

Retraction

Retracted: Using Clustering Analysis and Association Rule Technology in Cross-Marketing

Complexity

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

Copyright © 2024 Complexity. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Y. Cheng, M. Cheng, T. Pang, and S. Liu, "Using Clustering Analysis and Association Rule Technology in Cross-Marketing," *Complexity*, vol. 2021, Article ID 9979874, 11 pages, 2021.

Research Article

Using Clustering Analysis and Association Rule Technology in Cross-Marketing

Yang Cheng ¹, Ming Cheng,¹ Tao Pang,² and Sizhen Liu³

¹School of Journalism and Communication, Wuhan University, Wuhan, Hubei 430072, China

²Beijing Navroom Technology Co., Ltd., Beijing 100000, China

³Changjiang Ecological Environmental Protection Group Co., Ltd., Wuhan, Hubei 430000, China

Correspondence should be addressed to Yang Cheng; 2019101030024@whu.edu.cn

Received 30 March 2021; Revised 10 April 2021; Accepted 15 April 2021; Published 22 April 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Yang Cheng et al. 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 this paper, according to the perspective of customers and products, by using clustering analysis and association rule technology, this paper proposes a cross-marketing model based on an improved sequential pattern mining algorithm, where an improved algorithm AP (Apriori all PrefixSpan) is applied. The algorithm can reduce the time cost of constructing a projection database and the influence of the increase of support on the algorithm efficiency. The improved idea is that when the first partition is used to generate the projection database, the number of itemsets in the projection database is sorted from small to large, and when the second partition is used, the sequence patterns are generated directly from the mined sequence patterns, so as to reduce the construction of the database. The experimental results show that this method can quickly mine the effective information in complex data sets, improve the accuracy and efficiency of data mining, and occupy less memory consumption, which has good theoretical value and application value.

1. Introduction

With the continuous development and progress of science, technology, and economy, the worldwide industrial competition is becoming more and more fierce, and the business model, market environment, and competition model have undergone fundamental changes [1]. This change is more obvious in the information service industry. Providing new products and services to the existing customers, namely, cross-marketing, plays an important role in expanding profits. The key of cross-marketing is to provide the most suitable products and services to the existing customers so that the services accepted by customers can bring the greatest benefits to both the seller and the buyer, which is of great significance to the transformation of the customer-centered business philosophy. This reason is very easy to understand. However, only after we find a very accurate model, can we sell specific types of goods to the right customers and make profits from them [2].

Generally speaking, considering how to tap potential cross-marketing opportunities can start from two directions: one is from business and the other is from customers [3]. Identifying cross-marketing opportunities from customer analysis is based on the consumption characteristics of existing customers as the basis of forecasting cross-marketing, studying the purchase differences between different customer groups, so as to recommend specific types of commodity combinations. To identify cross-marketing opportunities from the perspective of business is to analyse the business characteristics to find out the existing users who meet the characteristics and recommend them [4].

This paper presents an improved PrefixSpan algorithm based on the Apriori algorithm (IPrefixSpan scheme). The idea of the algorithm is to generate the needed sequential patterns directly from the mined sequential patterns to reduce the construction of the projection database. The more the sequences that have been mined, the faster the mining speed, and there is no special requirement for the data form. The algorithm combines the advantages of the Apriori algorithm and reduces the influence of increasing support on efficiency.

In Section 2, the related works about cross-marketing is introduced, where potential characteristics model, NPTB model, and market mining model are classified to better describe the background of cross-marketing. Section 3 introduces the structure of cross-marketing model; in addition, an improved PrefixSpan algorithm combined with the Apriori method is combined. Section 4 is the simulation and analysis of the experiment. Section 5 is the conclusion.

2. Related Works

In recent years, more and more scholars at home and abroad study cross-marketing, and many scholars are committed to the research of cross-marketing identification methods. At present, the methods and models of cross-marketing opportunity identification mainly include the following three models: the potential characteristics model, NPTB model, and market mining model.

In the past, the research on cross-marketing was mainly concentrated in Europe, the United States, and other developed countries. The main reason is that the market competition in Europe and the United States and other developed countries is fiercer, and the traditional marketing model can no longer make enterprises maintain a greater advantage in the market competition. Enterprises need to find a new marketing model to participate in the market competition. At this time, cross-marketing will quickly enter the relevant enterprises and research. This also brings unprecedented opportunities to the research of cross-marketing [5]. Literature [6] points out that cross-marketing is to provide the right products for the right customers at the right time, and the original transaction data of customers can help us to achieve the above goals because such data can enable enterprises to realize the actual needs of customers through the purchase behaviour of similar customers. However, the database usually only contains transaction data but does not contain the data of related products in the market [7]. In addition, the information extracted from the database often relies on data mining technology, which also makes the data information lag far behind the data collection and storage, making part of the data missing, resulting in inaccurate prediction [8]. In view of these drawbacks, literature [9] proposes to apply a new data augmentation technology to predict customers' purchase of new products, that is, to select a mixed data factor analysis technology on the basis of existing customer transactions to predict the most valuable potential customers of products, so as to further implement cross-marketing. Sequential pattern mining [10] is to mine frequent sequential events or subsequence. Sequential pattern mining is widely used because it relies less on prior knowledge and can find unknown rules.

In [11], a spade algorithm based on vertical data format is proposed. The above algorithms will produce a large number of candidate sets. However, the FreeSpan algorithm proposed in [12] is based on the growth of sequential patterns and does not produce candidate sets. The PrefixSpan algorithm in [13] is an improvement of the FreeSpan

algorithm, which reduces the connection times of the projection database and subsequence and makes the database converge faster, and the efficiency of the algorithm is higher than that of previous algorithms. The PrefixSpan algorithm generates the corresponding projection database according to the prefix and then scans the projection database to avoid scanning the whole database, thus reducing the scanning time. The main time cost of the algorithm is to build the projection database, and with the increase of support, the efficiency will decrease. Due to the improvement of support, the convergence of the projection database is reduced. In [14], the construction of the database is improved, but the data form is too high. Wang et al. [15] improved the memory storage; when the support increases, the efficiency decreases.

