Hindawi
Complexity
Volume 2021, Article ID 5538677, 10 pages
https://doi.org/10.1155/2021/5538677

Research Article

Precision Marketing Method of E-Commerce Platform Based on Clustering Algorithm

Bei Zhang,[1] Luquan Wang,[1] and Yuanyuan Li[2]

[1]*School of Economics and Management, Shandong Xiandai University, Jinan 250104, Shandong, China*
[2]*Shandong Academy of Grape, Shandong Academy of Agricultural Sciences, Jinan 250100, Shandong, China*

Correspondence should be addressed to Yuanyuan Li; 000229@sdupsl.edu.cn

Received 2 February 2021; Revised 23 February 2021; Accepted 27 February 2021; Published 5 March 2021

Academic Editor: Wei Wang

In user cluster analysis, users with the same or similar behavior characteristics are divided into the same group by iterative update clustering, and the core and larger user groups are detected. In this paper, we present the formulation and data mining of the correlation rules based on the clustering algorithm through the definition and procedure of the algorithm. In addition, based on the idea of the *K*-mode clustering algorithm, this paper proposes a clustering method combining related rules with multivalued discrete features (MDF). In this paper, we construct a method to calculate the similarity between users using Jaccard distance and combine correlation rules with Jaccard distances to improve the similarity between users. Next, we propose a clustering method suitable for MDF. Finally, the basic *K*-mode algorithm is improved by the similarity measure method combining the correlation rule with the Jaccard distance and the cluster center update method which is the ARMDKM algorithm proposed in this paper. This method solves the problem that the MDF cannot be effectively processed in the traditional model and demonstrates its theoretical correctness. This experiment verifies the correctness of the new method by clustering purity, entropy, contour, and other indicators.

1. Introduction

By clustering analysis of users, we can find people with different interests and different behaviors, so that companies can analyze the characteristics of the core user groups of their products and provide help and basis for improving products and accurate marketing. In addition, user clustering analysis can also be applied to business decision-making, public opinion analysis, security warning, and other fields [1]. Data about users is usually mixed data. However, when using traditional clustering algorithms for user clustering analysis, it is impossible to dig deeper into the information of multivalued discrete features (MDF) [2]. This will lead to low data utilization and inaccurate user feature analysis. At the same time, the current user clustering analysis does not fully consider the association and importance between user data features, and most of the research treats different data features of users independently. Therefore, it is necessary to improve the utilization rate of user data and explore the association and importance of user features, which is important to improve the accuracy and quality of user clustering [3–10]. With the advent of the era of big data, Internet technology, database technology, and various data mining algorithms have developed rapidly [11]. Nowadays, Internet companies can obtain a large amount of data every day, and how to extract useful information from these data has been the direction of people's efforts. In order to solve this problem, researchers in Internet companies and institutions around the world have been actively drawing theoretical knowledge from various fields and conducting experimental validation [12].

In recent years, the competition of Internet enterprises has been quite fierce, and each enterprise is studying how to improve its own products to reduce the loss of users, how to be able to effectively tap into potential user groups, and how to analyze the interests and emotional state of users, which are all vital to the development of enterprises or even a matter of life and death [13]. Many Internet companies are

increasingly focused on how to use the data in hand to serve precision marketing; one of the typical methods is to cluster analysis of user data. User clustering analysis is to use different clustering algorithms to find groups of users with similar behavioral characteristics in different application scenarios, as a breakthrough in user behavior analysis [14]. By clustering analysis of users, we can understand users, infer their potential needs, explore potential user groups, enhance corporate influence, strengthen users' reliance on corporate products, and reduce the risk of user churn. At the same time, by grasping the current users' interests and needs, it helps to analyze and predict the users' future needs and provide the basis for the future development direction of the enterprise. Current research in the field of user clustering analysis generally uses traditional clustering algorithms such as *K*-means, which can characterize the audience of an enterprise's products and understand the interests and preferences of each group for data analysis, recommendation, and other works [15]. However, the user data obtained by enterprises are usually mixed types of data, including numerical types, categorical types, and MDFs, such as age, gender, interests, and hobbies. When using traditional clustering methods for user clustering analysis, generally only a single type of features can be handled, and the impact of various types of feature data cannot be considered comprehensively. The analysis of MDF only stays at the level of mathematical statistics, and it is impossible to use this kind of data to explore the hidden information of users, which leads to the low utilization of data [16]. At the same time, the current user clustering analysis method does not fully consider the association and importance of user features. Therefore, research on the above problems is of great significance for the development of enterprise products and even for the development of enterprises [17]. With the advent of the era of big data, how to filter the noisy information and discover valuable information from the large amount of user data collected has become a hot issue for Internet companies at home and abroad [18]. A variety of new products are constantly emerging in the market, which makes consumers' choice wider. At this time, how to discover the interests and behavioral characteristics of consumers becomes particularly important [19]. Among them, clustering analysis is a data mining technique widely used in the field of database knowledge discovery, and its operation can be shown in Figure 1. As early as the twentieth century, J. Mac Queen proposed the simple and efficient *K*-means clustering algorithm. Nowadays, using clustering methods to analyze the characteristics of the audience is one of the means for companies to conduct user analysis [20].

The source of user data is usually the activities of users in the online state, such as searching for information, shopping, interacting with microblogs, and watching videos. These activities generate a lot of data. With the growing number of users on the web today, the amount of data traffic generated on the web is increasing day by day. Enterprises can use various data mining algorithms to analyze a large amount of collected user data, which can well describe user characteristics, user behavior habits, and other pieces of important information, so as to further reason and predict user behavior and improve the existing system, which will greatly improve efficiency and bring great profits to enterprises. In this paper, we propose an unsupervised feature selection method combining *K*-means++ with random forest. The method uses *K*-means++ algorithm to perform preliminary clustering to obtain pseudolabels of user data; secondly, the user data with pseudolabels are selected by random forest to obtain feature importance ranking, and the importance is used as the weight parameter of user data features; finally, based on the idea of spectral clustering, the user data are analyzed by weighted clustering to obtain the final user clustering. Finally, based on the idea of spectral clustering, the final user clustering results are obtained by weighted cluster analysis of user data. The model integrates the weight relationship between user features, solves the problem of unlabeled user data, and can effectively improve the accuracy of clustering.

