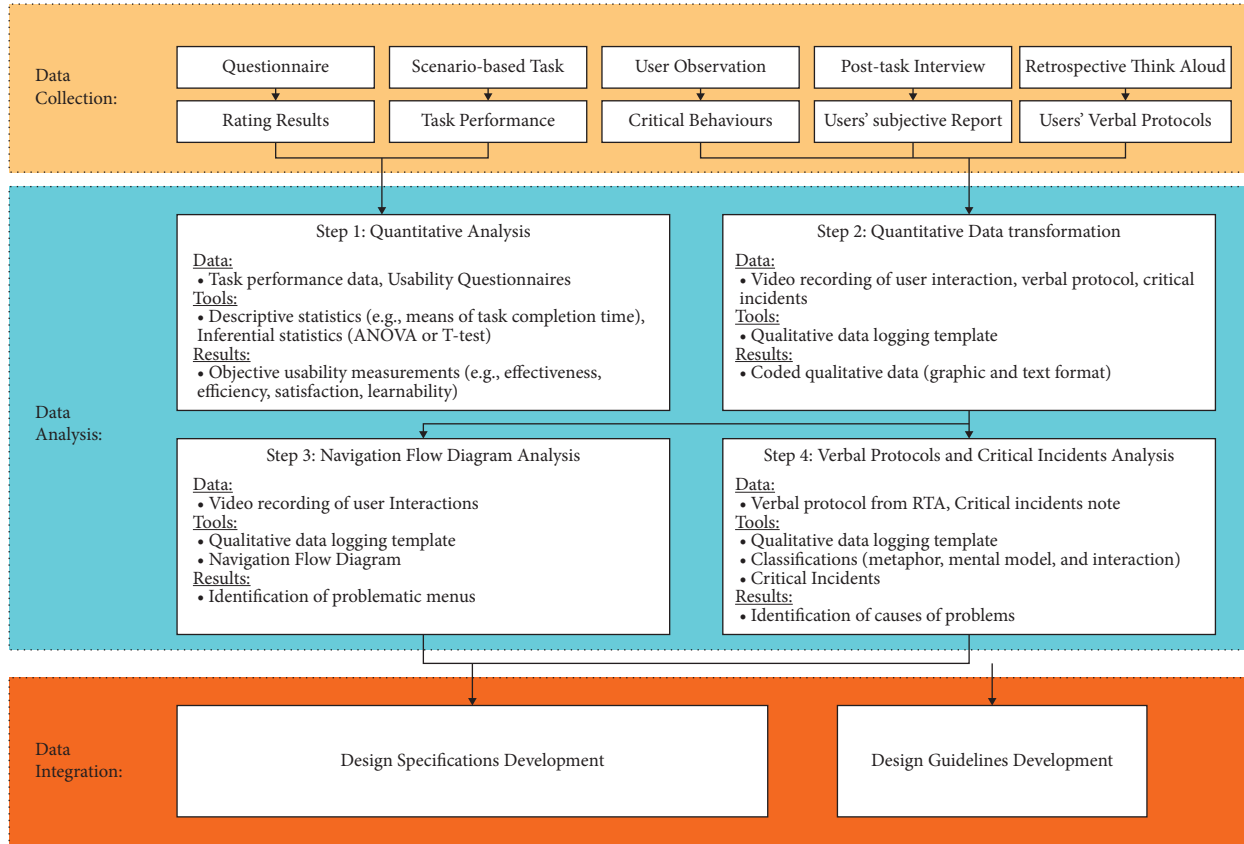


FIGURE 2: Data mining structure based on clustering algorithm.

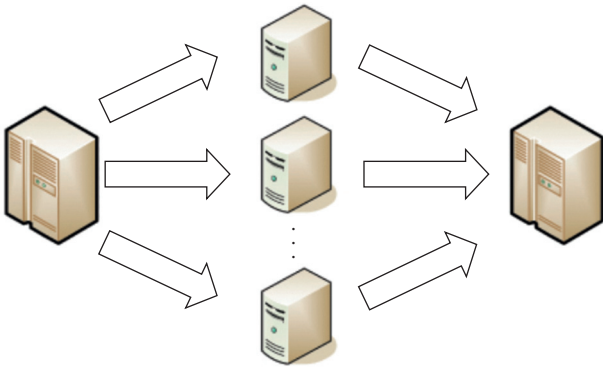


FIGURE 3: The hardware architecture of multimachine clustering.

The result of  $v_i^b$  in the iteration is

$$v_i^b = \frac{\sum_{k=1}^n (\mu_{ik}^b)^m x_k}{\sum_{k=1}^n (\mu_{ik}^b)^m}. \quad (4)$$

According to equation (4),  $U_{ij}^{(b+1)}$  is updated as  $u_{ij}^{(b+1)}$ , namely,

$$u_{ij}^{(b+1)} = \frac{1}{\sum \left( \|x_j - v_i^{(b)}\| / \|x_j - v_q^{(b)}\| \right)^2}, \quad (5)$$

where  $u_{ij}^{(b+1)}$  is the desired output.

