

Research Article

Algorithm of Apriori-Based Rural Tourism Driving Factors and Its System Optimization

Hongyi Xiao 

Henan University of Animal Husbandry and Economy, Zhengzhou 450000, China

Correspondence should be addressed to Hongyi Xiao; 81448@hnuah.edu.cn

Received 5 May 2022; Revised 10 June 2022; Accepted 13 June 2022; Published 1 August 2022

Academic Editor: Liping Zhang

Copyright © 2022 Hongyi Xiao. 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.

Rural tourism, as a vital component of tourism, is critical to the development of rural economies, farmers' income, rural civilization, new rural construction, and urban-rural interaction. Simultaneously, as the size and complexity of data sets grow larger, how to improve the efficiency of association rule algorithms for mining large data sets has become a hot topic in association rule mining. Rural tourism development that is cultural and creative not only contributes to rural revitalization, but also to the preservation and inheritance of rural culture. The Apriori algorithm is the most widely used and influential algorithm for mining Boolean association rules, and the majority of current algorithms are extensions of the Apriori algorithm. Demand, supply, marketing, and support forces of rural tourism, which are the core driving force of rural tourism development, are formed by the basic needs of each subsystem of rural tourism. One of the main methods is to promote the sustainable and healthy development of rural tourism in accordance with the nature, characteristics, and laws of rural tourism destination construction, in order to create a dynamic system for long-term development and establish a rural tourism development dynamic system. The study of rural tourism driving factors and their system optimization is proposed in this paper. The main tourism dynamic system is adopted by the association rule algorithm of Apriori, the driving factors of rural tourism development are analyzed in the paper, and the system optimization method is proposed, all based on the Apriori algorithm. In terms of support, the Apriori algorithm is 0.436 higher than the CD algorithm and 0.568 higher than the SVM algorithm, and the Apriori algorithm can greatly reduce database size and improve record reading speed. As a result, the findings of this paper can be used to improve the spatial layout of rural tourism and to develop urban-rural tourism.

1. Introduction

Rural tourism has a lot of resources, and it broadens the concept of tourism resources by turning a lot of villages, woods, pastures, and fishing grounds into tourist destinations [1]. It is a new type of industry that is based on agriculture, tourism as a means of transportation, urban residents as the primary target, and a combination of primary and tertiary industries [2]. Rural tourism, as a relatively new form of tourism, is rapidly expanding at a rate of more than 17% per year [3]. Many countries and regions regard it as an effective way to prevent agricultural decline and increase rural income, and it is a key driver of rural economic reform and development [4]. Analysing the dynamic mechanism of rural tourism development in China from a system dynamics perspective and dealing with the

relationship between accelerating rural development and adhering to the laws of rural tourism will undoubtedly improve the efficiency of rural tourism in China. For the positive role of the new countryside, a dynamic system to improve the efficiency of rural tourism in the construction of socialism tourism is very important [5].

With the rise in popularity and development of the Internet, as well as the emergence of e-commerce websites, online information searches have become a popular way for tourists to gather information prior to travelling [6]. Tourists, on the other hand, are frequently caught in a tangle of information searches and product choices as a result of severe information overload [7]. This massive amount of data is frequently stored across multiple devices, and the next step is not massive knowledge, but massive storage pressure. The relationship between rural tourism and new

urbanisation, rural cultural construction, precise poverty alleviation, rural ecological construction, and rural culture has become an important topic in today's research field, and this demand has promoted the act of developing effective tools for analysing the data [8] and mining information [9].

Local governments are using rural tourism as an effective means to revitalise the local economy and increase farmer income, thanks to the central government's policy guidance and strong promotion [10]. As a result, China's rural tourism industry is booming [11]. Rural tourism development is a complex phenomenon that necessitates the coordination and interaction of various aspects of tourism development, such as ensuring sustained demand and maintaining viable tourist attractions [12]. It also necessitates a favourable development environment and proactive policy guidance [13]. In terms of tourism spatial structure analysis for tourism system optimization, it is necessary to establish a regional spatial function system and identify tourism centres based on the comparative advantages of each district and country, in addition to focusing on the elements of tourism spatial structure. Define the spatial structure of tourism at a high level [14]. Simultaneously, it is critical to expand tourism area construction and compensate for the shortcomings of tourism transportation resourcefulness and product transformation research.

The innovations of this paper are as follows:

- (1) The proposed measures for improving and optimizing the dynamical system of rural tourism development are of reference value for decision making in promoting rural tourism development.
- (2) This paper uses the system dynamics analysis method to construct the dynamic system of rural tourism development, analyzes the current problems of rural tourism and the essential causes of relevant contradictions within the institutional framework, and proposes the system optimization strategies.
- (3) On the basis of determining the prediction index system of the dynamic system of rural tourism development, the algorithm of Apriori is used to optimize the dynamic system of rural tourism development, and the generation of frequent term sets is constructed and analysed in the algorithm of Apriori.

