Research Article

Precise Marketing Strategy Optimization of E-Commerce Platform Based on KNN Clustering

Benfang Yang [ID][1] **and Jiye Li**[2]

[1]*School of Economics and Management, Sichuan Technology and Business University, Chengdu, Sichuan 611700, China*
[2]*School of Computer Engineering, Chengdu Technological University, Chengdu, Sichuan 611700, China*

Correspondence should be addressed to Benfang Yang; yangbf@mail.cdtu.edu.cn

With the development of computer technology and the arrival of the era of artificial intelligence, the analysis of user demand bias is of great significance to the operation optimization of e-commerce platforms. Combined with CS domain signaling data, IP packet data of PS domain, and customer CRM data provided by operators, this research studies each dimension index of operator user portrait, after that the operator user portrait platform is divided into some individual subunits, and then the corresponding data mining technology is carried out to study the implementation scheme of each subunit. The system can process and mine multidimensional data of operators' users and form user portraits on the basis of user data aggregation. Finally, based on the operator user portrait platform studied in this paper, the operator user data are analyzed from both the user's mobile phone use behavior and user consumption behavior. Furthermore, the application value of this research in the precision marketing and personalized service of operators is illustrated.

1. Introduction

The combination of mobile communication and the Internet has created the current era of mobile Internet. With the development and growth of mobile Internet, operation data of large e-commerce platforms are generated [1]. Communication service operators are located in the center of information exchange and are the transmitters of all kinds of data. They can use convenient conditions to obtain a large amount of data. Therefore, many operators have a preliminary understanding of the application value of big data and try to use it for their own use and create benefits. For example, some operators conduct dynamic analysis on the network status of the business platform and the status of terminal devices through signaling data, so as to adjust the structure of the communication network and maximize the communication value. Some operators use cloud technology to mine users' personal characteristics, grasp users' main needs and consumption preferences accurately, and carry out more accurate marketing for different users [2–4].

However, after in-depth investigation, operators are not fully mining and analyzing communication data and are not accurate enough in the labeling display of user behaviors, and only a few influential products have been developed. Operators have more extensive data coverage than Internet companies or other companies in the industry. However, operators' business has been mainly in the field of communication for many years, which makes the data accumulated by operators also related to the consumption of communication charges, and there are certain restrictions on the application of these data [5]. At the same time, operators did not consider the importance of data at the beginning of their operations, and in the past, when storage devices were expensive, operators cleaned up a lot of data that they considered unimportant but actually had a lot of value. From the beginning, the establishment of the operator's business is mainly focused on mobile broadband and other aspects, and the user interaction is mainly to charge broadband fees and provide services. As a result, operators cannot fully understand the data of other aspects of the

society and cannot find good application scenarios [6–8]. As a result, the noncommunication data accumulated by operators are small, which causes great inconvenience for operators to expand other services. And how to excavate the useful information contained in operation data and optimize the precision marketing strategy of e-commerce platform is a challenging work [9–12].

The contribution of this paper is to design an improved customer segmentation system of e-commerce platform based on the RFM model. On this basis, the k-means method is used to further subdivide the customer groups of the e-commerce platform, and precise marketing strategies of the e-commerce platform are formulated according to the segmentation results of the customer groups. In addition, this paper studies marketing strategies of e-commerce platforms from the perspective of precision marketing, and relevant studies have not been involved before.

This paper consists of five parts. The first and second parts give the research background and related works. The third part is the methods and preliminary. The fourth part shows the data acquisition and analysis platform. The experimental results of this paper are introduced and compared and analyzed with relevant comparison algorithms in the fifth part. Finally, the conclusion of this paper is given in the sixth part. The study of this paper gives the user portrait system of operators and gives the realization scene of precision marketing. The paper can complete the analysis function according to the data provided by operators to form user portraits, which shows application value in precision marketing of operators.

2. Related Works

Nowadays, many scholars have studied the precise marketing strategy optimization of e-commerce platform. Ye and Feng [13] showed the effects of big data on the e-commerce precision marketing strategy and discussed that the e-commerce industry can determine the consumption needs of consumers and habit accurately. Huang [14] analyzed the concept of mobile marketing and given the current status of traditional dairy products brand of Yili district and obtained some important conclusions for the enterprises. Erdmann and Ponzoa [15] studied the cost-outcome relationship of grocery e-commerce in-store marketing behaviors. The strategy for research in this paper is based on the application of Dorfman and Steiner's optimal advertising budget model, which is applied to digital marketing and proved by the empirical statistical analysis method. The results show that the difference depends on the formats and countries. Zhu and Gao [16] declared that under the background of the digital marketing model, traditional retail industry is facing unprecedented impact, and the competitive advantage of traditional marketing is disappearing. Hence, on the basis of the digital marketing model, this paper makes a comprehensive exploration and analysis of retail precision marketing strategy and shows the differences between them. Since artificial intelligence (AI) is playing a more and more important role in marketing analysis field, Chiu and Chuang in [17] developed

an omni-channel chatbot which combines the iOS, Android, and Web components. The designed chatbot is to take the convolutional neural networks (CNNs) for personalized service and precision marketing. A case study of a shared kitchen is used to illustrate the advantages of the new method, which can be applied to other consumers application scenarios such as personalized services and clothing selection.