2. Related Work

In recent years, many Internet companies are paying more and more attention to how to use data for accurate marketing services, of which user clustering analysis methods are being studied by more and more companies. Through clustering analysis of user data, enterprises can discover the core audience and the potential behavioral information of users. At present, the main research direction of user clustering analysis is to analyze user behavior characteristics or support recommendation systems.

The previous authors analyzed the behaviors of groups with different consumption levels on the webcasting platform. It first used the user behavior data collected on the webcasting platform to construct a behavior feature dataset, used Gower distance to measure the similarity of hybrid features in the user data, and finally clustered the user groups of different consumption classes by Medoids clustering method. The researchers used the *K*-means clustering algorithm to analyze the heat map and charging time distribution of EV users' behaviors and summarize the behavioral characteristics of EV users. Researchers proposed a two-layer web user clustering method, which uses the DBSCAN algorithm to eliminate outliers, discovers irregular clusters using multiple features in user sessions, and finally uses a bottom-up hierarchical approach to cluster the initial clustering results. The source of user data is usually online user activity, such as information retrieval, shopping, microblog operation, and video viewing. These activities generate large amounts of data. As the number of users on the web increases, the amount of data traffic generated on the web increases daily. Companies can analyze large amounts of user data collected using various data mining algorithms. This allows you to properly describe user characteristics, user behavior habits, and other pieces of important information to further describe and predict user behavior, improve existing systems, and improve efficiency and bring great benefits to businesses.

The researchers propose a method to perform initial clustering using the SOM algorithm to obtain the initial clustering parameters. The above steps reduce the impact of

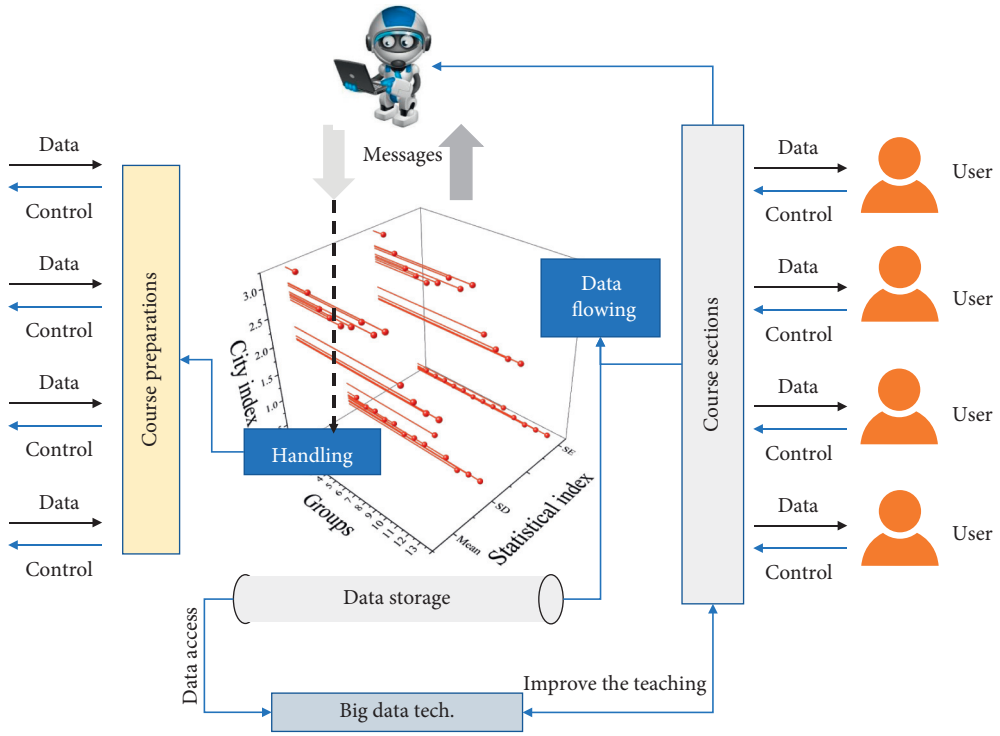

Figure 1: Clustering analysis algorithm.

improper initialization. The researchers propose a time-series-based method for classifying daily transactions of transit smart card users, which uses correlation distance, hierarchical clustering, and subgroups by metric parameters to understand the temporal patterns of users and to identify the daily behavior of different transit users. The researchers proposed a recommendation algorithm based on clustering and matrix analysis. First, the algorithm uses K-means to cluster user behavior data to find groups of users with similar behavioral characteristics and then uses similarities between user contexts to rank them to find the final set of similar users. Hsien-Ying Huang et al. proposed an adaptive clustering algorithm that introduces the topological potential field theory in physics and then combines the improved K-means algorithm with cluster users in order to help complete the later recommendation process. The researchers proposed a recommendation algorithm combining user clustering with scoring preferences, which preprocesses user behavior data by principal component analysis and K-means clustering and determines the weights of user behavior features by multiple linear regression. The researchers proposed a recommendation method based on user-item community detection. After obtaining clusters with tightly connected users and items, a traditional collaborative filtering model can be trained for each cluster.