2.1. Potential Characteristics Model. Jiang et al. [16] proposed to find potential users suitable for cross-marketing by analysing potential characteristics; the principle structure of the potential characteristics model is shown in Figure 1. Any theory of latent traits assumes that individual behaviour can be explained by specific personal characteristics and predict or explain the behaviour or performance in relevant situations through numerical calculation of these characteristics. This paper uses trait theory to predict cross-marketing opportunities and uses users' views on business or service characteristics and other characteristics related to business or service to predict their possibility of using the business or service. The latent trait model provides a mainstream research direction for the follow-up research of cross-marketing. However, the latent trait model proposed by Chen et al. [17] requires enterprises to understand the situation that each user consumes the business of their own enterprises and competitors, which is difficult to achieve in reality. Therefore, literature [18] puts forward a comprehensive data factor analysis model to deal with the extended data, mainly according to the sample survey data to deal with the investigated samples. In its extended model, four different types of power exponential function distribution models are used, namely, Bernoulli distribution is used to represent binary service use items, the binomial distribution is used to represent satisfaction ranking, the Poisson distribution is used to represent service use frequency, and the normal distribution is used to represent transaction volume. The concentration coefficient is used to summarize the ability of the model to predict cross-marketing opportunities.

2.2. NPTB Model. In literature [19], the NPTB (next product to buy) model is proposed to improve the effectiveness of cross-marketing. The empirical research results of Knott et al. show that the cross-marketing prediction results of the NPTB model (Figure 2) are more effective than the heuristic algorithm in improving the sales rate of enterprises.

Note: X_j represents the user data, including the user's current business, demographic variables, and other related variables. V_j represents the measured user's demand for

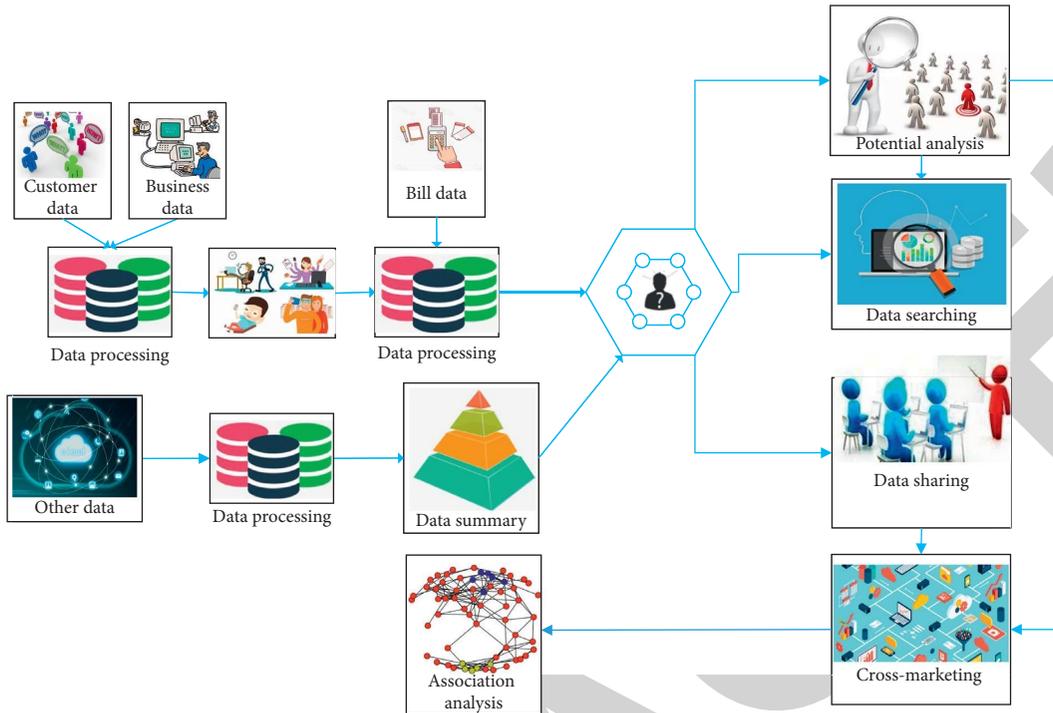


FIGURE 1: The structure of potential characteristics model.

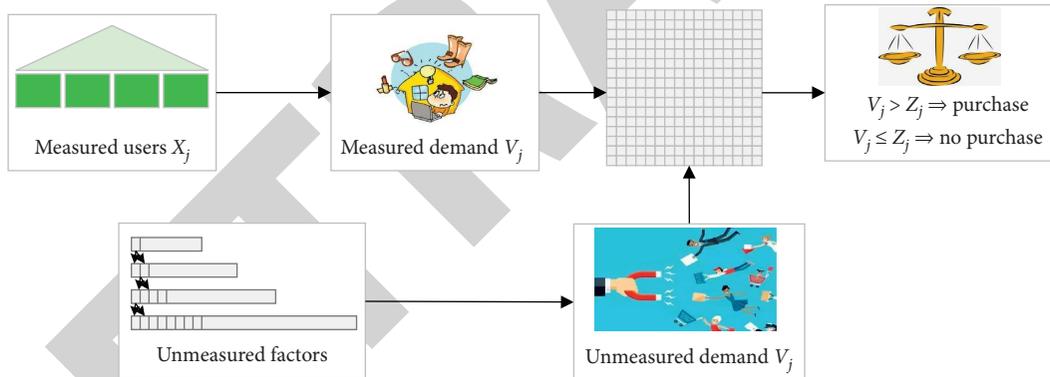


FIGURE 2: The structure of the NPTB model.

purchasing business, and Z_j represents the unmeasured factors that inhibit the purchase of business, such as the user’s failure to recognize this demand or the factors caused by the marketing efforts of competitors.

In addition, literature [20] proposed that retailers should tailor different sales plans for different customers and supplement and improve the NPTB model on the basis of preferential purchase model and NPTB model on the premise of continuously reducing sales cost in order to effectively recommend different products to different customers at different times. Literature [21] puts forward the methods of random forest, polynomial logarithm, and random polynomial logarithm in the retail results of home appliance retail enterprises to analyse and classify the large-scale customer purchase data of home appliance retail enterprises and study the cross-purchase behaviour of customers, so as to better help enterprises formulate cross-sales strategies and increase sales volume.

2.3. Market Mining Model Improved Sequential Pattern Mining Algorithm. The viewpoint of using the market segmentation method to forecast cross-marketing is put forward in literature [22]. Its choice of market segmentation variables is interactive psychological segmentation variables, including consumption motivation, consumption preference, attitude, and values. Based on the questionnaire survey of psychological variables of sample users randomly selected from the enterprise database, this paper subdivides the users, analyses the demographic characteristics of each subdivided group, and then establishes a scoring model to predict cross-marketing opportunities.