2. Related Work

2.1. Driving Factors of Rural Tourism and Its System Optimization. With the increasing role of rural tourism in rural economic development and the increase of farmers' income, the development of rural tourism in various regions has shown a boom. The combination of agriculture and tourism will promote the development of rural tourism industry, thus further promoting the development of new rural construction, narrowing the gap between urban and rural areas and accelerating the process of urban-rural integration. At present, significant progress has been made at home and abroad in research on the concept, influencing

factors, benefit analysis, and operation management of rural tourism.

Yue and Rong found in their study of community participation in rural tourism that the participation of community residents in the planning and decision making of rural tourism is a prerequisite for its sustainability if a portion of the economic effects can be obtained [15]. Yu used the Acadian region in eastern Canada as an empirical study to argue that the main drivers of rural cultural tourism differ in their development stages [16]. Singh et al. studied the joint brand hypothesis of rural tourism and used the western region of the United States as an example to argue that the joint brand effect can make the local area stronger in competitive tourism [17]. Wang, studying the Bidayuh community in Malaysia, found that the locals in the community wanted to participate in tourism activities by showing and sharing the customs and traditions of the community, but the government did not take effective measures to increase the importance of the Bidayuh community [18]. Based on the spatial scale and travel process involved in tourism transportation, Nilashi divided it into three levels: external transportation, transportation from urban tourism centers to scenic spots, and internal transportation. He argued that high quality is a characteristic of modern tourism transportation and is a decisive factor in the competitiveness of tourism [19].

The study of rural tourism in China started late, and nowadays, rural development models such as B&B, retirement resort, and agricultural leisure are becoming more and more mature under the huge political and socio-economic background. Therefore, the actual and potential consumption demand for rural tourism is very strong, which fits the consumption psychology of urban residents to return to nature, and also helps farmers to open their eyes, update their concepts, and change their lifestyles as well as living conditions.

2.2. Algorithm of Apriori. Systems theory holds that a system is an organic whole with specific functions composed of several interacting and interdependent components with mobility of matter, energy, and information. However, some contemporary rural areas have unique industrial structures and declining smallholder economies. Young people prefer to leave the rural land, leaving many dilapidated buildings, abandoned communities, and no commercial development, and the road to rural revitalization is a long one. In addition, there is an explosion of various data accumulated by people, and how to effectively use these data to improve production and people's lives is a major challenge at present.

Liu et al. first proposed the problem of association rule mining between sets of transaction items in customer transaction databases, and since then, many researchers have conducted a lot of research on association rule mining [20]. Wang and Cao generated candidate itemsets by encoding transaction records into bit tables, and then performed a dissimilarity operation on frequent itemsets and calculated the number of 1s in the dissimilarity result [21]. Xie proposed a dynamic hash pruning algorithm that not only

eliminates unsatisfied candidate sets in each iteration, but also eliminates records in the database that do not contain frequent sets, which can reduce the size of the database and the time required to scan the time required to scan the database [22]. Wang and Gao partitioned the transaction records based on 1-itemsets and calculated the support of candidate 1-itemsets by selecting the item with the lowest support in the 1-itemset, the candidate element set, and then dividing the support of the candidate element set of the transaction records that traverse the element based on the previous analysis of the 1-item frequent set [23]. Silva et al. proposed a partitioning algorithm that logically divides the horizontal database into several non-repeating modules and then reads each module with the IDs of the records in the database as a set of elements, and then intersects these ID element sets to produce a local frequent element set [24].

Algorithm of Apriori is the most classic and commonly used algorithm in data mining techniques. Association analysis still has considerable demand for application in commercial marketing data analysis, while fuzzy clustering algorithm is still inhibited in the field of massive data mining due to high resource overhead.

3. Driving Factors of Rural Tourism and System Optimization Ideas Based on Algorithm of Apriori

3.1. Driving Factors of Cultural and Creative Rural Tourism Development. The demand and supply mechanisms that drive rural tourism development can be summarised as follows [25]. For example, in the early stages of rural tourism, development is slow and only a few tourists visit, so this stage is more about local residents' commercial behaviour. The driving factors of cultural and creative rural tourism development are divided into three aspects: demand-driven, resource-driven, and economic-driven, based on the characteristics of cultural and creative rural tourism development and the political orientation of the country's integrated tourism development and regional economy [26]. Many people and organisations collaborate to help guide the long-term preservation of cultural tourism in the countryside, ensuring its long-term viability. A structural model is built based on the arrangement of cultural and creative tourism development drivers and the influencing factors of each driver. This is shown in Figure 1.

First of all, it is demand-driven. Contemporary rural areas have their own urgent needs for development, and the reform of the agricultural industry structure and the construction of new rural areas have become inescapable factors in the development of the cultural and creative rural environment. The algorithm of Apriori uses more support, which represents the proportion of transactions in the transaction log that contain this itemset. The mathematical formula is

$$S(X) = \frac{|\{t_i | X \subseteq t_i, t_i \in T\}|}{N}, \quad (1)$$

N —Number of transaction records.