A lot of advanced science and technology gradually enter into the field of marketing strategy optimization of e-commerce platform. For example, e-commerce platforms can not only establish their blockchain anticounterfeiting traceability platform but also play ball with third-party blockchain anticounterfeiting traceability platform [18, 19]. Guo et al. [20] developed the differential game model in four situations, and the relationship between the choice of sales mode and the anticounterfeiting traceability service strategy is discussed. The experimental results indicated that supplier's profit can be affected by many aspects and show different effects. Shahrel et al. [21] designed a Web application called Price Cop to help customers monitor product pricing, which helps users plan before they buy. A price forecasting model is established by using linear regression technique. LR is commonly used to determine prognosis and as a predictor [22]. The accuracy of least-squares support vector machine (LSSVM) is evaluated through artificial bee colony (ABC). LSSVM-ABC was originally proposed to predict the stock market price. Govindarajan and Chandrasekaran [23] proposed K-nearest neighbor (KNN) classifier that performs comparative cross-validation against existing K-nearest neighbor classifiers which is described. The feasibility and superiority of this method are showed in e-commerce platform. Recently, Kohli et al. [24] gave a review of sales prediction by linear regression and KNN algorithm.

From the above analysis, in the application of KNN, existing clustering methods based on KNN have many defects and deficiencies; for example, they fail to optimize the precise marketing strategies of e-commerce platforms; and there is no integration and analysis of e-commerce data, which will affect the proposal and optimization of marketing strategies.

3. Methods and Preliminary

3.1. The Concept of Precision Marketing. Precision marketing refers to a product and service marketing model that accurately locates customer needs on the basis of establishing the customer relationship system by means of information technology, which is different from the traditional concept and can not only effectively reduce costs for enterprises but also effectively improve the peer competitiveness [25, 26]. The core of precision marketing lies in mastering the consumption level and preference of target customers, which requires enterprises to fully understand customers and establish customer information database under the condition of conditions and push products and services most suitable for customers after analyzing and predicting their consumption preferences. With the growing maturity of big data technology, more and more enterprises have

reconstructed their customer relationship management mode through this precision marketing method and further upgraded their marketing thinking for core customer groups [27–29]. Based on the existing research results in the academic circle and the current situation of economic development, this paper preliminarily explores the concept of precision marketing, that is, on the premise that enterprises clearly grasp the market trends and customer needs, they build differentiation through big data and other scientific and technological means. Precise customer product service mechanism and customer relationship management system further reduce the marketing cost of enterprises and promote the rapid and effective development of business.

3.2. K-Means Clustering Method. As a common method in data mining technology, cluster analysis is to divide the analyzed sample data into several different groups according to a certain principle, which makes the similarity of each data point in the group as large as possible and the similarity of the sample data points between different groups as small as possible [30, 31]. There are many kinds of cluster algorithms. For example, the common algorithms include the analysis method based on hierarchy, the analysis method based on density, the analysis method based on division, and so on. Among many clustering algorithms, the K-means algorithm is one of the most basic and widely used clustering algorithms. The main connotation of the K-means clustering algorithm is to divide each sample data point into different groups by repeated iteration. By comparing the distance between the sample data point and the centroid of the group, the distance between each sample point in the same group and the distance between the sample data point of different groups are the minimum and the distance between the sample data point of different groups is the maximum. In the K-means clustering algorithm, Euclidean distance is generally used to measure the distance between data sample points; suppose the input data set is

$$T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}. \tag{1}$$

The above training set x is the feature vector of the sample, y is the category of the sample, and T is feature space containing the input samples. According to the given distance measure, find k points closest to X in the training set T, and the neighborhood of X containing these k points is called $N(k)$. Do not worry about what the neighborhood is here. In short, the neighborhood is the set domain of X points. In $N(k)$, category Y of X is determined according to classification decision rules (such as majority voting), and its formula is as follows:

$$y = \arg\max_{c_j} \sum_{x_i \in N_k(x)} I(y_i = c_j), \quad i = 1, 2, \ldots, N; j = 1, 2, \ldots, K. \tag{2}$$

In the K-means clustering algorithm, the following indicators are generally adopted to measure the distance between data sample points:

$$L_p(x_i, x_j) = \left(\sum_{l=1}^{n} \left| x_i^{(l)} - x_j^{(l)} \right|^p \right)^{1/p}, \tag{3}$$

where x_i represents the value of the ith dimension variable of the data point x and y_i represents the value of the ith dimension variable of the data point y. When $p = 2$, the distance is Euclidean distance. When $p = 1$, the distance is Manhattan distance.