3.2. *Simulation Experiment Analysis.* According to the general process of machine learning algorithm, first of all, the one dimensional discrete wavelet transform is adopted to data feature extraction, and the sample is divided into training set and testing set, using the training set to FCM clustering algorithm of control chart pattern recognition training, finally using the test set to test the accuracy of the algorithm, the specific technical route according to Figure 4 [26].

Two new individuals are formed, and the crossing method of  $m$ -th chromosome  $y_{mj}$  and  $n$ -th chromosome  $y_n$  at position  $j$  is as follows:

$$y_{mj} = y_{mj}(1-h) + y_{nj}h, \quad (6)$$

where  $h \in [0, 1]$  is a random number.

$$h = \frac{f_i}{\sum_{j=1}^N f_j}. \quad (7)$$

In order to enhance the diversity of the population, the continuity and trend model is used as

$$y_{ij} + (y_{\min} - y_{ij}) * h \left( 1 - \frac{h}{h_i} \right), \quad (8)$$

$$h_i = \frac{k}{y_i}. \quad (9)$$

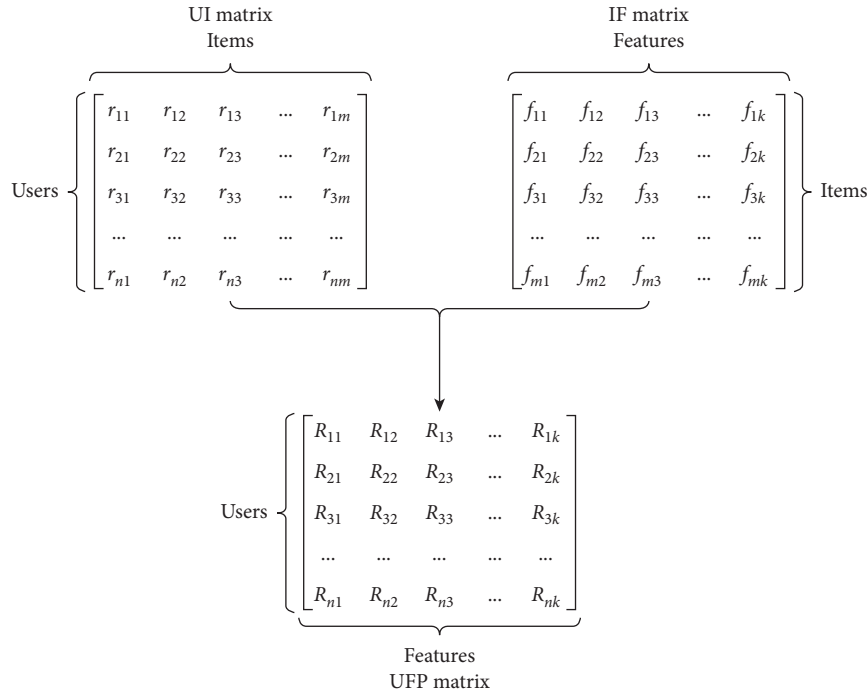


FIGURE 4: Basic flow of user preference feature matrix construction process.

Then, the step-index model is used as

$$y_{ij} + (y_{\min} - y_{ij}) + d \cdot w, \quad (10)$$

where  $d$  is the observed value of quality characteristic parameters in the manufacturing process.  $w$  is the average value of quality characteristic parameters under system control.

$$d = k \sum_{i=1}^n (y_i - y_0), \quad (11)$$

$$w = \frac{k}{d}. \quad (12)$$

Since FCM clustering algorithm does not require a large number of samples for training, monte Carlo method is used to generate 180 training samples, which meets the requirements of the number of training sets. The scoring vector of user for the project is  $r_u$ , and the membership vector of the corresponding feature of item is  $f_i$ .

$$R_{ui} = \frac{r_u \times f_i^T}{\sum_{j=0}^{j=m} f_{j,i}}. \quad (13)$$

**3.3. Properties of Predicting Algorithm.** A three-layer neural network of input layer, hidden layer, and output layer was constructed [27]. Both the input layer and the output layer of the network have one node. The hidden layer activation function uses nonlinear function to input and output data. In building a network of hidden layer nodes, in order to

guarantee the network error converge to the premise of as small as possible, to reduce the training time of network, and, at the same time, to avoid network over fitting phenomenon of samples, the experiment times are trained to adjust the network parameters, and finally we selected and used 150 nodes of neurons in hidden layer neural network structure for setup. In order to ensure the training accuracy, the number of network iterations is set as 500, and the error (MSE) converges to 0 during the training.

As a classical data mining clustering algorithm, C-means algorithm [28] is a simple and very effective algorithm [28]. In this paper, the clustering algorithm of user-space C-means algorithm in commodity recommendation system is divided into four steps by using classical C-means algorithm as follows.

- (1) Firstly, set  $n$  users and take these  $N$  users as the initial clustering center.
- (2) Calculate the remaining users, and assign each user to the cluster center set with the highest similarity to itself.
- (3) Calculate the newly generated clustering, and generate new clustering centers according to the user's score on the project.
- (4) Repeat the above 2 to 3 steps until there is no change in the clustering.