The concept of rural tourism resource should be any material and immaterial resource in the domain that is attractive to rural tourists and capable of bringing economic, social, and environmental benefits to the domain. In order to recommend accurate, diverse, and personalized tourism products to users, four areas of data information need to be conveyed: user needs, user preferences, restrictions, and tourism resources database. Given the weight of itemset $X = \{i_1, i_2, \dots, i_k\}$ is w_j , where $0 \leq w_j \leq 1, j = \{1, 2 \dots k\}$. The weighted support of itemset X is

$$WSup(X) = W(X) \times \frac{Sup(X)}{T}, \quad (2)$$

$W(X)$ — $\text{Max}\{w_1, w_2 \dots w_k\}$;

T —Total number of transactions in the database.

Under the perspective of whole area tourism, we deeply analyze the connotation of concepts related to tourism spatial structure, identify and analyze the elements and forms of municipal tourism spatial structure, and further guide the optimization of municipal tourism spatial structure. Given a transaction database, the number of support and the minimum support should satisfy if the t attribute set X is frequent if the user input has a minimum support of $w \min Sup$, respectively,

$$SC(X) \geq \frac{w \min Sup \times T}{MAX(w_i)}, \quad (3)$$

$$B(X) = \frac{w \min Sup \times T}{MAX(w_i)}. \quad (4)$$

When there are two m -dimensional vectors $X_1 = (x_{11}, x_{12}, \dots, x_{1n})$ and $X_2 = (x_{21}, x_{22}, \dots, x_{2n})$, the Euclidean distance between them is

$$\text{dist}(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2}. \quad (5)$$

It combines various resources from various industries to enhance product features, highlight regional cultural characteristics, and increase rural tourism's competitiveness, all on the premise of gathering its own rural tourism human resources. Rural tourism resources are diverse, encompassing not only resources related to farmer production and life, but also natural tourism resources such as terrestrial human landscapes, water features, and biological landscapes throughout the country [27]. Cultural tourism resources include leisure, knowledge, fitness, and shopping, as well as rural environmental areas and relics and architecture. As a result, demand drivers should be classified as objective or subjective, with objective demand being further subdivided into urban pollution and rural development [28]. The demand of urban residents and rural residents is divided into two categories. Natural and human conditions are separated in terms of resource demand [29]. Agricultural and tourism integration, on the other hand, is a key feature of rural tourism. The majority of farmers farm and travel within their

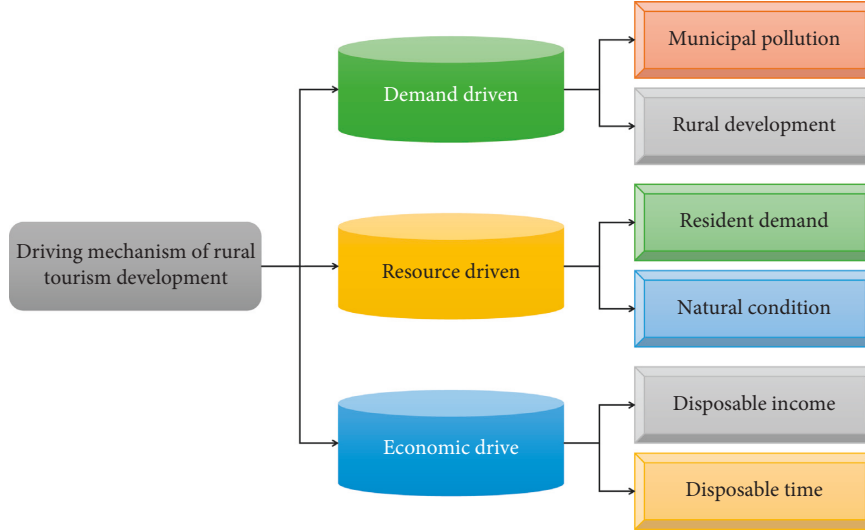


FIGURE 1: Structure model of influencing factors of driving mechanism of rural tourism development.

own communities. They have the ability to convert their living and productive assets to commercial assets. With small investment, low risk, flexible operation, correct planting time, and obvious localization, the field is very suitable for farmers to operate and is one of the best ways for farmers to get out of poverty and realize their dream of modernization. One property of k -itemset is the number of transaction records containing this k -itemset in the transaction record, called the number of supported transactions. The mathematical formula is

$$\sigma(X) = |\{t_i | X \subseteq t_i, t_i \in T\}|. \quad (6)$$

Finally, strong productivity and increased levels of modernization, driven by the economy, have significantly reduced the working hours of modern workers, leaving more time available for workers and providing the basis for the development of a tourist countryside. The countryside and investment are the two core factors that form the driving force behind the demand for rural tourism. Guided by the thinking concept of spatial planning and the basic theory of spatial design, the spatial level of the macro level of rural tourism is clarified with the help of rural tourism resource points, rural tourism product lines, and rural tourism zoning. Among the many driving factors, the construction and improvement of infrastructure becomes a fundamental factor. Roads, water, electricity, and other infrastructure construction improve the accessibility of rural areas, the convenience of tourism operations, so that the rural natural resources that are more accessible are transformed into tourism resources, and farmers, to develop tourism, participate in providing the most basic support forces.