$$L_1(x_i, x_j) = \sum_{l=1}^{n} \left| x_i^{(l)} - x_j^{(l)} \right|. \tag{4}$$

When $p = \infty$, the distance is the maximum distance of each coordinate.

$$L_\infty(x_i, x_j) = \max_l \left| x_i^{(l)} - x_j^{(l)} \right|. \tag{5}$$

The classification decision rule of the KNN algorithm is usually the majority voting rule; that is, the majority classes of K adjacent training instances of the input classes determine the category of the input classes; the relationship between them can be explained in this way, when the classification loss function is 0-1, formula (6) is the probability of classification:

$$f: \mathbf{R}^n \longrightarrow \{c_1, c_2, \ldots, c_K\}. \tag{6}$$

The probability of misclassification is as follows:

$$P(Y \neq f(X)) = 1 - P(Y = f(X)). \tag{7}$$

Assume that there are k training instance points nearest to form a set $N(x)$, then the error rate is given as follows:

$$\frac{1}{k} \sum_{x_i \in N_k(x)} I(y_i \neq c_j) = 1 - \frac{1}{k} \sum_{x_i \in N_k(x)} I(y_i = c_j). \tag{8}$$

In general, the application process of the K-means clustering algorithm includes the following four steps.

First, the sample data points were divided into K groups, each of which represented a different group category.

Second, the initial centers of different groups were determined according to the minimum Euclidean distance principle between sample data points.

Third, the new centroid of the group is determined. After determining the initial center of different groups, continuous optimization is needed to ensure that the initial center is more reasonable and reliable. The Euclidean distance between sample data points is calculated, and the new centroid with the mean value of Euclidean distance is determined.

Fourth, iterate Steps 2 and 3 by iterating repeatedly that all groups after operation of mass center and the edge of the boundary between groups no longer significantly change; that is to say, the K-means cluster analysis result was relatively stable, namely, clustering analysis process is complete, and the final clustering results are shown in Figure 1.
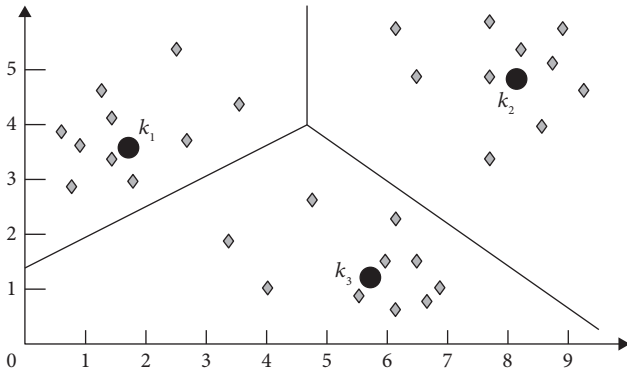
FIGURE 1: K-means clustering result under stable state.

4. Data Acquisition and Analysis

4.1. Data Source Analysis. Before analyzing and processing the data, you must know the data to be processed and the specific data structure. The original data provided by carriers mainly consist of three parts:

(1) CRM (Customer Relationship Management) data contain basic personal attribute information of users including basic personal information, ID information, address information, contact information, consumption information, package information, and terminal information.

(2) Signaling data in the Circuit Switch (CS) domain include phone call records, SHORT message sending records, and interaction records between terminals and the network.

(3) IP packet data in Packet Switch (S) domain mainly contain data packet records on the control plane and user plane when users use the network, such as Authentication, Authorization, and Accounting (AAA) Authentication data packets, Packet Data Protocol (PDP) establishment, deletion, and update. User data are mainly used by users on the network.

In view of the above data, it is necessary to filter the data and obtain useful information before further processing. The data are provided by the operator, and the collection of these data requires the design of effective data collection units.

The overall structure of the collection module is shown in Figure 2. Oracle DATABASE HDFS and Flume CRM are installed in the big data processing server. CRM data are collected directly to Oracle database through SQL Developer. For signaling data in CS domain and IP packet data in PS domain, multichannel Flume adopts Spooldir to collect user data in parallel. Agent will be started on each node to monitor the directory uploaded by FTP. When new data are uploaded, Source will format the captured data first, then push it into Channel buffer, and then the Channel will transfer the data to Sink. Sink of each node at the end will upload the data to Agent3 node, which will synthesize the data and submit it to HDFS to realize the concurrent data collection of user number.