When researching the user product recommendation system based on e-commerce, this paper adopts the combination of FCM and collaborative filtering algorithm to accelerate the data calculation speed, reduce the real-time

response time, and enhance the scalability. Thus, the accuracy, precision, and real-time of recommendation are improved (see Figure 5)

Then, save the optimal solution according to the fitness value, record the optimal solution, and replace the local solution with the global solution. Finally, determine whether the termination conditions are met. If not, repeat Step 3 and Step 4. If so, the optimal weight and threshold value of network are output [29].

The calculation formula of individual fitness  $F$  is as follows:

$$F = k \sum_{i=1}^n \text{asb}(y_i - o_i), \quad (14)$$

where  $n$  is the number of output nodes of the network,  $y_i$  is the expected output of the  $i$ -th node of the Backpropagation (BP) neural network,  $o_i$  is the actual output of the  $i$ -th node, and  $k$  is the coefficient.

The selection probability  $p_i$  of each individual  $i$  is

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j}, \quad (15)$$

$$f_i = \frac{k}{F}$$

where  $f_i$  is the fitness value of individual  $i$ , and  $N$  is the number of individuals in the population.

The error evaluation of simulation calculation adopts mean relative error, mean square error, and mean absolute error, namely,

$$\text{MAE} = \frac{1}{N_p} \sum_{i=1}^{N_p} |x_i - \hat{x}_i|, \quad (16)$$

$$\text{MRE} = \frac{\sum_{i=1}^{N_p} (x_i - \hat{x}_i)^2}{\sum_{i=1}^{N_p} (x_i)^2}, \quad (17)$$

$$\text{MSE} = \frac{1}{N_p} \sum_{i=1}^{N_p} (x_i - \hat{x}_i)^2, \quad (18)$$

where  $x_i$  is the actual value,  $\hat{x}_i$  is the predicted value, and the sample number is  $N_p$ .

The studied mean absolute error (MAE) is the real value of absolute difference between the value with real values and predicted ones. The average relative error (MRE) is the squares of mean absolute error absolute error to the sum of squares of true value, which is used to measure the deviation probability between predicted value and true value. Besides, the mean square error (MSE) is to the square of difference between the the expected value of true value. These values are used to measure the changing degree of data [30].

#### 4. Information of Evaluation Model

Next, the method of combining FCM and collaborative filtering algorithm in personalized recommendation process is

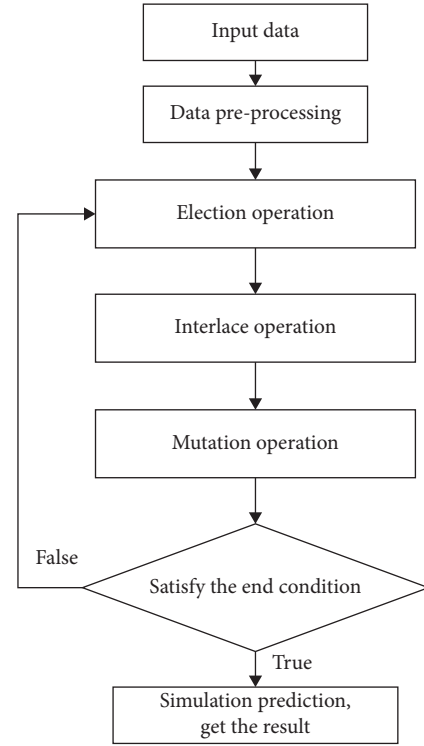


FIGURE 5: Flow chart of C-means algorithm.

introduced in detail, so as to facilitate weight analysis of the commodities that users are interested in, which are mined from historical data and commodity information. The data processing module mainly processes user data and commodity data from two aspects. The first kind of data is what other kinds of products users browse when browsing a certain kind of products. In the given recommendation scenario, according to the analysis between commodities and target users, the commodities that users are interested in are mined and analyzed by data mining technology and matched according to the types of commodities. According to the established purchase memory function, by combining FCM and collaborative filtering algorithm, weight analysis is carried out on the commodities that users are interested in, which are mined from historical data and commodity information [31]. Cluster analysis is carried out on the commodities according to the weight analysis results, and personalized commodities are recommended. Therefore, personalized recommendation of goods mainly includes three parts: type matching of goods, weight analysis of goods that users are interested in, and cluster analysis of goods according to the weight analysis results.

This study mainly introduces the design process of user commodity recommendation system based on e-commerce. Firstly, the design principles adopted in the design of the system are introduced, and then the architecture design and the functional design of each part of the user commodity recommendation system are introduced [32]. Finally, the database design on the recommendation system is explained and introduced. The design of user product recommendation system based on e-commerce provides the basis for the realization of the system.