Feature selection, also known as feature subset selection, extracts an optimal subset of features from all features in the sample data to make the constructed model more generalizable and effective. Unlike feature extraction, feature selection can retain as much information on the original features as possible while eliminating redundant features. Feature selection methods are broadly classified into supervised feature selection and unsupervised feature selection. Among them, supervised feature selection methods need to use the label information of samples to perform feature selection by measuring the correlation between sample features and labels, such as Relief, m RMR, and CFS. However, in real life, data with labels are difficult to obtain, and it is time-consuming and laborious to use manual labeling. Therefore, researchers are increasingly interested in the study of unsupervised feature selection methods.

3. Precision Marketing Based on Clustering Algorithm

3.1. Association Rule Mining. Association rule mining (ARM) is an important data mining technique, which is a process of identifying frequent patterns, associations, or causal structures from various types of datasets. ARM can be used to identify frequent item sets between uncertainties and generate powerful association rules from large datasets, especially auxiliary datasets for engineering operations. Currently, many scholars apply clustering methods to the discovery of association rules. However, they aim to improve the quality of the generated association rules. In this paper, association rules are applied to the process of user similarity metrics in clustering to improve the accuracy of user similarity metrics in clustering.

An association rule is an inference of the form $X > Y$, where X and Y are nonempty sets that do not contain the same elements, representing the antecedent and consequent parts of the rule, respectively. Three metrics are generally used to measure association rules, namely, support, confidence, and lift. The support of an association rule is the

percentage of all transactions that contain both X and Y, which can also be expressed as the probability $P(X > Y)$; the confidence is the percentage of transactions that contain both X and Y, which can also be expressed as the conditional probability $P(Y|X)$. The lift is the ratio of the probability of containing Y when X is included to the probability of containing Y when X is not included. If the lift of an association rule is greater than 1, it means that X and Y are positively correlated; if the lift is less than 1, it means that X and Y are negatively correlated; if the lift is equal to 1, it means that X and Y are not correlated.

The K-mode clustering algorithm uses the plural instead of the mean and is based on the idea of simple matching, using the Hamming distance to calculate the distance between two objects. The dissimilarity between an object and the clustering center is the number of features with different values corresponding to the features. Finally, all the 1's are summed and the cumulative value represents the dissimilarity between the object and the cluster center, and each object belongs to the cluster center with the least dissimilarity to it. It is defined as follows.

Let $U = \{x_1, x_i, x_n\}$ be a typed dataset containing n objects. The object x is denoted as $[x_{i1}, x_{i2}, x_{im}]$, where m is the characteristic number. Let x_i and x_{im} be the two objects represented by $[x_{i1}, x_{i2}, x_{im}]$ and $[x_{i1}, x_{i2}, x_{im}]$, respectively. x_i and x_{im} are defined by the following equation for calculating the distance between x_i and x_{im}:

$$\mathrm{Dis}\left(x_i, x_{im}\right) = \frac{\sum_{i=0}^{n} \Phi\left(x_i, x_{im}\right)}{x_i + \cdots + x_{im}}, \tag{1}$$

where $\Phi(x)$ is the indicator function.

$$\Phi\left(x_i, x_{im}\right) = \begin{cases} 1, & x_i < x_{im}, \\ 0, & x_{im} < x_i. \end{cases} \tag{2}$$

The optimization model of the K-mode algorithm when the formula is used as a distance metric for the object is defined as

$$Q(X, Y) = \frac{\sum_{i=0}^{k} \sum_{j=0}^{n} x_{ij} \left(\Phi_i + \cdots + \Phi_{im}\right)}{u_{ij}}, \tag{3}$$

subject to

$$\sum_{i=0}^{k} u_{ij} \sum_{i=0}^{n} \left(\Phi_i + \Phi_j\right) = 1, \quad i, j \in [0, 1] \tag{4}$$

where the affiliation matrix U is an $n*k$ binary matrix. At each iteration, if object i belongs to cluster p, then let $i_{\mathrm{pu}} = 1$; otherwise, $i_{\mathrm{pu}} > 0$. $Z = \{z_1, z_2, z_k\}$ denotes the set of k centers. $w = \{w_1, w_2, w_m\}$ is the weight vector of all features in the dataset.

The purpose of cluster analysis is to classify objects using the nature of the data itself, to be able to calculate the similarity between sample points according to specific definitions and to discover internal patterns in the data through iterations in order to classify the data. The above process does not require labeling information, so clustering analysis is unsupervised learning in machine learning. By clustering, objects in the same cluster tend to be similar in some sense, while objects in different clusters tend to be different. Cluster analysis allows macroanalysis of data without data mining for a particular individual. Usually, the similarity of samples is measured based on calculating the distance between sample data.

The distance is calculated differently for different scenarios. The closer the distance between two sample points is, the higher the degree of similarity is. Cluster analysis is widely used in various fields, such as group classification of target users. The target audience group is divided into several user groups with distinct characteristics of difference, so that personalized recommendations and services for the audience group can be carried out at a later stage, which ultimately improves the efficiency and business effect of enterprise operations, as well as discovering the value combinations of different products.

Enterprises can perform cluster analysis for a large number of product categories according to different application scenarios and purposes and according to specific evaluation indicators, in order to segment the product system and develop marketing programs that meet the current situation of the enterprise and so on. Through clustering analysis, it is possible to identify minority groups whose behavioral characteristics differ significantly compared to other groups, which may be system anomalies or irregularities of fraudulent groups and should be dealt with appropriately and, if necessary, fed back to and monitored by the relevant supervisory authorities. Common clustering algorithms can be divided into five categories based on their accumulation rules: division-based clustering, hierarchy-based clustering, grid-based clustering, density-based clustering, and model-based clustering.

3.2. Spectral Clustering. Spectral clustering is a clustering method that draws on the idea of graph theory, which converts the problem of classifying data categories into a problem of cutting undirected graphs. The idea of spectral clustering algorithm can be simply explained as the original high-dimensional feature space is downscaled to obtain a low-dimensional feature space, and then other traditional clustering algorithms are used in the low-dimensional data for clustering analysis, so as to achieve the purpose of clustering on data sample space of different shapes, as shown in Figure 2.