The Bayesian network is proposed to classify products in cross-marketing. The Bayesian network represents the joint probability distribution of a group of discrete random variables: it is considered to be a box probability model composed of partial qualitative conditional dependence

between a variable and the conditional probability of the group of variables specified by its quantitative part. Then, literature [23] proposes to use the advantages of the dynamic Bayesian network to support the cross-marketing behaviour of financial service companies. The dynamic Bayesian network establishes a dynamic system model based on the Bayesian network, so as to develop the independence of conditions. On this basis, it optimizes the effectiveness of obtaining information in the process of cross-marketing and increases the success rate of cross-marketing. Literature [24] proposes a personalized recommendation model based on domain knowledge, which is applied to the cross-marketing strategy of enterprises. In the application process of this model, we should first preprocess the customer domain knowledge cluster (collaborative filtering method can be used), then propose to combine the related products in the domain knowledge and form a recommendation list, and finally refine the recommendation list to find out the most favourable products for cross-marketing. Furthermore, literature [25] proposed the use of the multiple credit method to comprehensively predict the sales risk of related products, so as to help financial enterprises choose customers with profit expectations to cross-marketing products.

3. Cross-Marketing Model Research Based on Improved Sequential Pattern Mining Algorithm

3.1. The Structure Design of Cross-Marketing Model. The sequential pattern mining model based on clustering in marketing business data is composed of a data acquisition module, data preprocessing module, data storage module, decision support module, and user recommendation model. The structure of the model is shown in Figure 3.

User recommendation layer (also called customer layer): it is the user interaction interface for users to use the functions and services of the analytical CRM system. Its function is to accept user requests and deal with the platform of user interaction. The dynamic page is automatically generated by the web service layer, and the web browser is used to submit the user's request and display the page generated by the web layer. But it does not perform the functions of querying database and complex business rules.

Database layer: the backstage database server of the analytical CRM system, which represents enterprise information resources, including transaction monitor, relational database, and various customized applications. Its function is to manage the metadata of each part, provide the corresponding interface, and complete the creation, maintenance, and access of data sources such as data warehouse. In the design of this system, the relational database SQL Server is used as the background database of analytical CRM, and the data warehouse is established based on it. At the same time, in order to better extract the required basic data and meet the requirements of data backup, a data extraction, conversion, and loading (ETL) server is added between the data warehouse and the data warehouse management server. The server extracts the required data from the data centre;

standardizes the name, code, number, and form of data items; and eliminates duplicate data.

Data preprocessing layer: it mainly includes data extraction, data cleaning, data reduction and normalization, user identification, and path identification. The main task of this layer is to preprocess the collected structured data, semistructured data, and irregular data and remove the duplicate data and invalid data.

Pattern discovery layer: it includes clustering mining, sequence mining, and OLTP module; decision support is to analyse and evaluate the results. The main work of this layer is to deeply mine users' needs and potential needs through association, clustering, and OLTP operations and recommend urgently needed products or services for users.

Information collection layer: according to the needs of this paper, a large number of customer basic information data and customer behaviour data are collected from the unified customer resource subsystem, billing subsystem, and integrated business accounting subsystem. The data contains the detailed business behaviour and accounting information of customers; through statistical analysis of these data, we can get the relevant attributes needed for the research.

3.2. Data Mining Algorithm Based on Improved Sequential Pattern Mining Algorithm. Data mining algorithm based on improved sequential pattern mining algorithm is explained in this section. The fusion idea of the two algorithms is to use PrefixSpan to generate projection and use the Apriori-all algorithm to further process in the projection area. Next, the improved scheme of the data mining algorithm is illustrated in this section.

Suppose that the transaction database DB is a set of sequences, and the data sequence is (cid, ds) . $I = \{i_1, i_2, \dots, i_n\}$ is the set of all items; itemset $t = \{e_1, e_2, \dots, e_m\}$ is a subset of I . A sequence is an ordered list of itemsets, denoted as $s = \langle s_1, s_2, \dots, s_w \rangle$, where s_j represents an itemset. The number of times an item appears in a sequence is called the length of the sequence. Usually, an item can appear at most once in any itemset of any sequence, but it can appear in different itemsets of the sequence. A sequence of length k is called a k -sequence.

Defining: $a = \langle a_1, a_2, \dots, a_n \rangle$, $\beta = \langle b_1, b_2, \dots, b_m \rangle$. The transaction of a client in DB is called the data sequence. If data sequence a is a subsequence of sequence s , then s contains a . The support of a is the ratio of the number of sequences containing a in DB to the total number of sequences in DB, denoted as a_{sup} . Furthermore, minimum support $\min sup$ is defined as the threshold specified by the user. If $a_{sup} \geq \min sup$ is satisfied, then a is called the sequential pattern. Sequential pattern mining is to find out all the sequential patterns in DB.

In this paper, an improved sequential pattern mining algorithm is proposed based on the Apriori-all algorithm and PrefixSpan algorithm. The idea of this method: if the sequence pattern set with a sequence pattern $\langle a \rangle$ as the prefix is known as β and the corresponding projection database $s|a$, the sequence pattern in the sequence pattern set β is taken as the candidate set, and the projection database $s|$