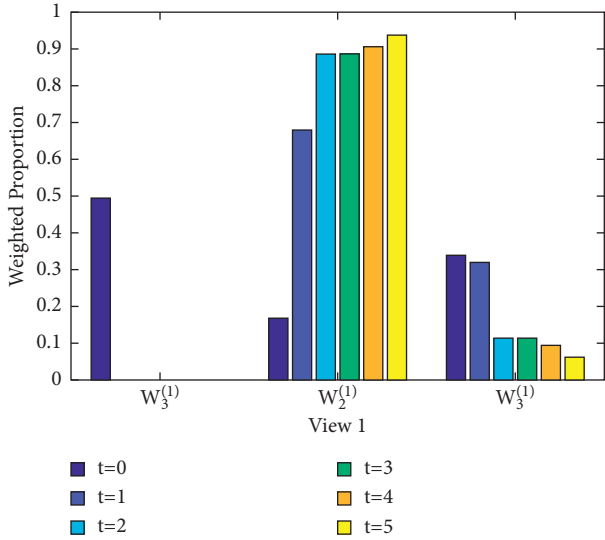


FIGURE 6: Comparison of prediction results of training samples.

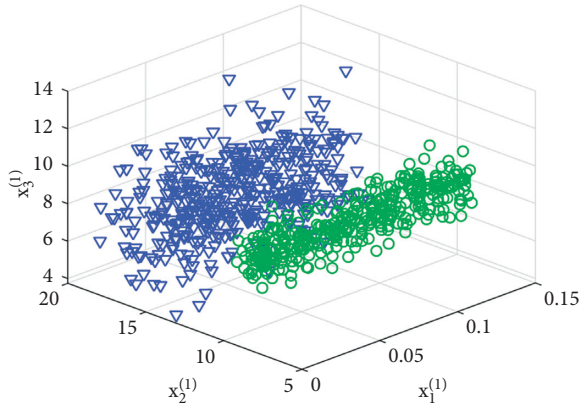


FIGURE 7: Comparison of the cluster data set with a noise feature.

### 5. Calculation Results Are Analyzed and Discussed

According to certain rules, the product recommendation module selects the product that the target user is most likely to be interested in from the alternative product set for recommendation, which mainly includes three parts: recommendation scene construction, weight analysis, and personalized product recommendation. The method in this paper, the clustering analysis method of information entropy, and the clustering analysis method based on neural algorithm are, respectively, used for clustering analysis of the selected 12 data sets, and the clustering accuracy is shown in Figure 6.

It can be seen from Figure 7 that the average clustering accuracy of the proposed method is higher than that of the other two methods. As illustrated in Figure 8, this paper demonstrates the predicting behaviors by using a numerical analysis. This example proposed in the paper is also applied in explaining the selected indexes in handling distribution of view weights. Furthermore, there is the effect on learning

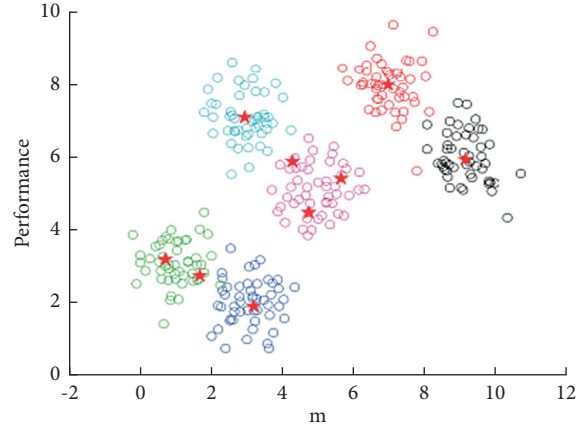


FIGURE 8: Performance of the test using C-FCM on data set.

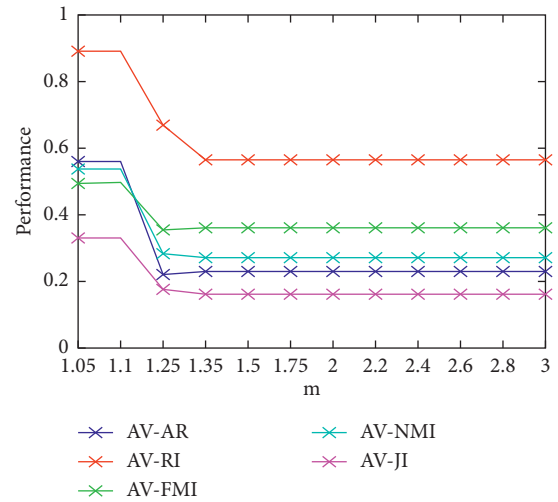


FIGURE 9: Performance comparison of different time conditions using C-FCM data set.

process with each index in performing the final predicting results.

Based on the data of nearly 10,000 registered users on an e-commerce website in three years, 7418 users who have logged in at least once and have purchase records are selected as the research objects to conduct data classification verification using C-FCM data set and Minimax-FCM on data set. After initial statistical analysis of users in the early stage, it is found that the clustering number and SSE corresponding to purchase times and purchase amount of users are shown in Figures 8 and 9 respectively. In the process of multidimensional user analysis with a large number of levels, the selection of C-means clustering clusters can be controlled by a key dimension in the user group for dimensionality reduction. In the classical RFM model of user classification, the last purchase time  $R$ , that is, the time interval from the last purchase time to the current, is the most important indicator. In this paper, the total error sum of  $R$  is selected to distinguish the cluster number of user clustering. By the last time of the index processing, two cluster numbers of 10 and 100 are selected as the reference.

The cluster number and corresponding SSE of the running results are shown in Figure 9.