4.2. Data Preprocessing and Analysis

4.2.1. IP Packet Data Extraction. At present, the most common text category analysis techniques include LDA (Latent Dirichlet Allocation) algorithm. The LDA semantic analysis model is an unsupervised algorithm realized through clustering. In the training process, the K value of the clustering class needs to be specified. If the K value is selected improperly, the result will be adversely affected. The multiclassification algorithm is a supervised machine learning algorithm, which requires a lot of training data to train the model and a lot of manpower to mark the categories of text data, and the final accuracy is not well guaranteed. Through many comparative experiments, this paper selects a text classification scheme based on the TF-IDF (Term Frequency Inverse Document Frequency) algorithm. Keywords of text content are extracted through TF-IDF, and the formula of TF-IDF is as follows:

$$\text{tfidf}_{i,j} = \text{tf}_{i,j} \times \text{idf}_{i,j}, \tag{9}$$

where $\text{tf}_{i,j}$ is the term frequency, which represents how often a given word appears in a document. The variable $\text{idf}_{i,j}$ represents the inverse document frequency index, which can be described as follows:

$$\text{idf}_{i,j} = \log \frac{|D|}{1 + |\{j : t_i \in d_j\}|}, \tag{10}$$

where $|D|$ is the total number of files for corpus and $\{j : t_i \in d_j\}$ is the number of files that contain t_i; this keyword extraction method is more representative and accurate than word frequency extraction.

4.2.2. Data Cleaning. To keep the accuracy of clustering, data cleaning is needed for the data to be analyzed before data clustering analysis. This work mainly discusses the numerical transformation of data and the processing of statistic singularities and standardization. In this paper, the signaling data extracted from IP domain data packets and the data in user CRM data are cleaned and filtered to obtain useful data information as shown in Figure 3.

The specific processing process is as follows.

First, data are read from the distributed storage system including IP domain packet classification information that has been processed by IP packet classification, signaling data in the CS domain, and original user CRM data.

After that, the read data are numerically transformed, and the data columns are divided into three categories for processing. The first category is the columns originally of numerical type, including voice, phone charge, traffic, age, number of short tenets, Internet time, and other information, which can be directly processed in the next step. The second category is the column representing category data, including information such as gender and regional classification, which needs to be mapped between category and number. The third category is the column of other
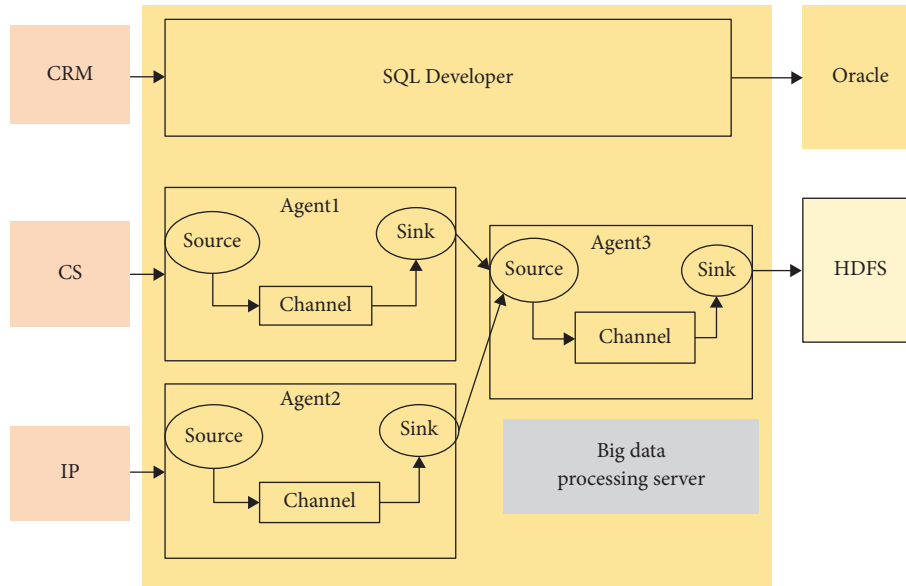
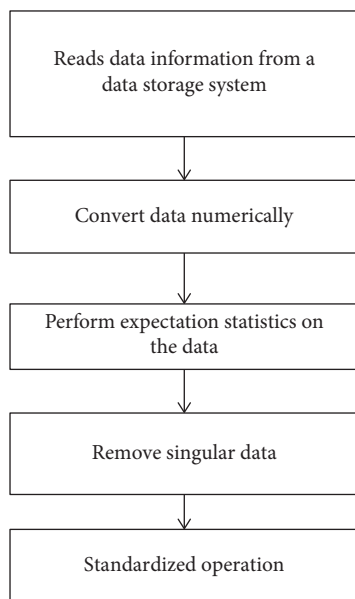FIGURE 2: Design of the user data acquisition module.

FIGURE 3: Flow chart of data cleaning process.

information, mainly including address information, interests, and hobbies, which can be processed according to label coding.

The user's usage information in the last two years is taken for statistical analysis, and the expected value is calculated according to month, week, and day, respectively, for subsequent processing.