3.2. Optimization of Dynamic System Based on Algorithm of Apriori. Tourism recommendations include six aspects of tourism products: tourism, entertainment, shopping, food, accommodation, and travel. However, the optimization of the rural tourism system should not be developed blindly and hastily. Rather, it should be deep in the tourism market,

taking into account the location of rural transportation, resource sharing, social and economic development, etc., to determine the direction of each region's tourism development. However, limited by time and consumer spending and other restrictions, the actual number of scenic spots visited by users in a year is only a handful, and the number of scenic spots where users travel together is also very small, so tourism data is relatively scarce. Therefore, a priori association rule algorithm is used. The flow chart of algorithm of Apriori is shown in Figure 2.

The first is staging. Scenic operations in tourism development are about bringing a part of the life of local residents to the foreground to showcase them, basically taking a closed approach to the backstage. The premise of determining the type of rural tourism destination is to find the influencing factors that can correctly reflect the characteristics of rural tourism destinations. For the evaluation of rural tourism places, most of the factors of tourism resources, tourism infrastructure conditions, location and transportation conditions, and external support conditions are selected and numerous evaluation factors are listed. We use an n -dimensional random variable to represent X with a weighting factor of A_i , i.e., the weighted sum of n basis vectors to represent X .

$$X = \sum_{i=1}^n \partial_i \phi_i. \quad (7)$$

Since users' needs, product preferences, and constraints may be reflected in the travel information that users see, a text classification method is used to extract frequent sets of historical keywords searched and viewed by users to generate a user preference dictionary. The absolute value function of the factor coefficients in the regression equation is added to the model as a penalty term to make some regression coefficients smaller. By regression, coefficients whose absolute values are not sufficient to explain the dependent variable can be changed directly to 0. The expression of the LASSO method can be written as:

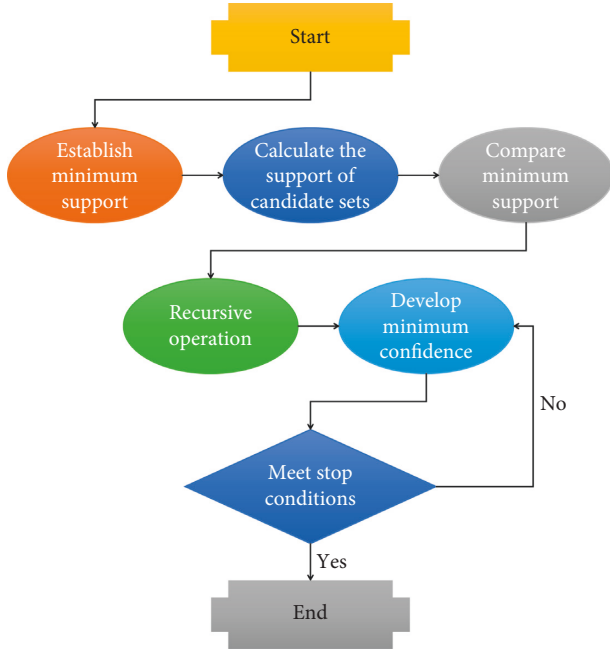


FIGURE 2: Algorithm of Apriori flow chart.

$$\bar{\beta} = \arg \min \left\{ \|Y - X\beta\|^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}. \quad (8)$$

The main idea is to divide the instances in the database into corresponding subsets based on the difference of attribute values. Then, the classification rule attributes are generated based on the approximate up-down relationship between the subset divided by the conditional attributes and the subset divided by the decision. The algorithm of Apriori-based tourism dynamic system optimization is shown in Figure 3.

The second is profit-oriented, through scientific planning and strong guidance, so that most farmers can find a combination of their own interests between modernization and the countryside, making them aware that the countryside is the core attraction of rural tourism and the origin of valuable resources, committed to rural tourism, so that the rural characteristics can be inherited and optimized. The basic theory of graph-based ranking algorithm is to establish the connection between words with semantic relations, construct a model of Text Rank, and recursively calculate the value of words based on their mutual “votes”. The value of each word depends on its word voting score and the size of the voting word itself, and the word with the larger value is considered to be the important word. This also indicates that the solution to the linear programming problem is an integer solution. Converting the diversity constraint to an upper and lower bound constraint for classification, the optimization problem is as follows:

$$\max \sum_{i,j} w_{ij} x_{ij}. \quad (9)$$

w_{ij} — ij preference scoring matrix;

x_{ij} —Final match.

The database of things is scanned and the items are counted, and then the items that satisfy \min - Support are filtered out so that the set of frequent 1-itemsets can be obtained. Thus, the feature vectors are used to describe the relevant data in the database that are more similar. The variation parameters of the different data attribute features can be calculated using the following equation.

$$e(y, z) = \left[\sum_{j=1}^p |y_j - z_j|^s \right]^{1/s}. \quad (10)$$

y —The amount of data in the database;

z —Number of data attributes;

s —Differences in data attribute characteristics.