Euclidean distance is the most basic definition of the distance between two samples in an n-dimensional data space or the modulus of a vector. The Euclidean distance is chosen as the distance measure in most clustering algorithms. It is defined as follows:

$$\mathrm{dci} = \sqrt{\sum_{i=0}^{n} \left(\Phi_i + \Phi_j\right)}. \tag{5}$$

Cosine similarity measures the similarity of two vectors by calculating the cosine of the angle between them. In the field of text classification, cosine similarity is more widely used. It is defined as follows:

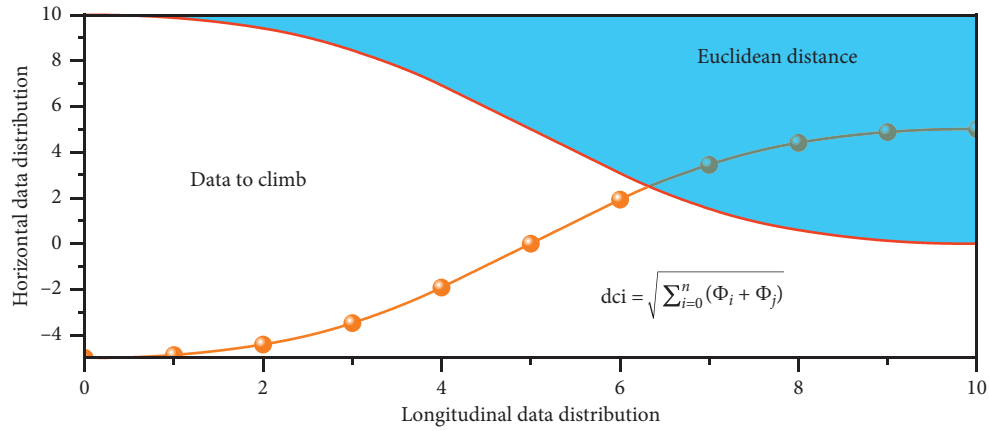

Figure 2: Spectral clustering classification diagram.

$$\mathrm{sim}_a(q) = \{k = 1 | x_o(k) - x_i(k)\}, \tag{6}$$

$$\lambda_a(i) = \frac{\max \min \Delta_a(q)}{\Delta_a(q) + \delta \max\left(\Delta i(q)/\min \Delta_a(q)\right)}. \tag{7}$$

Paul Jaccard introduced the Jaccard index, also known as the Jaccard similarity coefficient, and used it to analyze the distribution of alpine flora. The Jaccard index is used to measure the similarity between two sample sets. The Jaccard distance is a complement to the Jaccard index, and its objective is to calculate the dissimilarity between two sample sets. Its generalized formula is defined as follows:

$$q_a = \frac{\sum_{i=1}^{n} v_i \lambda}{n}. \tag{8}$$

In recent years, many Internet companies are paying more and more attention to how to use data for precision marketing services. Internationally, companies such as Walmart and Amazon have a pivotal position in the field of user behavior analysis; in China, e-commerce companies such as Taobao, Jingdong, and Jindoduo are conducting research on user behavior prediction and product recommendation for precision marketing. Through the clustering analysis of users, we can find people with different interests and behaviors, so that companies can analyze the characteristics of the core user groups of their products and dig out hidden customers to help and provide a basis for improving products and precision marketing. However, the existing user clustering algorithms cannot effectively explore the information contained in the MDF of user data. This leads to a decrease in the utilization of user data and the accuracy of similarity calculation among users. To address this problem, this paper proposes a user clustering algorithm that combines association rules with MDF. Firstly, association rules are introduced into the Jaccard distance calculation process to calculate the similarity between users, and this method improves the data utilization and the accuracy of the similarity measure. The update method of clustering centers is improved based on the idea of the *K*-mode clustering algorithm to accommodate complex data types.

4. Cluster Analysis-Based User Group Discovery Applications

4.1. Data Acquisition. With the continuous evolution of the Internet and the popularity of smart devices, people can access the Internet in various ways. While the Internet brings convenience to life, it also accumulates a large amount of data, and how to mine and analyze these data to bring out the value of data is a hot issue for research. On the other hand, in the current situation, users play a very important role. Only with the users will the enterprise have revenue and long-term development. How to effectively analyze the characteristics of users so as to provide customized services for different user groups is also a very meaningful research problem. In this chapter, two-layer structured user clustering (TL-FIUC) algorithm is applied to the real user characteristics data, and the user characteristics data are clustered and analyzed to discover the main user groups, which provide a reference for the audience user analysis of enterprises.

The dataset has been officially desensitized and the feature types are mostly ordered discrete features. Therefore, these features can be treated as numerical features when calculating user similarity and clustering centers. To verify the validity of the algorithm TL-FIUC, user_profile is first preprocessed to remove samples containing missing and abnormal values, and user features with data amounts of 1000 (Dataset6) and 10000 (Dataset7) are randomly selected from the dataset respectively as the validation dataset in this application scenario. Since the original dataset is unlabeled data, it is necessary to determine the value of the clustering number k first. In this paper, Sum of Squared Error (SSE) is used to make the judgment. First, we draw the SSE line graph of the clustering results with different values of k and then observe where the inflection point of the image is the best value of the clustering number k. In practical applications, the number of user groups found is generally not too large. This is because the purpose of clustering analysis of users by enterprises is to understand several user-product audience groups with different characteristics; if too many central users are found, it will lead to smaller differences between

each user group and cannot analyze the differences in characteristics between user groups more intuitively. Therefore, in this application for dataset Dataset6, the experiments set the value range of k to (1, 9); for dataset Dataset7, the experiments set the value range of k to (1, 20) in order to find the value of the clustering number k that has a better effect on the division of user groups. The SSE fold diagram is shown in Figure 3.