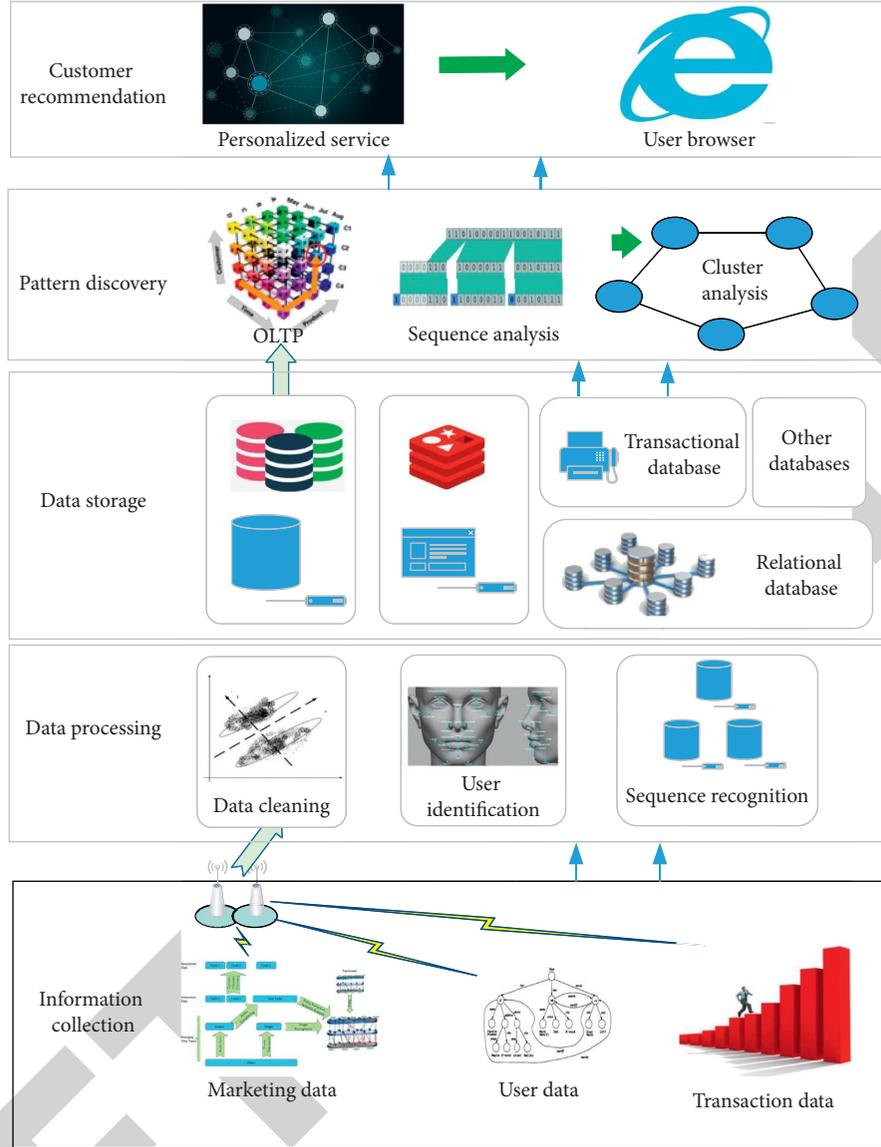


FIGURE 3: The structure of the cross-marketing strategy based on data mining algorithm.

A is scanned to verify whether the number of each sequence pattern in the candidate set is greater than the support to generate the sequence pattern set with sequence pattern $\langle a \rangle$ in the candidate set. According to the characteristics of sequence patterns generated by PrefixSpan, if a sequence does not satisfy the support degree, then the sequence prefixed with the sequence does not satisfy the support degree. Therefore, when verifying the candidate set, if a sequence does not meet the support, then the sequence prefixed with the sequence does not need to be verified.

Giving $s = \langle s_1, s_2, \dots, s_w \rangle$, the weight is calculated as follows:

$$\omega(s_i) = \alpha f(s_i) + (1 - \alpha)t(s_i), \quad (1)$$

where $\omega(s_i)$ is the weight of record s_i , $f(s_i)$ is the usage frequency of record s_i in usage record s , $t(s_i)$ is the usage duration of record s_i in usage record s , and α is the weight

parameter, which is used to weigh the usage frequency and usage duration.

Given a sequence $s = \langle s_1, s_2, \dots, s_w \rangle$ and weight set $\omega = (\omega(s_1), \omega(s_2), \dots, \omega(s_n))$, then the weight of sequence s is calculated as follows:

$$W(s) = \frac{\sum_{i=1}^k \omega(s_i)}{n}. \quad (2)$$

The whole structure of the improved sequential pattern mining algorithm can be designed as shown in Figure 4. As shown, for sequential pattern mining, let DB represent the original database; let DB represent the incremental database, that is, the new data added in the database, including the new transaction and data sequence. ud stands for the updated database. The customer number cids in ud may already exist in the DB, or it may be a new customer. In addition, start mining from the smallest projection database, scan the

projection database to get the corresponding length-2 sequence pattern, and then divide it with length-2 as the prefix. At this time, scan the result data set to determine whether the length-1 prefixed by the sequence is mined. If it has been mined, use the YZ method to generate the required sequence pattern directly from the length-1 sequence set. If there is no PrefixSpan algorithm. When finding the sequence pattern set prefixed with length-2, the sequence pattern set prefixed with length-2 is included in the sequence pattern set prefixed with length-1, so it can be generated directly from the YZ method of length-1. Then, the time to generate a sequence from length-1 is less than that of the PrefixSpan algorithm.

The detailed evaluation indexes are defined as follows:

Index 1: accuracy is the ratio of the number of correct predictions in the app prediction process to the total number of predictions. The calculation formula is as follows:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \quad (3)$$

where TP means the prediction is positive, and in fact the prediction is correct, that is, the correct rate of judging as positive; TN means the prediction is negative, but actually the prediction is correct, that is, the correct rate of judging negative; FP means positive prediction, false prediction, false positive rate, that is, negative prediction is positive; FN means the prediction is negative, in fact, the prediction error, the missing rate; that is, the positive judgment is called negative. In general, the higher the accuracy of the model, the better the effect of the model.

Index 2: training time means that, in the experiment, the shorter the running time of model training, the less the resource occupied, the less the impact on users, and the better the algorithm.

4. Simulation Results and Performance Analysis

4.1. Data Sources and Simulation Setting. The experimental data is the standard synthetic transaction data, and the generation process is the same as that in literature [9]. The relevant parameters of the test data set are $|D|$ for the number of customers, set to 10K; $|C|$ for the average number of transactions of customers, set to 10; $|t|$ for the average number of transactions, set to 2.5; and $|n|$ for the total number of items set to 1000. $|s|$ denotes the average length of the longest frequent sequence, set to 4; $|ns|$ denotes the number of the longest frequent sequences, set to 1000; and $|N_i|$ denotes the number of the longest frequent itemsets, set to 5000. The above parameters are used to generate UD.

First, the updated database UD with the number of customers is generated. We set three parameters to simulate the various updating situations of real transaction data more in line with the actual situation.

Parameter 1: update rate r_{inc} , $|DB| = |UD| \times r_{inc}$. Generate $|DB|$ nonrepeating random numbers in the range of 1 to $|UD|$ and use these random numbers as the customer numbers appearing in DB. Use $|UD| - |DB|$

customer numbers as customer numbers in DB.