Finally, organization. Through the establishment of organisations such as rural tourism associations and rural tourism cooperatives, the scattered disadvantaged groups are organized to provide support and guidance for the disadvantaged groups to cultivate development opportunities and capabilities through scientific planning and reasonable institutional arrangements. Also, when comparing and analysing the factor data of different dimensions, they need to be standardized. The value of i factor j in all samples is:

$$Z_{i,j} = \frac{X_{i,j} - \mu_j}{\sigma_j}. \quad (11)$$

μ_j —The mean value of factor i ;

σ_j —The standard deviation of factor j .

Then, based on the user’s current search keywords, the algorithm of Apriori is used to correlate and extract the user’s trend dictionary, thus recommending travel information that matches the user’s current interest trends. In this stage, those candidates that are frequent sets in each module are eliminated because they are already global frequent sets, and those candidates that cannot be frequent sets are also eliminated. Then, a new mean value is calculated for each cluster and the process is repeated until the criterion function converges. Using the squared error criterion, which is defined as follows:

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2. \quad (12)$$

E —Sum of squared errors of all objects in the dataset

p —Points in space

i —Given object

m_i —Mean value of cluster C_i .

The transaction log is then traversed and the number of supported transactions from candidate sibling itemset nodes is calculated. For a transaction log, if it does not contain a prefix tree node, then the transaction log must not contain

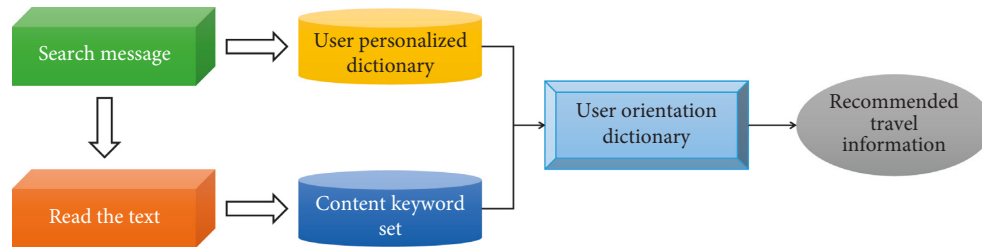


FIGURE 3: Optimization of tourism power system based on algorithm of Apriori.

the set of candidate elements represented by all nodes in that node tree to improve the efficiency of statistical support for the prefix set.

4. Application and Analysis of Algorithm of Apriori in Optimization of Rural Tourism Power System

4.1. Generation and Analysis of Frequent Itemsets in Algorithm of Apriori. In parallel environments, we often encounter the problem that before a task starts, we try to distribute the task equally to each processor, hoping that each processor will work at full capacity. Therefore, the focus of mining association rules is on the first step, generating all the frequent sets. The basic idea is to start with itemsets, and the generated candidate itemsets are judged by the minimum support of the generated frequent itemsets. Continue to combine complex item frequent sets to generate item candidate sets, continue to judge the minimum support of the generated frequent itemsets, and so on, until the maximum item frequent set is found. In order to verify the effectiveness of the algorithm of Apriori, the support of the algorithm of Apriori and the SVM algorithm were implemented on a computer running the Windows 11 operating system, and comparative experiments were conducted with different numbers of things. The results are shown in Figures 4 and 5.

First, the support of the itemset in the candidate itemset is calculated, i.e., compared with each transaction in the transaction database. If a transaction in the transaction database contains a candidate itemset, the candidate itemset holder is added by 1. The support of an itemset is the frequency of the itemset, i.e., the number of records in the entire dataset that contain the itemset. If the frequency of an itemset is greater than some frequency threshold, it becomes a frequent itemset. However, as each processor runs for a period of time, it is often the case that some processors finish quickly while others finish slowly. These processors become bottlenecks and affect the overall task completion. Therefore, in practice, the content of the database is generated from the transaction data of the e-commerce website, and then based on these data, the aforementioned data mining techniques are used to generate frequent itemsets, which in turn extract knowledge and eventually become the recommendation routes recommended to users. The pruning effect would be better if the granularity of the database is finer, but of course within a suitable range. The chunking support is shown in Table 1.

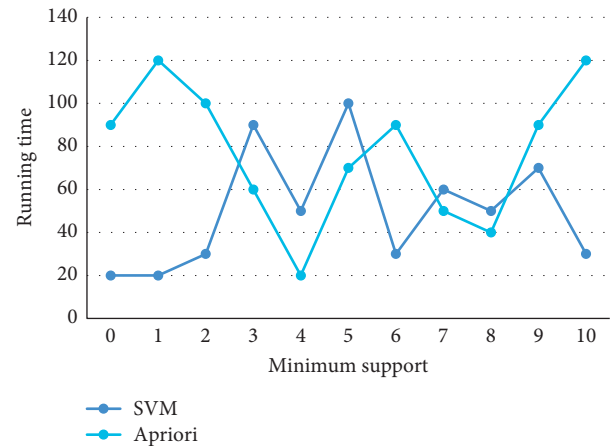


FIGURE 4: Comparison of support between algorithm of Apriori and SVM algorithm.