For Dataset6, it can be observed that the SSE decreases faster when the number of clusters k is in the range of (1, 4) and slows down significantly when k is in the range of (4, 9), so the value of k is set to 4; for Dataset7, it can be observed from Figure 3 that the SSE decreases faster when k is in the range of (1, 6) and slows down significantly when k is in the range of (6, 20). In practice, the number of clusters can be set according to the actual needs of the enterprise.

4.2. Precision Marketing of E-Commerce Platform. In this paper, the TL-FIUC algorithm is used to cluster Dataset6 and Dataset7 separately. The results were generated with k clusters and k clustering centers. The clustering effect of the TL-FIUC algorithm on the two datasets is shown in Figure 4. User clustering analysis is to classify users with the same or similar behavioral characteristics into the same group by means of clustering and then discover the core, larger user groups by iterative update of clusters. This chapter briefly introduces the relevant theoretical foundation and preparatory knowledge involved in the research of this paper through the algorithm definition and steps. The main topics include user clustering methods, definitions and methods related to ARM, and random forest-based feature selection methods required when considering the importance of user features. We propose a clustering method that combines association rules with MDF based on the idea of the K-mode clustering algorithm. First, this paper constructs a method to calculate the similarity between users using Jaccard distance and combines association rules with Jaccard distance to improve the similarity between users; then, a clustering center update rule for MDF is proposed; finally, the similarity measurement method combines association rules, Jaccard distance, and clustering. Finally, the basic K-mode algorithm is improved by using a similarity measure combining the association rule with Jaccard distance and the clustering center update method, which is the ARMDKM algorithm proposed in this paper. The method solves the problem that the traditional model cannot deal with MDF effectively and proves its theoretical correctness. The experiment verifies the correctness of the new method by the purity of clustering, entropy value, contour coefficient, and other indexes.

The TL-FIUC algorithm can effectively classify users with different characteristics, and the effect of this algorithm on clustering user data is significant. In summary, through cluster analysis, companies are able to obtain several major user groups of large scale. Analyzing the characteristics of these groups can provide a more intuitive understanding of users and uncover potential user groups. The TL-FIUC algorithm is applied in the cluster analysis of real data sets to discover the main user groups and the feasibility of the algorithm in the field of user clustering analysis is verified by visualization. In addition, this application experiment verifies that the optimal number of clusters for users generally does not exceed 10, as shown in Figure 5. On the one hand, the number of clusters is influenced by the number of user features: generally, the higher the number of user features, the higher the optimal number of clusters; on the other hand, it is affected by the similarity measure. The larger the number of clusters, the smaller the difference between user groups. Therefore, the number of clusters should not be set too large when companies perform clustering analysis on users in practical applications.

5. Results and Discussion

User behavior analysis is a user-centered analysis of their historical behavioral data or even ongoing behavioral actions, using techniques such as mathematical statistics or data mining. Among them, clustering analysis technology is more widely used in user behavior analysis by data analysts and researchers of various enterprises, the application scenario of clustering algorithm has been gradually expanded, and good results have been achieved. However, there are still some problems in the application of clustering algorithms in the field of user behavior analysis which need to be solved, and there are still many shortcomings in the use of clustering analysis in the field of user behavior analysis. The purpose of this paper is to solve the current problems of user clustering analysis and try to explore more application scenarios of user clustering in the process of solving the problems.

This paper solves the problem of low accuracy of user similarity calculation due to the current low data utilization. The method introduces association rules into the calculation process of Jaccard distance, constructs a user similarity measure, and improves the update method of clustering centers based on the idea of the K-mode clustering algorithm. It is verified through experiments that the ARMDKM algorithm outperforms the traditional clustering algorithm in several evaluation criteria, not only improves the utilization of data but also improves the quality of user clustering, and solves the problem that the clustering algorithm cannot effectively analyze the MDF in user data.

In addition, a TL-FIUC algorithm is proposed experimentally to consider the influence of the importance of user data features. In the field of user behavior analysis, the behavior of different users varies and the importance of different data features for user analysis varies. In order to comprehensively consider the weight relationship between user data features, this paper first uses K-means++ to analyze the data in one clustering to generate pseudolabel features, as shown in Figure 6. Then, the OOB error in the random forest algorithm is used to evaluate the feature importance to obtain the weight parameters of the features. Finally, based on the idea of spectral clustering, the weighted user data are analyzed by clustering to obtain the final clustering results. The experimental results show that the algorithm effectively improves the clustering accuracy. The TL-FIUC algorithm