Each cluster in e-commerce subsystem and express subsystem is composed of multiple index variables. Therefore, it is necessary to eliminate redundant index variables through grey correlation analysis method and select a representative index variable from each cluster. The results of grey correlation analysis of indicator variables of e-commerce subsystem and the results of grey correlation analysis of indicator variables of express subsystem are shown in this study. With the increase of cluster value, the grouping of users will be more accurate, and the degree of aggregation in the cluster will also become higher; that is, SSE will gradually decrease. When the cluster number reaches the true value, SSE will decrease slowly and become stable; that is, subdivision will have no practical significance. It can be seen from Figures 3 and 4 that as the number of family  $K$  continues to increase, when  $K$  value is greater than 4, SSE decreases significantly and basically becomes stable. In other words, when  $K=4$  is used for analysis by C-means algorithm, it is suitable for the accuracy requirements of current users' clustering classification. Two special clusters appear in the C-means analysis results; that is, each cluster contains only one value, and the purchase amount is RMB 4820,248 and RMB 1352426, respectively. The number of users in these two clusters is very small, but from the perspective of purchase amount, the users are extremely important to the e-commerce enterprises, so the e-commerce enterprises need to invest more resources for key maintenance. The other two clusters contain a large number of users, and the maximum purchase amount is only 912380 yuan. Therefore, C-means clustering analysis can be performed again for the user groups whose purchase amount is less than one million yuan, and the results are shown in Figure 6. In order to meet the requirements of e-commerce enterprises for reasonable classification of users, the above theories and methods are applied again, and the analysis results obtained after running the program are shown in Figures 10 and 11, respectively. The method in this paper, the clustering analysis method of information entropy and the clustering analysis method based on neural network algorithm are, respectively, used for clustering analysis of the selected 12 data sets, and the clustering accuracy is shown in Figure 2. It can be seen that the average clustering accuracy of the proposed method is higher than that of the other two methods.

### 6. The Optimized Algorithm Based on Simulated Data

The successive application of C-means clustering analysis can, on the one hand, identify the singularities in the data but attach great importance to the users and, on the other hand, make the classification degree of users controllable and the structure clear. The user classification characteristics in each cluster are shown in the following table. Within the cluster 0, users purchase amount and purchase frequency correlation is extremely low; namely, the increase of number of such users to buy will not bring users to buy the increase

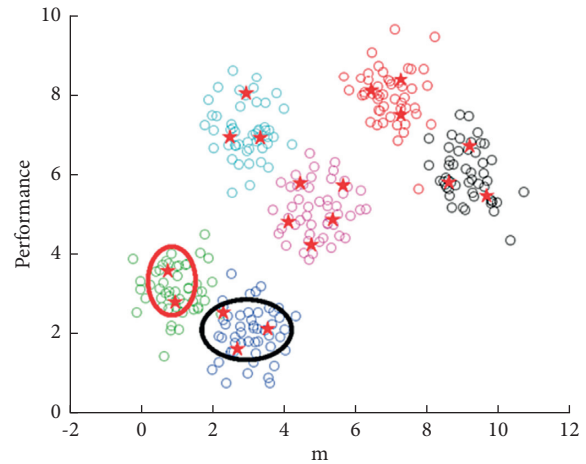


FIGURE 10: Performance of the test using Minimax-FCM on data set.

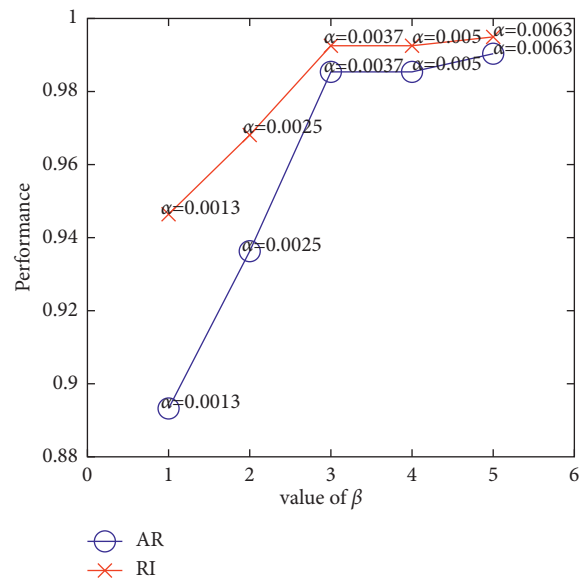


FIGURE 11: Relations of the parameter and performance of the proposed method.

of the amount, but by electricity enterprise single more than \$one hundred for free distribution policy influence, such user to price sensitive, purchase frequency is particularly low, and the number of products needs to increase the low value, guiding such users to purchase the marketing strategy of high-value products of e-commerce enterprises, so as to improve the contribution of such users. There are very few users in cluster 1, but the user's purchase amount is close to that of all users in cluster 0, and its purchase frequency and average single amount are extremely high. It is a strategic enterprise user of e-commerce, and e-commerce enterprises need to send professionals to maintain the relationship with this user. The users of cluster 2 have a large purchase amount and a higher purchase frequency, but the average single amount is not high. They should be the commission users of the e-commerce enterprises, and the e-commerce