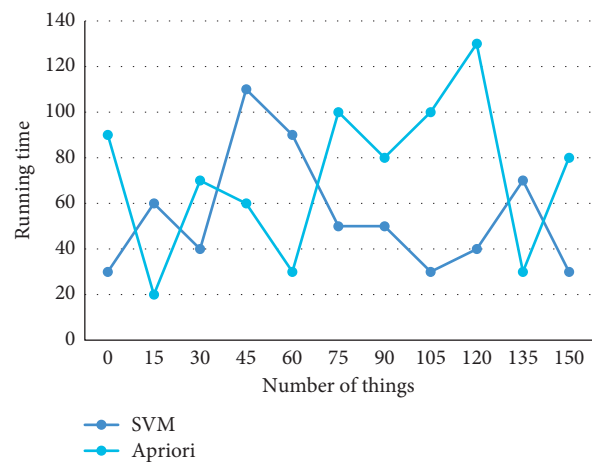


FIGURE 5: Comparison between algorithm of Apriori and SVM algorithm under different things.

Second, by adding partitioning techniques to the association rule method, the efficiency of data reading is improved, which helps to improve the efficiency of association analysis operations and also provides the possibility of parallel processing. The extraction of frequent itemsets is done by iterative level-by-level search, adding the transaction log attributes to each node in the sorted tree, thus reducing the number of transaction logs that must be traversed when calculating the support of candidate element

TABLE 1: Block support.

Itemset	Block 1 support	Block 2 support	Global support
A	36	28	64
B	41	37	68
C	55	45	100

sets. It is obtained by extracting and loading the data from the original database and processing it using a data cleaning tool. The next task is to extract association rules using the above mining algorithm and store the generated mining results (knowledge) in the knowledge base. For the user (tourist), a reasonable and cheap travel itinerary is available without multiple comparisons and tedious queries.

Finally, the frequent itemset is the set of all itemsets with support greater than the minimum support in the candidate itemset, but it is not enough to have support, but also to set a standard set of elements for their confidence. A useful association rule must satisfy two conditions: support and confidence. In the process of counting candidate itemsets, binary-coded itemsets and transaction records are used, and operations between binary sets are used instead of operations between sets. That is, the k -itemset probing itemset $(k + 1)$ -is used to exhaust all frequent itemsets in the dataset until no more frequent itemset k -can be found. With the development of rural tourism, the rural economy participates in the market economy, improving the competitiveness of the countryside and gaining more advanced cultural concepts and ideas that will give back to the development of the new countryside. Therefore, when customers visit a tourism website, their visit traces are recorded and these records are very useful, and these records can be used as the original database for the data mining module to provide data sources for data mining.

4.2. Load Balance Optimization Analysis. The focus of this section is to investigate how to dynamically allocate data so that each processor balances the workload as much as possible throughout the run and maximizes parallelization.

First, the database is virtually partitioned using k -clustering techniques in conjunction with the requirements of virtual partition removal techniques to make the database partitioning as skewed as possible. Association rules are not only based on the support and trust provided to simply find the relevant rules, we can also extend them. The database is scanned and the D transaction matrix is further constructed based on the definition, and the behavioral things are enumerated as the set of elements during the construction process. In terms of space complexity, the improved algorithm processes the data from scratch and assigns it to the Boolean array data table, so the data in the source database is represented by 1s and 0s and takes up more space when stored than the original data storage. Direct domino chain reaction of parallel nodes, if the task of one node fails, the whole task cannot be completed. The method is implemented on the basis of sampling methods and will play an active role in data mining tasks for classical association algorithms. Optimization experiments for

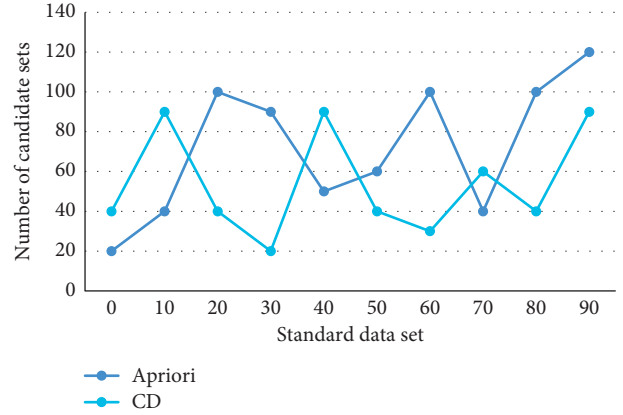


FIGURE 6: Algorithm of Apriori, CD algorithm.

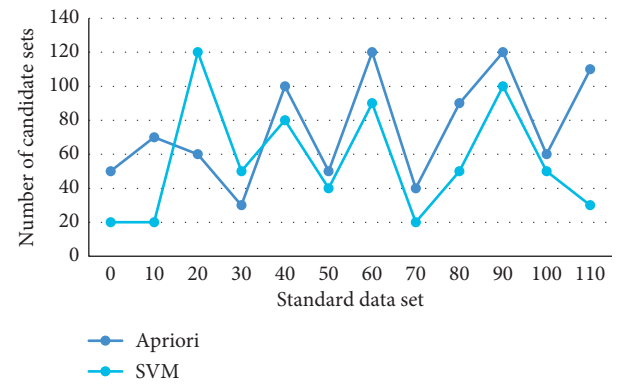


FIGURE 7: Comparison between algorithm of Apriori and SVM algorithm.