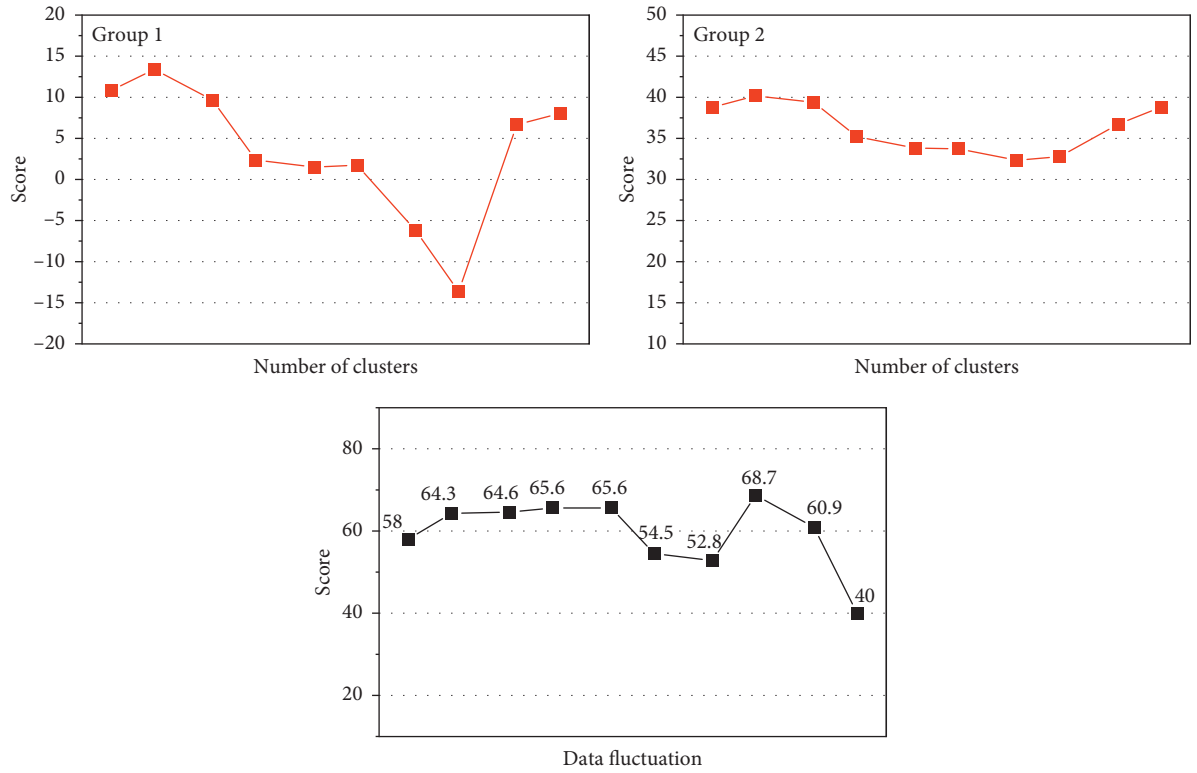

Figure 3: The SSE fold diagram.

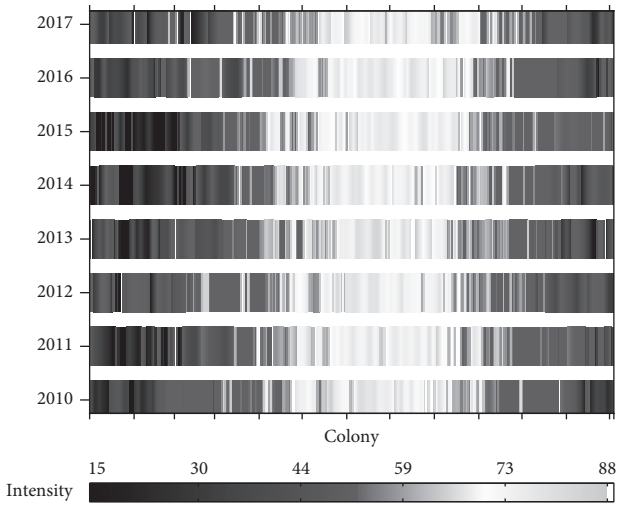

Figure 4: Clustering diagrams on the two datasets.

also performs well in the clustering analysis of real user information datasets.

Due to the nature of clustering itself, it is able to divide data according to the characteristics of the data itself, aggregating similar clusters and separating dissimilar ones. Currently, cluster analysis is widely used in business decision-making, precision marketing, and other fields. Cluster analysis can discover hidden information among users and can be applied to build more detailed user profiles and potentially discover hidden target user groups. In many data mining tasks, cluster analysis can be used as a data preprocessing approach, as shown in Figure 7. Particularly when the data are not labeled or when a simple understanding of the data distribution pattern is needed, cluster analysis can effectively accomplish the above tasks. Therefore, it is worthwhile to explore the application scenarios of clustering algorithms under different domains. The dataset used in the experiment contains only four user features selected from the user feature dataset to participate in the similarity calculation, the main purpose is to verify the feasibility and effectiveness of the algorithm in this paper, and more features can be selected for analysis in practical applications. In the clustering analysis of the LR dataset, the performance of all seven algorithms improved in the NMI index, and the TL-FIUC algorithm performed the best. On the other hand, the TL-FIUC algorithm proposed in this paper is based on spectral clustering, which can better handle sparse matrices due to its natural dimensionality reduction ability, thus improving the clustering effect. This part of the experiment shows that the TL-FIUC algorithm is a feasible method to be applied in the field of image processing.

Subsequent improvements such as parallel computing or approximation algorithms can be considered to improve the efficiency of the algorithm. In addition to the above-mentioned algorithm improvement directions, this paper can also consider the combination of fuzzy clustering of the clustering analysis of users. The basic clustering algorithms utilized in this paper are all hard clustering, and hard clustering does not reflect the characteristics of variable and flexible user behavior well. The use of soft clusters may be