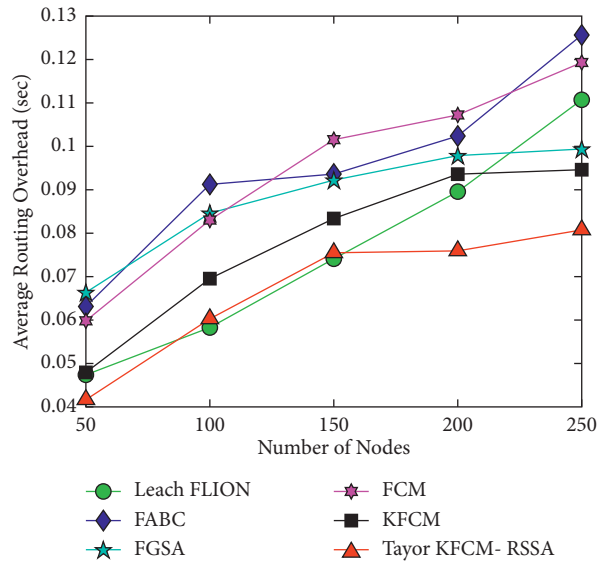


FIGURE 12: The value of mean square error predicted by C-means clustering algorithm.

enterprises need to send technical personnel to provide technical guidance or help. Users of other clusters also show their own characteristics, providing data basis for service allocation and precision marketing of e-commerce enterprises.

Relationship of the parameter and performance of the proposed method are shown in Figure 11, the clustering accuracy of the proposed method is the highest, and the clustering analysis method based on genetic algorithm has a higher accuracy than that based on information entropy. Therefore, in the comparative analysis of clustering efficiency, only the clustering efficiency of the proposed method and that based on genetic algorithm are compared, and the results are shown in Figure 12. Therefore, the overall time of the iterative process presented in this paper is significantly reduced and has significant efficiency advantages.

- (1) Most users purchase products through the home page and product introduction page, so they are excluded from the characteristics of user behavior category.
- (2) In the user group of category I, the pages of commodity classification, commodity sale, and other functions account for a low proportion, while the pages of commodity search and shopping cart account for a high proportion. Therefore, category I can be classified as search-type users who search and purchase commodities through search pages.
- (3) For category II user groups, user behavior in the classification of goods page proportion is significantly higher than other pages, and at the same time of basic commodity search page or shopping cart page, this says that the user buys goods through the platform home page to choose goods types and choose commodities they see in the introduction page. Then, they add the goods to the shopping cart for purchase, the order of goods selected in the

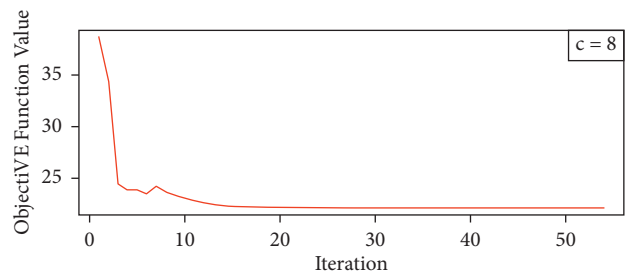


FIGURE 13: Convergence study on the predicting data.

shopping cart is consistent with the order set by sales, and in three different categories, this category accounts for more than 50%, and users of this category can be defined as ordinary users.

- (4) In category III, the page for special sale has the highest proportion of user behavior, while other pages have a low proportion, indicating that this user group pays more attention to the page for special sale, which can be defined as promotional users. The convergence study on the predicting data is shown in Figure 13.

As can be seen from Table 1 and Figure 12, the values of mean square error predicted by C-means clustering algorithm for economic benefits of enterprises all show a gradual upward trend after adopting different methods to cluster the user behaviors of the objects. The method in this paper has the most significant improvement range, and with the extension of using time, the economic benefits of e-commerce platform are more significant and have greater application value.

Internet companies use big data technology to collect all kinds of data about customers, and through big data analysis to establish a system of user portrait to describe the overall information of a user abstractly, so that users

TABLE 1: Comparative results of different algorithms.

Methods	Leach flion	Taylor KFCM	FABC	FGSA	FCM	KFCM
Alive nodes	2	25	30	7	1	31
Delay	0.1245	0.1124	0.1265	0.1375	0.1611	0.1214
Energy	0	0.0187	0.0333	6.20	0	0.0947
Throughput	0.1667	0.2552	0.0220	0.2000	0.2000	0.2857
Routing overhead	0.0474	0.0632	0.0632	0.0598	0.0479	0.0418

can be personalized recommendation, precision marketing and advertising [33–35]. When a user logs into the website, the system can predict the user’s intention and then find out the right goods from the commodity library and recommend to consumers. The core of E-commerce recommendation supported by big data is to push the business of enterprises to the users who need this business most at the right time, through the right carrier and in the right way. In the Internet era, users’ consumption behavior is prone to change in a short time, so big data marketing can be implemented in a timely manner when users’ demand is the greatest [36]. Therefore, the big data and machine learning technique in e-commerce can achieve one-to-one marketing for subdivided users according to users’ interests and needs at a certain point in time and timely adjust marketing strategies according to real-time effect feedback.