For data standardization operation, due to the selection of voice traffic and call fee data, there is a relatively large gap in the data range; for example, the voice range is generally tens to hundreds of minutes, and the traffic usage may be hundreds to thousands of gigabytes, so it is necessary to carry out data standardization

operation. The standardization method selected in this paper is deviation standardization, and its realization is described as follows:

$$y_i = \frac{x_i - \min_{1 \le j \le n}\{x_j\}}{\max_{1 \le j \le n}\{x_j\} - \min_{1 \le j \le n}\{x_j\}}. \tag{11}$$

5. Experimental Results and Analysis

5.1. User Marketing Analysis. The main goal of analyzing users' consumption behavior is to facilitate the accurate formulation of voice data plans for specific users to provide personalized services. User consumption behavior is mainly based on user consumption history records, traffic usage records in IP packets, and voice call records extracted from CS domain signaling data. The following describes the clustering analysis of user traffic flow. The data flow and the usage of call charges are analyzed by clustering, and several different K values are selected for experiments. Figures 4–7 show the clustering effect when K value is 2, 4, 5, and 6, respectively. X-axis represents monthly call time (single bit: minutes), Y-axis represents monthly traffic usage (unit: MB), and Z-axis represents monthly call fee (unit: yuan).

As you can see from the figures that with the increase in K value, the clustering effect will be more detailed. Here, the sum of squares of errors within clusters (SSE) is needed to select K value. Figure 7 shows the SSE-K curve from 0 to 20 of K value, where the abscess represents the value of K and the ordinate represents the sum of squares of errors within clusters (SSE). That is to say, the K is selected by experience.

As can be seen from Figure 8, with the increase in K value, the sum of squares of error in the cluster presents a decreasing trend. When K value is less than 6, SSE decreases rapidly; when K value is 6, SSE decreases slowly. Therefore, K value is selected as 6. When K value is set to 6, 6 central
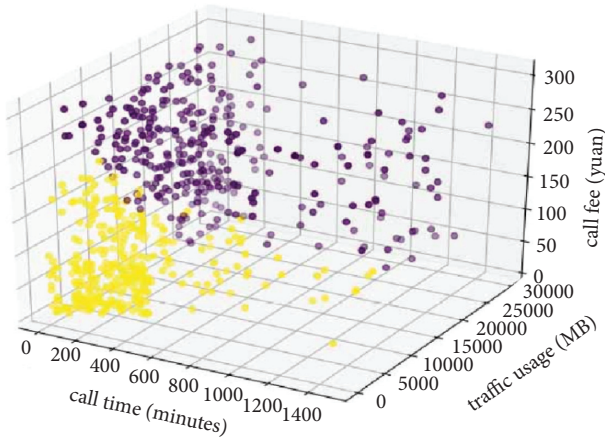
Figure 4: Clustering analysis of user traffic flow: ($k = 3$).

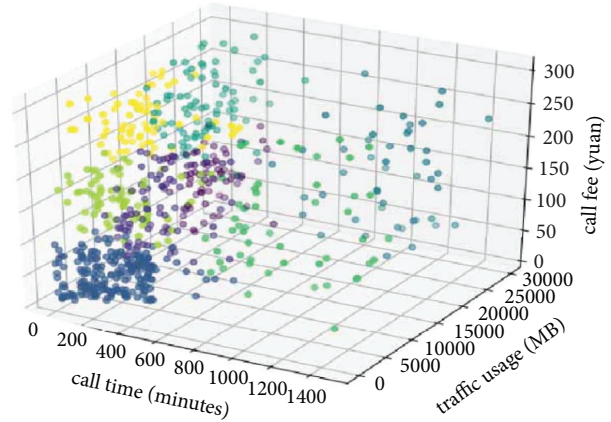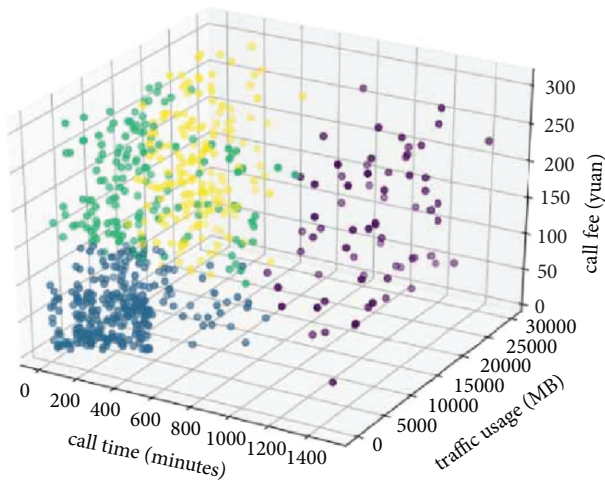Figure 7: Clustering analysis of user traffic flow: $k = 5$.