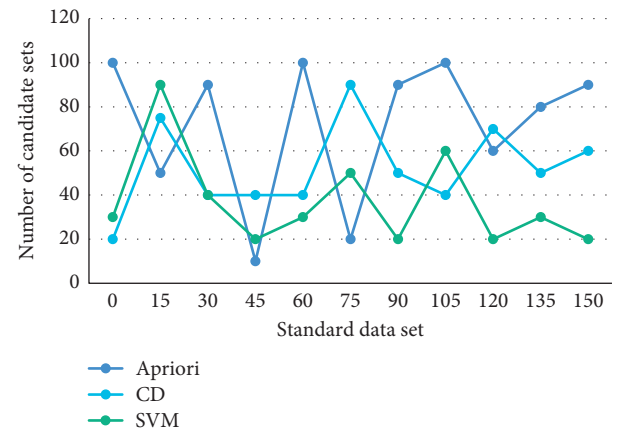


FIGURE 8: Comparison of algorithm of Apriori, CD algorithm, and SVM algorithm.

candidate set generation are performed using simulated generated standard datasets. The standard datasets were tested in algorithm of Apriori, CD algorithm, and SVM algorithm. The experimental results of candidate set generation are shown in Figures 6–8.

Next, the database is divided into multiple intervals of the same size. Assuming that the number of one-element sets

TABLE 2: Database after statistics.

Dimension	Support 1	Support 2	Total
Apriori	0.557	0.672	1.229
CD	0.415	0.378	0.793
SVM	0.372	0.289	0.661

in the database is m , when counting the 1-itemsets, the support vector of 1-itemsets for each interval can be obtained. Then, a two-dimensional array is input, and the combination of subscripts and the frequency of the combination are stored in this array to facilitate finding the frequent element sets. If the number of tasks in the cluster is particularly large or small, it still causes the problem of uneven load. So, we need to ensure homogeneity not only between control modules, but also within modules. Then, traverse the transaction log and insert all possible subsets contained in the transaction log into the binomial tree, as the transaction log should increase the number of transactions supported by the set of items in all subsets of the log by one transaction. The items in each transaction in the transaction database operate in descending order according to the established support of the generated one-dimensional target items, and each transaction requires a new link at the root node. The support of each element set of the algorithm of Apriori, CD algorithm, and SVM algorithm can be obtained by counting the support of all 1 element sets, respectively, as shown in Table 2.

Finally, assuming that each module can obtain a similar number of intervals from each group, the distribution of element sets within each module should also be fairly uniform, usually better than the random partitioning. This includes processes such as candidate itemset generation, transaction log matching, and candidate itemset support counting. However, as the mining database grows, it takes a long time to generate the candidate itemsets and calculate their support at each iteration when applying the algorithm of Apriori. Therefore, when a 4-item candidate C_4 set is generated, the algorithm eventually terminates because no new frequent itemsets can be found. Among them, the values of attributes are categorical data processing, which generally uses Boolean association rules because it can show the relationship between different attributes. For numerical attribute processing, it usually uses numerical association rules. Only one scan of the database is required to generate a clustering matrix to replace the original transactional database. When generating frequent itemsets, only part of the grouped arrays need to be operated accordingly, which can greatly improve the operational efficiency of the algorithm.

5. Conclusion

The terms “rural tourism” and “sustainable development” refer to tourism activities that take place in rural areas, with “rural” and “sustainable development” as criteria, and rural natural and cultural tourism resources as the draw. The current driving system of China’s tourism development can be divided into four subsystems based on a comprehensive analysis of rural tourism development: gravitational system,

demand system, intermediary system, and support system. Rural tourism development can be broken down into three categories: demand-driven, resource-driven, and economic-driven. The algorithm of Apriori-based method of rural tourism driving factors and its system optimization is proposed for the special problems faced by dynamical systems in the field of tourism. The frequent itemset can be generated by calculating the partial clustering matrix using this algorithm, which only requires scanning the database once and generating a series of different clustering matrices. Meanwhile, the Apriori algorithm’s generation of frequent element sets is investigated from the standpoint of dynamic interaction of factors, and a load balancing optimization analysis is performed. The system is able to extract information related to the tourist attractions preferred by consumers and generate feedback information that is instructive to the tourism management system after extracting, comparing, and analysing the data in the tourism management system using the Apriori algorithm. As a result, the improved Apriori algorithm improves search efficiency while reducing the workload. Rural tourism can perform a wide range of functions, including revitalising the economy, coordinating society, improving the environment, and promoting urban-rural integration, using the rural tourism driving force factor based on the Apriori algorithm and its system optimization method. Furthermore, in a positive interaction, it can effectively promote the development of new socialist countryside construction and rural tourism.