## 7. Conclusion

In this paper, a clustering classification method of e-commerce user behavior based on combinatorial optimization is proposed to improve the clustering analysis effect of e-commerce user behavior. Combining BP neural network algorithm and C-means algorithm, clustering analysis is carried out on the data set of e-commerce user behavior. Based on combinatorial optimization theory, clustering analysis of e-commerce user behavior is realized. Experiments verify that the proposed method can obtain high-precision clustering results. The collected data is preformatted and used as the input of data mining. Through the user’s behavior record, analyze the user’s potential like and like degree of goods and establish the user preference model. Through commodity analysis, commodity similarity, commodity collocation, and target user tag are analyzed. Then, it uses data mining technology to mine and analyze the commodities that users are interested in, makes matching according to the types of commodities, and recommends the commodities that users are interested in under a given scene according to the established purchase memory function. Finally, through the combination of FCM and collaborative filtering algorithm, recommend the commodities that users are interested in, which are mined from historical data and commodity information.

This study also briefly reviews the application of big data in e-commerce. Big data is so widely used in e-commerce that the development of big data will be faster and faster in the future. Moreover, the influence of big data will become more and more profound, so for e-commerce, the attention to big data can not be reduced. The data-driven era combined with machine learning has come, and big data will

surely become a huge source of power influencing this era in the future.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no known conflicts of financial interest or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] K. Carling, M. Han, J. Håkansson, X. Meng, and N. Rudholm, “Measuring transport related CO<sub>2</sub> emissions induced by online and brick-and-mortar retailing,” *Transportation Research Part D: Transport and Environment*, vol. 40, pp. 28–42, 2015.
- [2] C. Leong, S. L. Pan, S. L. Pan, S. Newell, and L. Cui, “The emergence of self-organizing e-commerce ecosystems in remote villages of China: a tale of digital empowerment for rural development,” *MIS Quarterly*, vol. 40, no. 2, pp. 475–484, 2016.
- [3] S. Meyer, B. Hans-Peter, and S. Sebastian, “Retail investor information demand-speculating and investing in structured products,” *European Journal of Finance*, vol. 22, no. 10, pp. 1063–1085, 2016.
- [4] Y. Chang, R. J. Ries, and Y. Wang, “Life-cycle energy of residential buildings in China,” *Energy Policy*, vol. 62, pp. 656–664, 2013.
- [5] A. Perry, R. M. Lynch, I. Rusyn, and D. W. Threadgill, “Long-term combinatorial exposure to trichloroethylene and inorganic arsenic in genetically heterogeneous mice results in renal tubular damage and cancer-associated molecular changes heterogeneous mice results in renal tubular damage and cancer-associated molecular changes,” *G3 Genes Genomes Genetics*, vol. 9, no. 5, pp. 1729–1737, 2019.
- [6] J. Zhang, H. Qiu, X. Li et al., “Real-time nowcasting of microbiological water quality at recreational beaches: a wavelet and artificial neural network-based hybrid modeling approach,” *Environmental Science & Technology*, vol. 52, no. 15, pp. 8446–8455, 2018.
- [7] S. Xiao and M. Dong, “Hidden semi-markov model-based reputation management system for online to offline (O2O) e-commerce markets,” *Decision Support Systems*, vol. 77, pp. 87–99, 2015.
- [8] Y. Wang and L. Chen, “Multi-view fuzzy clustering with minimax optimization for effective clustering of data from multiple sources,” *Expert Systems with Applications*, vol. 72, pp. 457–466, 2017.