Parameter 2: return rate r_{nc} , number of old customers $|old| = |DB| \times r_{nc}$, which is randomly selected from $|DB|$. This part of the data sequence will be further divided into two parts, namely, the transactions of the same customer in the DB and the transactions in the new DB. And this ratio is controlled by parameter 3, transaction additional ratio.

Parameter 3: weight parameter α . It is used to weigh the frequency and duration of use.

We use VC++ 6.0 to implement fspm and prefix span algorithm on a machine with 512 M memory, 866 mhz CPU, and Windows 2000 operating system. The IPrefixSpan algorithm is compared with PrefixSpan algorithm.

4.2. The Optimal Selection of Algorithm Parameter. The weight parameter is an important parameter in the whole algorithm. In this paper, we use the experimental method to get the optimal parameter. First of all, initialize $\alpha = 0.1$ and $0 \leq \alpha \leq 1$, and use the polling method to obtain $\alpha = 0.1$ with the highest accuracy. At this time, α is the best parameter of usage frequency and app usage duration. The simulation results are as follows.

It can be seen from Figure 5 that the accuracy changes with the change of weight parameter α (i.e., the measurement variable of usage time and frequency). When it is less than 0.5, the accuracy rate increases with the increase of time. When the parameter is greater than 0.5, the accuracy shows a downward trend. Therefore, when $\alpha \approx 0.5$, the accuracy rate is the highest, which is the best value to measure the proportion of users' use time and frequency.

In order to get the optimal return rate and update rate parameters, we have carried out the following experiments. A series of different parameters, return rate, and update rate are designed under different support degrees, and the optimal parameters are obtained by using two indicators of accuracy and running time. The simulation results are shown in Figure 6.

As is shown in Figure 6, the support degrees are set as 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, and 9%; return rate are set as 20%, 30%, 40%, and 50%; and the update rate are set as 20%, 30%, 40%, 50%, and 60%.

As is shown in Figure 6, simulation results show that, with the increase of support, the execution time of the algorithm decreases first and then increases. And when the return rate remains unchanged, no matter how large the update rate is selected, the execution time is almost the same, indicating that the update rate has no great impact on the

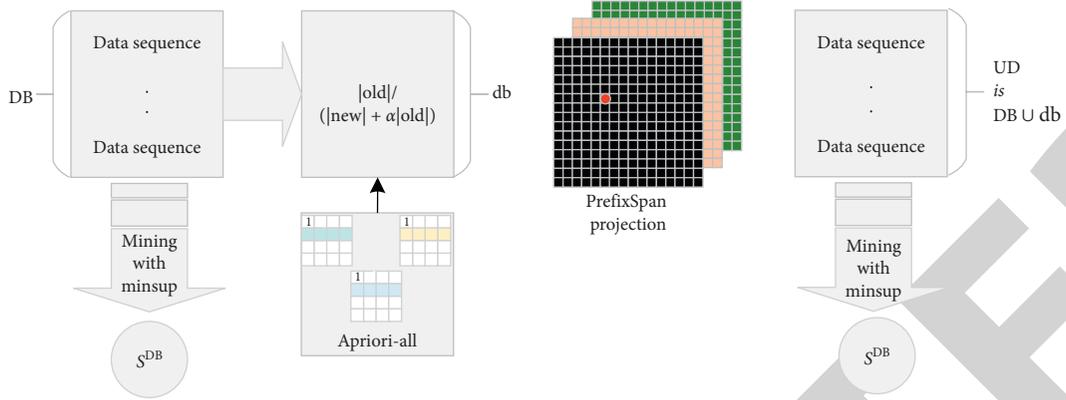


FIGURE 4: The structure of improved sequential pattern mining algorithm.

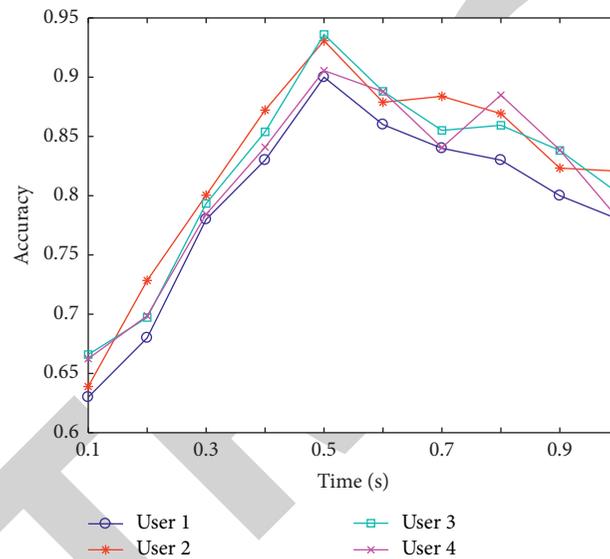


FIGURE 5: The structure of improved sequential pattern mining algorithm.

execution time. In addition, the minimum execution time is selected, and the update rate and return rate are 30% and 40%, respectively. The support was 2%. What needs to be specially explained is the data obtained by this group of laboratories when the weight parameter is 5%.

4.3. The Accuracy Validation of Sequence Mining. In order to further verify the performance of the algorithm, we use VC++ 6.0 to implement the improved sequence mining method on the host computer with 512 MB memory, Pentium iii-733 mhz CPU, and Windows 2000 Professional operating system. Taking the above data set as an example, after data mining, the data display diagram is as follows: among them, Figure 7(a) is part of the original data, Figure 7(b) is the data mining results figure of the improved PrefixSpan, and Figure 7(c) is the data thermal energy diagram after mining. Figure 7(d) is the probability figure of user habits.

It can be seen from the figure that this paper combines the PrefixSpan method to design the appropriate weight

coefficient. After obtaining the appropriate parameters, the data mining results have achieved good results. It can be seen from Figure 7(b) that this method can quickly classify and mine data from complex data. The energy figure and probability figure in Figure 7(c) and Figure 7(d) further illustrate the high accuracy of the algorithm. It can make the data mining results more accurate.