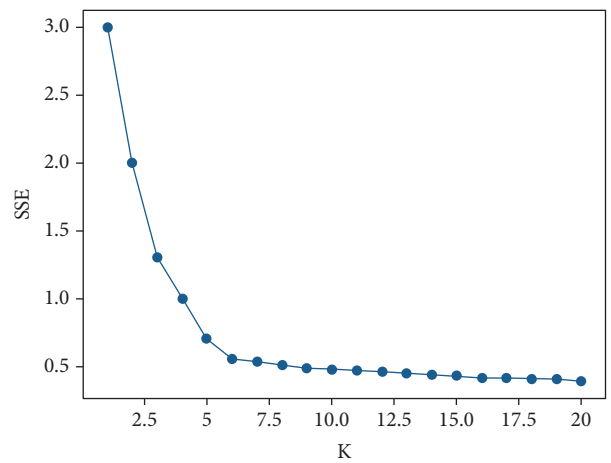Figure 5: Clustering analysis of user traffic flow: ($k = 3$).
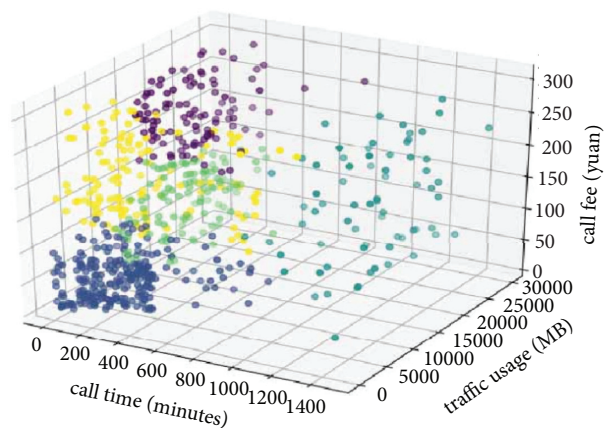

Figure 8: Diagram of SSE.

Figure 6: Clustering analysis of user traffic flow: $k = 4$.

points can be selected, the coordinates of each central point are analyzed and the crowd characteristic information is marked out as shown in Table 1. Based on these consumption data, the voice traffic usage trend and consumption of users can be calculated, and adding these

features to the user label system can facilitate operators to develop personalized user packages.

It can be seen from Table 1 that the basic call charges can be divided into six categories according to the user's call duration consumption level and traffic usage label. These six categories can be used as the consumption category of the user as a level 1 label to represent the user. Users can be more accurately labeled according to the specific call consumption level traffic and other labels to achieve a fine description of the user portrait and facilitate more accurate marketing.

Here, due to the small amount of selected data, there are few classifications. When there is a large amount of data, the selection of K value should be differentiated according to the specific situation. This section conducts cluster analysis on the three dimensions of voice traffic and call fee information and obtains the analysis results, which confirms the feasibility of selecting the K-means ++ cluster analysis method in the research process of this topic.

5.2. Analysis Based on Mobile Phone Use Behavior. The analysis of users' mobile phone usage behavior can optimize operators' services in the following aspects:

TABLE 1: Cluster analysis results.

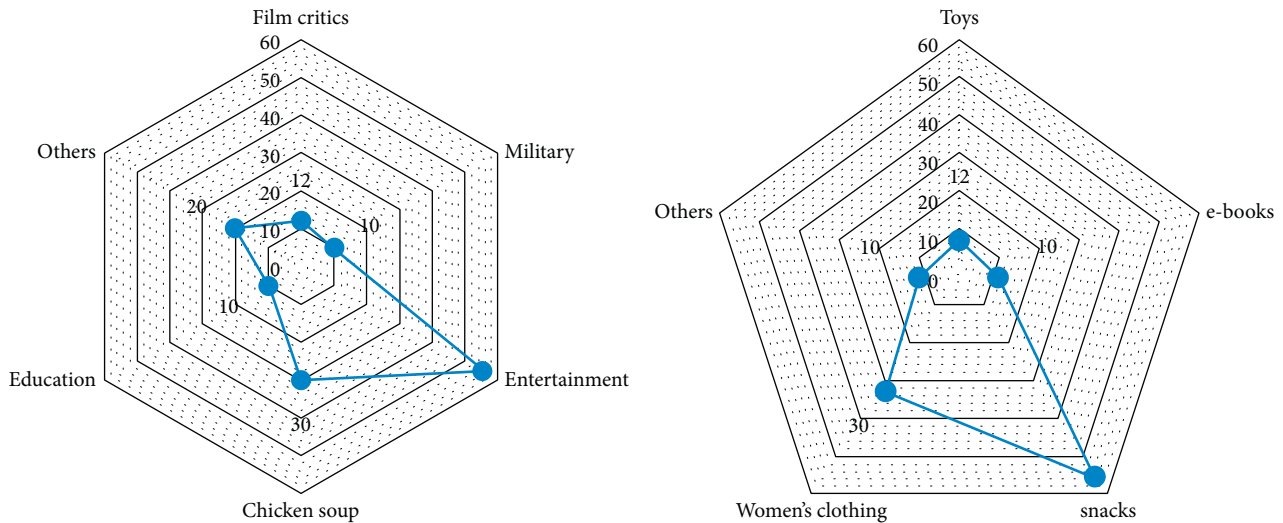| Call duration	Flow	Telephone fee	Call label	Consumption level	Flow label	Usage type
85.36	19836.20	251.07	Less	High	More	I
185.46	2442.75	22.73	Medium	Low	Less	II
1242.46	12730.42	124.24	More	High	More	III
71.24	17154.61	164.31	Less	Middle	Medium	IV
23.26	12.42	15.51	Less	Low	Less	V
252.51	5441.51	152.53	Medium	Middle	Medium	VI