- [9] A. Rabani, J. Madariaga, C. Bouvier, and D. Axinte, "An approach for using iterative learning for controlling the jet penetration depth in abrasive waterjet milling," *Journal of Manufacturing Processes*, vol. 22, no. 4, pp. 99–107, 2016.
- [10] E. Png, S. Srinivasan, K. Bekiroglu, C. Jiang, R. Su, and K. Poolla, "An internet of things upgrade for smart and scalable heating, ventilation and air-conditioning control in commercial buildings," *Applied Energy*, vol. 239, no. 4, pp. 408–424, 2019.
- [11] C.-L. Liao and S.-J. Lee, "A clustering based approach to improving the efficiency of collaborative filtering recommendation," *Electronic Commerce Research and Applications*, vol. 18, pp. 1–9, 2016.
- [12] F. Da, S. Gai, Y. Huang, and L. Zeng, "Research on a hole filling algorithm of a point cloud based on structure from motion," *Journal of the Optical Society of America, A. Optics, image science, and vision*, vol. 36, no. 2, pp. A39–A46, 2019.
- [13] H. R. Richard and M. E. Arthur, "Comparing the utility of decision trees and support vector machines when planning inspections of linear sewer infrastructure," *Journal of Hydroinformatics*, vol. 16, no. 6, pp. 1265–1279, 2014.
- [14] X. Guo, J. Xiao, and Y. Wang, "A survey on algorithms of hole filling in 3D surface reconstruction," *Visual Computer*, vol. 34, no. 1, pp. 93–103, 2018.
- [15] M. Majdi, I. Alijarah, H. Faris, A. I. Hammouri, A. M. Al-Zoubi, and S. Mirjalili, "Binary grasshopper optimization algorithm approaches for feature selection problems," *Expert Systems with Applications*, vol. 117, no. 11, pp. 267–286, 2019.
- [16] S. Ding, X. Xu, H. Zhu, J. Wang, and F. Jin, "Studies on optimization algorithms for some artificial neural networks based on genetic algorithm (GA)," *Journal of Computers*, vol. 6, no. 5, pp. 939–946, 2011.
- [17] O. Hasan, T. Erzurumlu, and F. Erzincanli, "Prediction of minimum surface roughness in end milling mold parts using neural network and genetic algorithm," *Materials & Design*, vol. 27, no. 9, pp. 735–744, 2006.
- [18] D. Mackin, X. Fave, L. Zhang et al., "Measuring computed tomography scanner variability of radiomics features," *Investigative Radiology*, vol. 50, no. 11, pp. 757–765, 2015.
- [19] S. Nanda and R. A. Hansen, "Forensics as a service: tier architecture for cloud based forensic analysis," in *Proceedings of the 2016 15th International Symposium on Parallel and Distributed Computing*, IEEE, Fuzhou, China, 2016.
- [20] M.-F. Tsai and C.-J. Wang, "On the risk prediction and analysis of soft information in finance reports," *European Journal of Operational Research*, vol. 257, no. 1, pp. 243–250, 2017.
- [21] E. Fields, C. Osorio, and T. L. Zhou, "A data-driven method for reconstructing a distribution from a truncated sample with an application to inferring car-sharing demand," *Transportation Science*, vol. 55, no. 3, pp. 616–636, 2021.
- [22] S. De Luca and R. Di Pace, "Modelling users' behaviour in inter-urban carsharing program: a stated preference approach," *Transportation Research Part A*, vol. 71, no. 1, pp. 59–76, 2015.
- [23] J. K. Tarus, Z. Niu, and G. Mustafa, "Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning," *Artificial Intelligence Review*, vol. 50, 2017.
- [24] F. Ebrahimi and M. R. Barati, "Vibration analysis of biaxially compressed double-layered graphene sheets based on non-local strain gradient theory," *Mechanics of Advanced Materials and Structures*, vol. 26, no. 2, pp. 1–12, 2018.
- [25] K. Nie, H. Al-Hallaq, X. A. Li et al., "NCTN assessment on current applications of radiomics in oncology," *International Journal of Radiation Oncology, Biology, Physics*, vol. 104, no. 2, pp. 302–315, 2019.
- [26] S. Jiang, A. Alves, F. Rodrigues, J. Ferreira Jr., and F. C. Pereira, "Mining point-of-interest data from social networks for urban land use classification and disaggregation," *Computers, Environment and Urban Systems*, vol. 53, pp. 36–46, 2015.
- [27] C. Sabrina, M. Ali, M. I. Ben, and N. Ladhari, "Colour recipe prediction using ant colony algorithm: principle of resolution and analysis of performances," *Coloration Technology*, vol. 135, no. 5, pp. 349–360, 2019.
- [28] E. Riza, Y. Tu, J. Benjamas, and M. Li, "Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending," *Applied Economics*, vol. 47, no. 1, pp. 54–70, 2015.
- [29] P. Guido, M. Rosano, M. Saint-Guillain, and P. Rizzo, "Simulation-optimisation framework for city logistics: an application on multimodal last-mile delivery," *Intelligent Transport Systems*, vol. 12, no. 4, pp. 262–269, 2018.
- [30] B. Arpit and A. Tiwari, "Breast cancer diagnosis using genetically optimized neural network model," *Expert Systems with Applications*, vol. 42, no. 10, pp. 4611–4620, 2015.
- [31] H. Izakian, W. Pedrycz, and I. Jamal, "Fuzzy clustering of time series data using dynamic time warping distance," *Engineering Applications of Artificial Intelligence*, vol. 39, pp. 235–244, 2015.
- [32] V. Borjalilou and M. Asghari, "Small-scale analysis of plates with thermoelastic damping based on the modified couple stress theory and the dual-phase-lag heat conduction model," *Acta Mechanica*, vol. 229, no. 6, pp. 1–16, 2018.
- [33] J. Wang, Y. Yang, T. Wang, R. Sherratt, and J. Zhang, "Big data service architecture: a survey," *Journal of Internet Technology*, vol. 21, no. 2, pp. 393–405, 2020.
- [34] M. Duan, K. Li, X. Liao, and K. Li, "A parallel multi-classification algorithm for big data using an extreme learning machine," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2337–2351, 2018.
- [35] J. Zhang, S. Zhong, T. Wang, H.-C. Chao, and J. Wang, "Blockchain-based systems and applications: a survey," *Journal of Internet Technology*, vol. 21, no. 1, pp. 1–14, 2020.
- [36] B. Xu, D. Huang, and B. Mi, "Research on e-commerce transaction payment system based on c4.5 decision tree data mining algorithm," *Computer Systems Science and Engineering*, vol. 35, no. 2, pp. 113–121, 2020.