4.4. The Superiority Validation of the Proposed Scheme. The experimental environment and test data set are the same as experiment one, and the test data sets are supported by 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, and 11%, respectively. The experimental results are shown in Figure 8. It can be seen from Figure 8 that the improved prefix span algorithm is obviously better than the prefix span algorithm when the support is between 4% and 8%. Section 4.2 shows that, with the increase of support, the ratio of the time used by the improved prefix span method to the time used by the prefix span method is getting smaller and smaller, while Figure 8 shows that the time

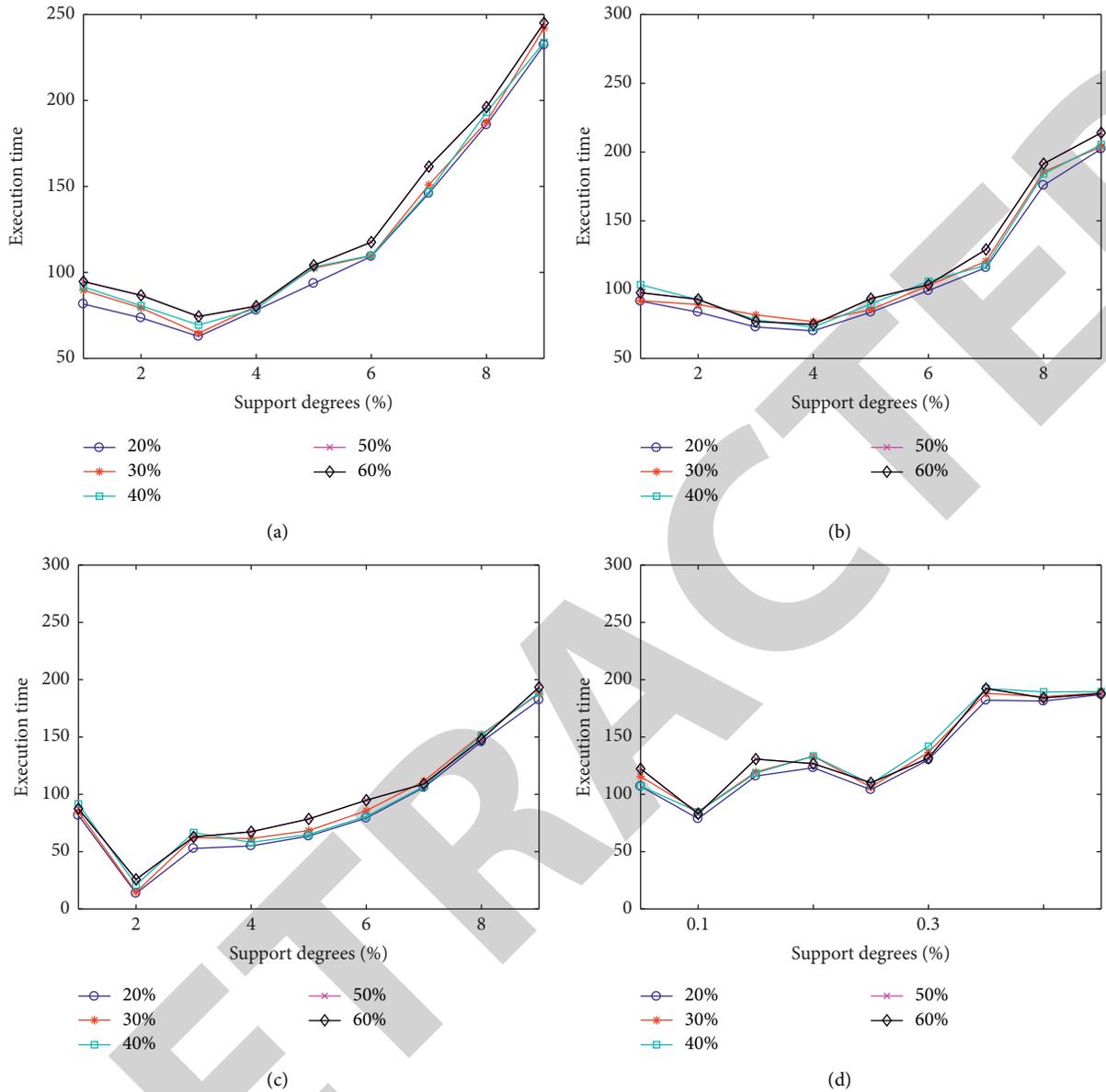


FIGURE 6: The execution time under different parameters. (a) 20% return rate. (b) 30% return rate. (c) 40% return rate. (d) 50% return rate.

gap between the prefix span algorithm and the improved prefix span algorithm is decreasing after the support exceeds 8%. The main reason is that, with the increase of support, the number of sequential patterns decreases, the total time used by the algorithm decreases, and the time gap between the two algorithms becomes smaller. The

experimental results show that the improved prefix span algorithm is better than the prefix span algorithm. In addition, the accuracy of data mining under different supports is shown in Figure 9. From Figure 9, it can be seen that the two have the same trend, and the accuracy of IPrefixSpan is significantly higher than that of PrefixSpan.