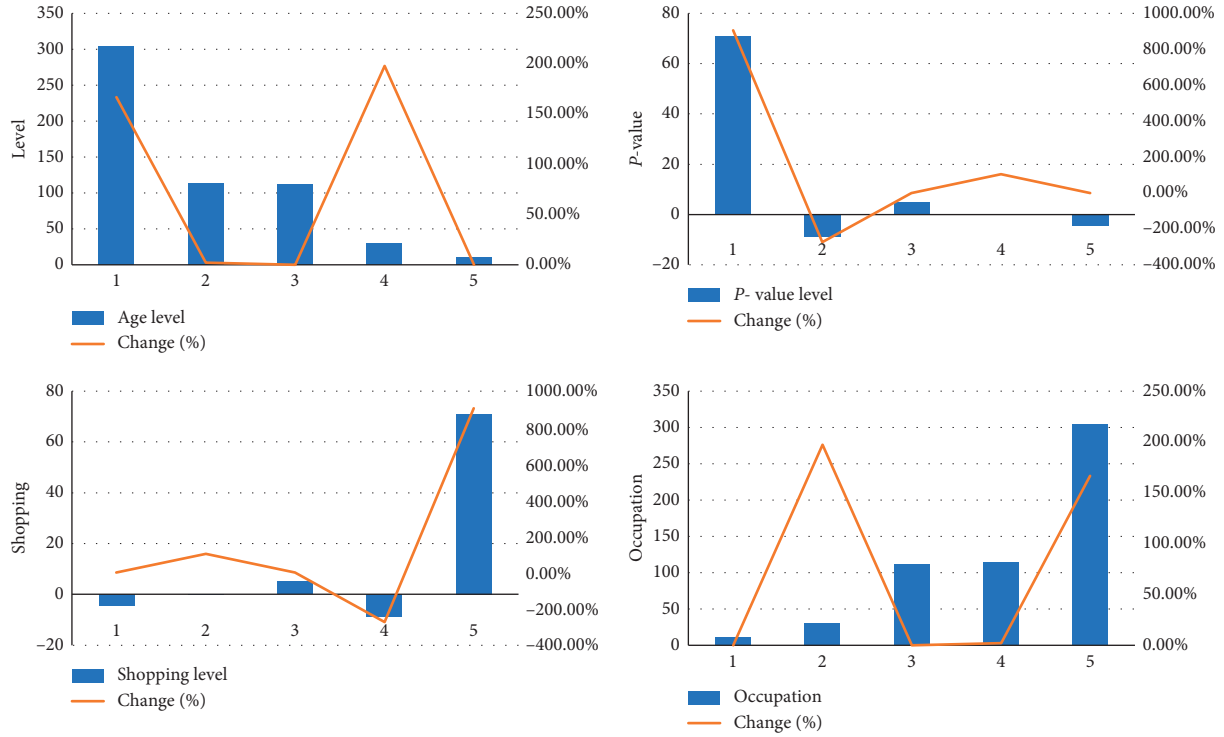

Figure 5: Six central users obtained from TL-FIUC.

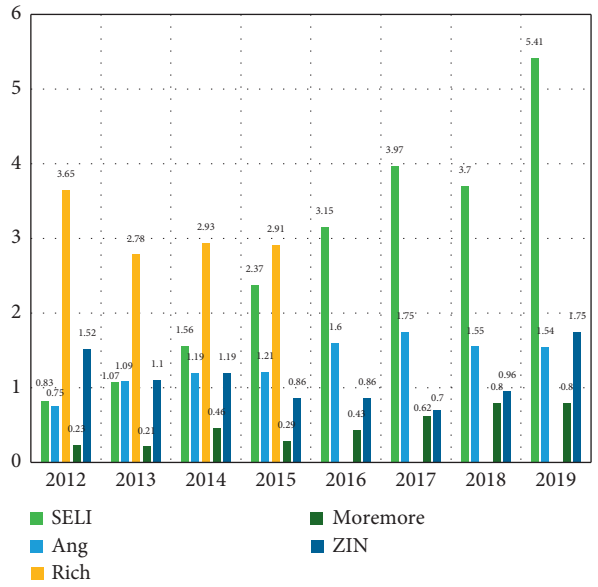

Figure 6: Performance of FMI indicators for clustering results under numerical dataset.

able to improve the quality of the clusters, which is to be verified by subsequent theories and experiments. The purpose of cluster analysis is to classify objects using the nature of the data itself, to be able to calculate the similarity between sample points according to specific definitions, and to discover internal patterns in the data through iterations in order to classify the data. The above process does not require labeling information, so clustering analysis is unsupervised learning in machine learning. By clustering, objects in the same cluster tend to be similar in some sense, while objects in different clusters tend to be different. Cluster analysis allows macroanalysis of data without data mining for a particular individual. Usually, the similarity of samples is measured based on calculating the distance between sample data.

The distance is calculated differently for different scenarios. Cluster analysis is widely used in various fields, such as group classification of target users. The target audience

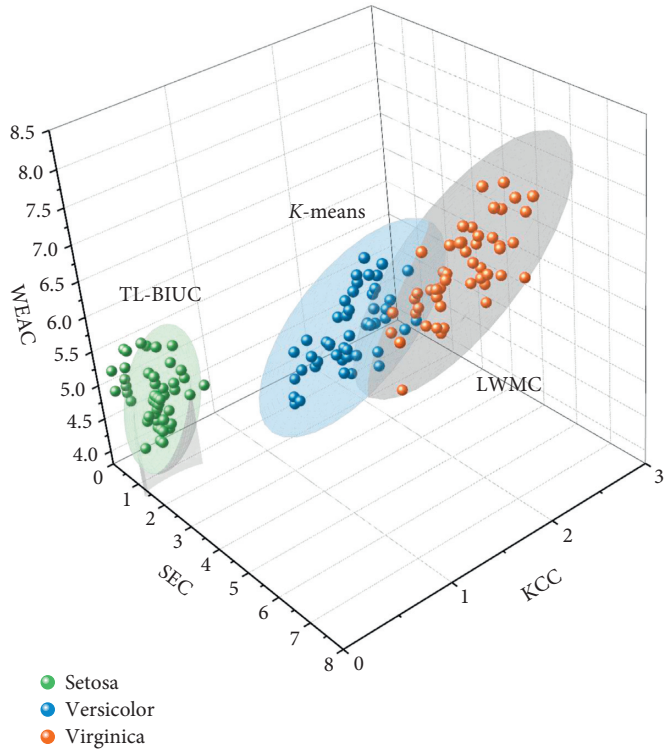

Figure 7: NMI metric performance comparison of TL-FIUC algorithm on image dataset.

group is divided into several user groups with distinct characteristics of difference, so that personalized recommendations and services for the audience group can be carried out at a later stage, which ultimately improves the efficiency and business effect of enterprise operations, as well as discovering the value combinations of different products.