FIGURE 9: Diagram of user interest.

(1) By analyzing the online time period of users, the peak period of data traffic usage can be counted to strengthen operation optimization

(2) By analyzing users' browsing preferences and shopping preferences, personalized products can be developed for users and reasonably pushed to users

(3) By analyzing the user's common location information, the layout of the base station information can be reasonably optimized

The user's mobile phone usage behavior is mainly to analyze the data generated during user's mobile phone usage, which is mainly obtained based on the Internet access record information in IP domain packet information. Based on the analysis of various dimensional characteristics of users, this paper draws a user interest distribution map for accurate product formulation and push. User interest distribution is in the stage of analyzing the dimension of user interest and preference. It calculates the weight value of user preference according to the length and times of user access to the product category and stores it in the database, thus forming the user interest distribution. Figure 9 describes user interest from two aspects of user browsing preference (left) and shopping preference (right). Notably, the browsing preference and shopping preference are the main two aspects of user interest.

6. Conclusions

This work investigates the specific application cases of the user portrait system based on big data technology in the field of precision marketing of operators, analyzes the research status of user portrait research KNN clustering, and clarifies the meaning and focus of the research.

In the analysis of the processing results of the user portrait platform, in order to facilitate the display of the results, this paper designed the user portrait result display platform and conducted data analysis of the results. Finally, the application scenarios of user portrait in precision marketing of operators are analyzed from two aspects of user's mobile phone usage behavior and user's consumption behavior.

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 conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] K. Zheng, Z. Zhang, and B. Song, "E-commerce logistics distribution mode in big-data context: a case analysis of JD.COM," *Industrial Marketing Management*, vol. 86, pp. 154–162, 2020.

[2] J. Gao, Y. Qin, X. Cheng, T. Zhang, J. Guan, and L. Xu, "A best-marketing time prediction algorithm based on big data analytics," *Signal and Information Processing, Networking and Computers*, Springer, Singapore, 2020.

[3] R. Alabdulrahman and H. Viktor, "Catering for unique tastes: targeting grey-sheep users recommender systems through one-class machine learning," *Expert Systems with Applications*, vol. 166, Article ID 114061, 2021.

[4] I. O. Pappas and A. G. Woodside, "Fuzzy-set qualitative comparative analysis (fsQCA): guidelines for research practice in information systems and marketing," *International Journal of Information Management*, vol. 58, Article ID 102310, 2021.

[5] M. Stöhr, G. S. Michelitsch, J. C. Tully, K. Reuter, and R. J. Maurer, "Communication: charge-population based dispersion interactions for molecules and materials," *The Journal of Chemical Physics*, vol. 144, no. 15, Article ID 151101, 2016.

[6] T. S. Ustun, S. M. S. Hussain, and H. Kikusato, "IEC 61850-based communication modeling of EV charge-discharge management for maximum PV generation," *IEEE Access*, vol. 7, pp. 4219–4231, 2018.

[7] T. Y. Lin Yejian, M. Anghel, and D. Livescu, "Data driven learning of Mori-Zwanzig operators for isotropic turbulence," *Physics of Fluids*, vol. 33, no. 12, Article ID 125118, 2021.

[8] Y. Xu, M. Zhang, and X. Xu, "A fast detection method of network crime based on user portrait," *Journal of Information Hiding and Privacy Protection*, vol. 3, no. 1, pp. 17–28, 2021.

[9] P. Lenti, S. Kottmair, S. Stock, A. Shukri, and D. Müller, "Predictive modeling to identify potential participants of a disease management program hypertension," *Expert Review of Pharmacoeconomics & Outcomes Research*, vol. 21, no. 2, pp. 307–314, 2021.