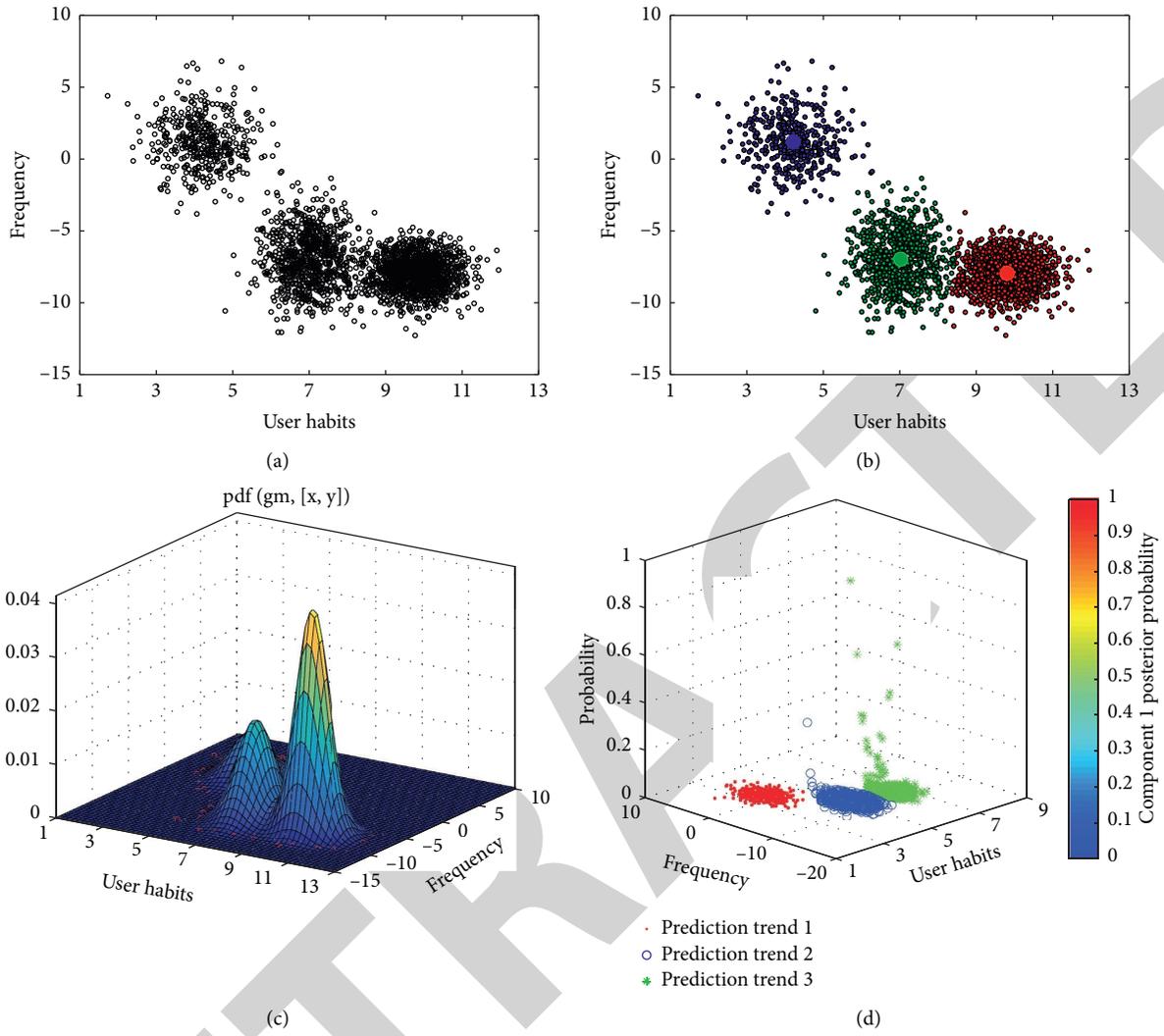
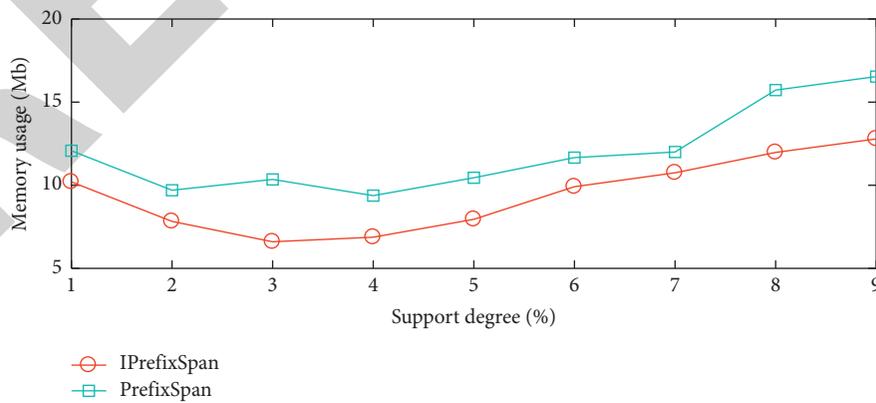


FIGURE 7: The simulation results of the proposed method. (a) Original data. (b) Data mining results. (c) Thermal energy diagram. (d) Probability figure of user habits.



(a)

FIGURE 8: Continued.

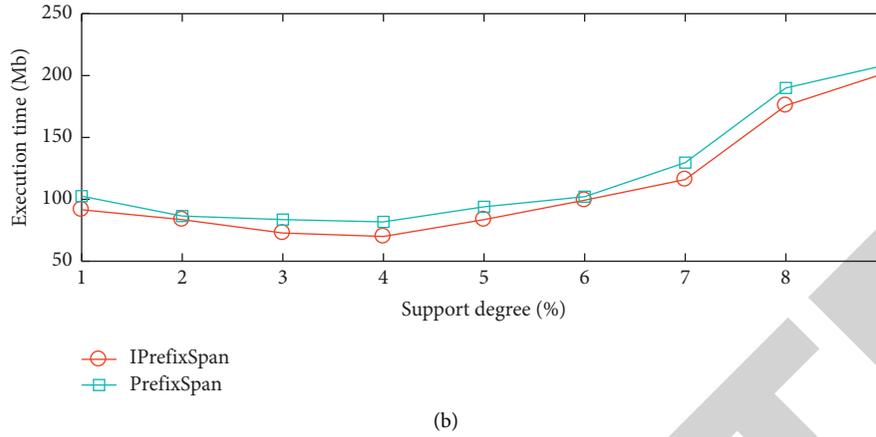


FIGURE 8: The comparison results compared with the PrefixSpan method.

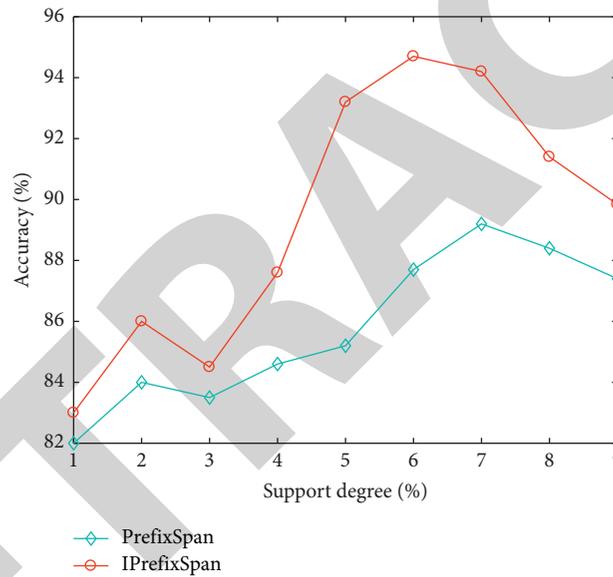


FIGURE 9: The comparison results compared with the PrefixSpan method.