6. Conclusion

The posterior pieces of association rules mined in this experiment contain only one element. Subsequent attempts can be made to combine association rules containing multiple posterior elements with similarity metrics to improve the accuracy of user similarity calculation in practical applications. The ARMDKM algorithm is based on the basic *K*-mode algorithm, which is not improved for the initialization problem and requires multiple runs to obtain better results. In addition, the data features of this experiment are selected by hand, and the ARMDKM algorithm has no ability to select features. Subsequent methods can be combined with existing methods to solve the initialization and feature selection problems to achieve better results. The algorithm itself contains two major steps: one is unsupervised feature selection and the other is overall clustering analysis; the unsupervised feature selection method uses one clustering analysis and one classification model construction, the algorithms used are relatively primitive, and there is still a lot of room for optimization. The algorithm can be improved by referring to the latest papers on the direction of unsupervised feature selection; the overall clustering analysis part uses the spectral clustering algorithm, which is more effective in dealing with high-dimensional data, but because spectral clustering needs to calculate the similarity between each sample, resulting in excessive time overhead in dealing with large sample data.

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 competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] N. Huang, "Analysis and design of university teaching evaluation system based on jsp platform," *International Journal of Education and Management Engineering*, vol. 7, no. 3, pp. 43–50, 2017.

[2] Q. Wang, C. Wu, and Y. Sun, "Evaluating corporate social responsibility of airlines using entropy weight and grey relation analysis," *Journal of Air Transport Management*, vol. 42, pp. 55–62, 2015.

[3] T. S. Riall, J. Teiman, M Chang et al., "Maintaining the fire but avoiding burnout: implementation and evaluation of a resident well-being program," *Journal of the American College of Surgeons*, vol. 226, no. 4, pp. 369–379, 2017.

[4] T. Singh, A. Patnaik, and R. Chauhan, "Optimization of tribological properties of cement kiln dust-filled brake pad

using grey relation analysis," *Materials & Design*, vol. 89, pp. 1335–1342, 2016.

[5] R. Tan, W. Zhang, and C. Shengqun, "Decision-making method based on grey relation analysis and trapezoidal fuzzy neutrosophic numbers under double incomplete information and its application in typhoon disaster assessment," *IEEE Access*, vol. 8, pp. 3606–3628, 2019.

[6] T. Wang, "Study on adhesion property of asphalt and aggregate based on grey relation," *Theory*, vol. 48, no. 14, pp. 40–42, 2019.

[7] J. W. Boland, M. Brown, A. Duenas, G. M. Finn, and J. Gibbins, "How effective is undergraduate palliative care teaching for medical students?," *A Systematic Literature Review*, vol. 10, no. 9, pp. 036458-036459, 2020.

[8] J. Teaching and L. Practice, "A practice-based study of Chinese students learning—putting things together," *Journal of University Teaching and Learning Practice (JUTLP)*, vol. 16, no. 2, pp. 12–18, 2019.

[9] X. Wang, "Application of grey relation analysis theory to choose high reliability of the network node," *Journal of Physics Conference Series*, vol. 1237, no. 3, pp. 032055-032056, 2019.

[10] M. Dinerstein, L. Einav, J. Levin, and N. Sundaresan, "Consumer price search and platform design in internet commerce," *American Economic Review*, vol. 108, no. 7, pp. 1820–1859, 2018.

[11] L. Chen, S. Qiao, N. Han et al., "Friendship prediction model based on factor graphs integrating geographical location," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 3, pp. 193–199, 2020.

[12] Q. Wang, Y. Yu, H. Gao et al., "Network representation learning enhanced recommendation algorithm," *IEEE Access*, vol. 7, pp. 61388–61399, 2019.

[13] Y. Cen, J. Zhang, G. Wang et al., "Trust relationship prediction in alibaba E-commerce platform," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 5, pp. 1024–1035, 2020.

[14] E. Cristobal-Fransi, Y. Montegut-Salla, B. Ferrer-Rosell, and N. Daries, "Rural cooperatives in the digital age: an analysis of the internet presence and degree of maturity of agri-food cooperatives' E-commerce," *Journal of Rural Studies*, vol. 74, pp. 55–66, 2020.

[15] Z. H. Borbora, M. A. Ahmad, J. Oh, K. Z. Haigh, J. Srivastava, and Z. Wen, "Robust features of trust in social networks," *Social Network Analysis and Mining*, vol. 3, no. 4, pp. 981–999, 2013.

[16] P. M. Carron, K. Kaski, and R. Dunbar, "Calling Dunbar's numbers," *Social Networks*, vol. 47, pp. 151–155, 2016.

[17] J. Kwak, Y. Zhang, and J. Yu, "Legitimacy building and e-commerce platform development in China: the experience of Alibaba," *Technological Forecasting and Social Change*, vol. 139, pp. 115–124, 2019.

[18] Z. Almeraj, F. Boujarwah, D. Alhuwail, and R. Qadri, "Evaluating the accessibility of higher education institution websites in the state of Kuwait: empirical evidence," *Universal Access in the Information Society*, vol. 1, pp. 11–18, 2020.

[19] R. Gonçalves, T. Rocha, J. Martins, F. Branco, and M. Au-Yong-Oliveira, "Evaluation of E-commerce websites accessibility and usability: an E-commerce platform analysis with the inclusion of blind users," *Universal Access in the Information Society*, vol. 17, no. 3, pp. 567–583, 2018.

[20] A. Ismail, K. S. Kuppusamy, and S. Paiva, "Accessibility analysis of higher education institution websites of Portugal," *Universal Access in the Information Society*, vol. 19, no. 3, pp. 685–700, 2020.