[10] N. Jing, H. Zhang, X. Deng, B. Zhang, and Z. Ye, "Accelerate personalized IoT service provision by cloud-aided edge reinforcement learning: a case study on smart lighting," in *Proceedings of the 18th International Conference on Service-Oriented Computing*, pp. 69–84, Springer, Dubai, UAE, December 2020.

[11] A. Higashihara, T. Ono, G. Tokutake et al., "Hamstring muscles' function deficit during overground sprinting in track and field athletes with a history of strain injury," *Journal of Sports Sciences*, vol. 37, no. 23, pp. 2744–2750, 2019.

[12] H. Cao, "Personalized web service recommendation method based on hybrid social network and multi-objective immune optimization," *Journal of Information Processing Systems*, vol. 17, no. 2, pp. 426–439, 2021.

[13] X. Ye and L. Feng, "On E-commerce precision marketing strategy based on big data," *Big Data and Cloud Innovation*, vol. 3, no. 1, pp. 1001–1012, 2019.

[14] L. Huang, "Research on mobile marketing strategy based on SICAS model—a case study of Yili group," *American Journal of Industrial and Business Management*, vol. 9, no. 4, pp. 1059–1075, 2019.

[15] A. Erdmann and J. M. Ponzoa, "Digital inbound marketing: measuring the economic performance of grocery e-commerce in Europe and the USA," *Technological Forecasting and Social Change*, vol. 162, Article ID 120373, 2021.

[16] G. Zhu and X. Gao, "Precision retail marketing strategy based on digital marketing model," *Science Journal of Business and Management*, vol. 7, no. 1, pp. 33–37, 2019.

[17] M. C. Chiu and K. H. Chuang, "Applying transfer learning to achieve precision marketing in an omni-channel system–a case study of a sharing kitchen platform," *International Journal of Production Research*, vol. 59, no. 24, pp. 7594–7609, 2021.

[18] Y. Yalan and T. Wei, "Deep logistic learning framework for e-commerce and supply chain management platform," *Arabian Journal for Science and Engineering*, vol. 23, pp. 1–15, 2021.

[19] Z. Wang and H. Zhu, "Optimization of e-commerce logistics of marine economy by fuzzy algorithms," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 4, pp. 3813–3821, 2020.

[20] F. Guo, D. Ma, J. Hu, and L. Zhang, "Optimized combination of e-commerce platform sales model and blockchain anti-counterfeit traceability service strategy," *IEEE Access*, vol. 9, 2021.

[21] M. Z. Shahrel, S. Mutalib, and S. Abdul-Rahman, "PriceCop-price monitor and prediction using linear regression and LSVM-ABC methods for e-commerce platform," *International Journal of Information Engineering and Electronic Business*, vol. 13, no. 1, 2021.

[22] A. N. Hidayanto, M. S. Adha, M. A. Jiwanggi, and T. Melia, "A study of impact of search engine optimisation to internet marketing strategy," *International Journal of Services, Economics and Management*, vol. 4, no. 4, pp. 298–316, 2012.

[23] M. Govindarajan and R. Chandrasekaran, "Evaluation of *k*-nearest neighbor classifier performance for direct marketing," *Expert Systems with Applications*, vol. 37, no. 1, pp. 253–258, 2010.

[24] S. Kohli, G. T. Godwin, and S. Urolagin, "Sales prediction using linear and KNN regression," *Advances in Machine Learning and Computational Intelligence*, Springer, Singapore, 2021.

[25] Z. You, Y.-W. Si, D. Zhang, X. Zeng, S. C. H. Leung, and T. Li, "A decision-making framework for precision marketing," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3357–3367, 2015.

[26] C. Yu, Z. Zhang, C. Lin, and Y. J. Wu, "Can data-driven precision marketing promote user AD clicks? Evidence from advertising in WeChat moments," *Industrial Marketing Management*, vol. 90, pp. 481–492, 2020.

[27] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, "Learning *k* for KNN classification," *ACM Transactions on Intelligent Systems and Technology*, vol. 8, no. 3, pp. 1–19, 2017.

[28] X. Zhu, S. Zhang, R. Hu, Y. Zhu, and J. Song, "Local and global structure preservation for robust unsupervised spectral feature selection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 3, pp. 517–529, 2017.

[29] F. Restuccia, S. D'Oro, and T. Melodia, "Securing the internet of things in the age of machine learning and software-defined networking," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4829–4842, 2018.

[30] J. Jagannath, N. Polosky, A. Jagannath, F. Restuccia, and T. Melodia, "Machine learning for wireless communications in the internet of Things: a comprehensive survey," *Ad Hoc Networks*, vol. 93, Article ID 101913, 2019.

[31] M. H. Alsharif, A. H. Kelechi, K. Yahya, and S. A. Chaudhry, "Machine learning algorithms for smart data analysis in internet of things environment: taxonomies and research trends," *Symmetry*, vol. 12, no. 1, p. 88, 2020.