5. Conclusion

Based on the research of the PrefixSpan algorithm, this paper studies that the cost of the PrefixSpan algorithm mainly lies in the construction of subdatabases and also studies the idea of the Apriori algorithm. The Apriori algorithm is efficient in verifying candidate sets. Based on the characteristics of the sequences generated by the PrefixSpan algorithm, this paper improves the verification method, improves the PrefixSpan algorithm, and reduces the impact on the efficiency of the algorithm with the increase of support. In addition, the weight coefficient can significantly improve the efficiency of the algorithm when updating data, and it can also reduce the running time of the algorithm by adjusting the weight coefficient. Simulation results show that this method can achieve a good data mining effect in marketing data. The algorithm can

reduce the time cost of building projection database and reduce the impact of support increase on the efficiency of the algorithm. The improved idea is that when the first partition is used to generate the projection database, the number of itemsets in the projection database is sorted from small to large, and when the second partition is used, the sequence patterns are generated directly from the mined sequence patterns, so as to reduce the construction of the database. Furthermore, this paper presents a basic algorithm of time series data mining, which can be used for any time series data set, for example, transportation, weather, and other fields.

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

References

- [1] C. Sun, B. Gu, and J. Zhang, "Ecosystem service value assessment based on clustering analysis and ESV algorithm," *World Scientific Research Journal*, vol. 6, no. 5, pp. 54–61, 2020.
- [2] J. Jokinen, T. Rty, and T. Lintonen, "Clustering structure analysis in time-series data with density-based clusterability measure," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 45–56, 2019.
- [3] S. Li, X. Song, H. Lu, L. Zeng, F. Liu, and M. Shi, "Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm," *Expert Systems with Applications*, vol. 139, Article ID 112839, 2019.
- [4] C. Torrecillas, I. B. Guerrero, and R. P. Jiménez, "Clustering applied to the analysis of ground deformation," in *XIV International Congress on Graphic Expression Applied to Building*, Sevilla, Spain, February 2019.
- [5] P. Sadhukhan, "Performance analysis of clustering-based fingerprinting localization systems," *Wireless Networks*, vol. 25, no. 5, pp. 2497–2510, 2019.
- [6] H. Rehioui and A. Idrissi, "New clustering algorithms for twitter sentiment analysis," *IEEE Systems Journal*, vol. 14, no. 1, pp. 530–537, 2020.
- [7] M. Charytanowicz, K. Perzanowski, M. Januszczak, A. Wołoszyn-Gałęza, and P. Kulczycki, "Application of complete gradient clustering algorithm for analysis of wildlife spatial distribution," *Ecological Indicators*, vol. 113, Article ID 106216, 2020.
- [8] J. Jokinen, T. Raty, and T. Lintonen, "Clustering structure analysis in time-series data with density-based clusterability measure," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1332–1343, 2019.
- [9] P. Charoensukmongkol and P. Tarsakoo, "Dimensions of social media marketing capabilities and their contribution to business performance of firms in Thailand," *Journal of Asia Business Studies*, vol. 14, no. 4, pp. 441–461, 2019.
- [10] S. Venaik and D. F. Midgley, "Archetypes of marketing mix standardization-adaptation in MNC subsidiaries," *European Journal of Marketing*, vol. 53, no. 2, pp. 366–399, 2019.
- [11] M. D. Mooij, "Fairy tales of global consumer culture in a polarizing world," *International Marketing Review*, vol. 36, no. 4, pp. 581–586, 2019.
- [12] M. Chen, S. Lu, and Q. Liu, "Uniform regularity for a keller-segel-navier-stokes system," *Applied Mathematics Letters*, vol. 107, Article ID 106476, 2020.
- [13] D. Hristovski, J. Stare, B. Peterlin, and S. Dzeroski, "Supporting discovery in medicine by association rule mining in medline and UMLS," *Studies in Health Technology & Informatics*, vol. 84, no. Pt 2, pp. 1344–1348, 2001.
- [14] D. Kunkle, D. Zhang, and G. Cooperman, "Mining frequent generalized itemsets and generalized association rules without redundancy," *Journal of Computer Science and Technology*, vol. 23, no. 1, pp. 77–102, 2008.
- [15] S.-L. Wang, D. Patel, A. Jafari, and T.-P. Hong, "Hiding collaborative recommendation association rules," *Applied Intelligence*, vol. 27, no. 1, pp. 67–77, 2007.
- [16] F. Jiang, K. K. R. Yuen, and E. W. M. Lee, "Analysis of motorcycle accidents using association rule mining-based framework with parameter optimization and GIS technology," *Journal of Safety Research*, vol. 75, pp. 292–309, 2020.
- [17] M.-C. Chen, S.-Q. Lu, and Q.-L. Liu, "Global regularity for a 2D model of electro-kinetic fluid in a bounded domain," *Acta Mathematicae Applicatae Sinica, English Series*, vol. 34, no. 2, pp. 398–403, 2018.
- [18] W. Zhang, Y. Liu, Z. Liu, S. Wang, Z. Zhen, and Y. Li, "Research on fault monitoring technology of distribution network based on fuzzy association rules mining," *Journal of Physics: Conference Series*, vol. 1626, no. 1, p. 9, Article ID 012072, 2020.
- [19] D. Dahmani, S. A. Rahal, and G. Belalem, "A new approach to improve association rules for big data in cloud environment," *The International Arab Journal of Information Technology*, vol. 16, no. 6, pp. 1013–1020, 2019.
- [20] C.-H. Wang, "Association rule mining and cognitive pairwise rating based portfolio analysis for product family design," *Journal of Intelligent Manufacturing*, vol. 30, no. 4, pp. 1911–1922, 2019.
- [21] T. A. Rather, S. Kumar, and J. A. Khan, "Multi-scale habitat modelling and predicting change in the distribution of tiger and leopard using random forest algorithm," *Scientific Reports*, vol. 10, no. 1, Article ID 11473, 2020.
- [22] Z. He, S. Zhang, and J. Wu, "Significance-based discriminative sequential pattern mining," *Expert Systems with Applications*, vol. 122, pp. 54–64, 2019.
- [23] M. J. Alibasa, R. A. Calvo and K. Yacef, "Sequential pattern mining suggests wellbeing supportive behaviors," *IEEE Access*, vol. 7, pp. 130133–130143, 2019.
- [24] Y. Liu, S. Huang, S. Mei, and X. Zhang, "Analysis on pattern of power system cascading failure based on sequential pattern mining," *Dianli Xitong Zidonghua/Automation of Electric Power Systems*, vol. 43, no. 6, pp. 34–40, 2019.
- [25] B. P. Jaysawal and J. W. Huang, "Psp-AMS: progressive mining of sequential patterns across multiple streams," *ACM Transactions on Knowledge Discovery from Data*, vol. 13, no. 1, Article ID 5, 2019.