Data Availability

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

Conflicts of Interest

The author does not have any possible conflicts of interest.

Acknowledgments

This study was supported by Young Backbone Teachers Project of Colleges and Universities in Henan Province, Research on the integration of culture and tourism driving the high-quality development of rural tourism in Henan Province, Grant No. 2019GGJS260.

References

- [1] X. Ma, Z. Yang, and J. Zheng, “Analysis of spatial patterns and driving factors of provincial tourism demand in China,” *Scientific Reports*, vol. 12, no. 1, 2260 pages, 2022.
- [2] L. Chen, T. Brijesh, K. Jinwon, and L. Yi, “Landscape optimization in a highly urbanized tourism destination: an integrated approach in Nanjing, China,” *Sustainability*, vol. 9, no. 12, pp. 1–20, 2017.
- [3] Y. Kim, S. Park, Y. Kwon, S. Lee, W. Choi, and C. Lee, “Multi-city flight route optimization system using big-data analytics,” *Journal of the Korean Institute of Industrial Engineers*, vol. 44, no. 2, pp. 92–101, 2018.

- [4] J. Ma, H. Wu, X. Zhang, Z. Huijuan, and C. Yufei, "Research on customized passenger route optimization based on improved genetic algorithms," *IOP Conference Series: Earth and Environmental Science*, vol. 638, no. 1, Article ID 012011, 2021.
- [5] L. Hulianytskyi and A. Pavlenko, "Ant colony optimization algorithms with diversified search in the problem of optimization of airtravel itinerary," *Cybernetics and Systems Analysis*, vol. 55, no. 6, pp. 978–987, 2019.
- [6] J. M. Su, F. Zhang, and W. P. Jing, "KM-A algorithm based on clustering mixed Apriori for the study of ZROM model," *Journal of Computer Science*, vol. 28, no. 2, pp. 165–175, 2017.
- [7] L. Zhu, "Implementation of web log mining device under Apriori algorithm improvement and confidence formula optimization," *International Journal of Information Technology and Web Engineering*, vol. 15, no. 4, pp. 53–71, 2020.
- [8] E. Q. Wu, D. Hu, P. Y. Deng et al., "Nonparametric Bayesian prior inducing deep network for automatic detection of cognitive status," *IEEE Transactions on Cybernetics*, vol. 51, no. 11, pp. 5483–5496, 2021.
- [9] E. Q. Wu, C. T. Lin, L. M. Zhu, Z. R. Tang, Y. W. Jie, and G. R. Zhou, "Fatigue detection of pilots' brain through brains cognitive map and multilayer latent incremental learning model," *IEEE Transactions on Cybernetics*, vol. 1, pp. 1–13, 2021.
- [10] L. Zhou, "Analysis of factors that tourists attach importance to rural tourism image—taking rural tourism in Ningde region as example," *Journal of Contemporary Educational Research*, vol. 4, no. 9, 7 pages, 2020.
- [11] C. Ren, T. Cui, and X. Yang, "Study on the development of rural tourism in Liaoning province," *Asian agricultural research: English version*, vol. 9, no. 4, pp. 51–53, 2017.
- [12] S. Cornelissen, "The global tourism system: governance, development and lessons from South Africa," *Tourism Management*, vol. 28, no. 2, 650 pages, 2017.
- [13] H. R. Wang, Y. C. Luo, and H. L. Shang, "A study of the spatiotemporal evolution and the driving factors of A-level tourism attractions in Qiandongnan Prefecture, China," *Ecological Economy: English version*, vol. 17, no. 2, 11 pages, 2021.
- [14] Y. Shi, C. Shao, and Z. Zhang, "Efficiency and driving factors of green development of tourist cities based on ecological footprint," *Sustainability*, vol. 12, no. 20, 8589 pages, 2020.
- [15] P. Yue and P. Rong, "Spatial patterns and driving factors of urban residential embedded carbon emissions: an empirical study in Kaifeng, China," *Energy and Power Engineering*, vol. 11, no. 02, pp. 58–75, 2019.
- [16] H. Yu, "Apriori algorithm optimization based on Spark platform under big data," *Microprocessors and Microsystems*, vol. 80, no. 11, Article ID 103528, 2021.
- [17] S. Singh, R. Garg, and P. K. Mishra, "Performance Optimization of MapReduce-Based Apriori Algorithm on Hadoop Cluster," *Computers & Electrical Engineering*, vol. 67, pp. 348–364, 2018.
- [18] M. Wang, "Applying Internet information technology combined with deep learning to tourism collaborative recommendation system," *PLoS One*, vol. 15, no. 12, Article ID e0240656, 2020.
- [19] M. Nilashi, K. Bagherifard, M. Rahmani, and V. Rafe, "A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques," *Computers & Industrial Engineering*, vol. 109, no. jul, pp. 357–368, 2017.
- [20] X. Liu, Y. Zhao, and M. Sun, "An improved Apriori algorithm based on an evolution-communication tissue-like P system with promoters and inhibitors," *Discrete Dynamics in Nature and Society*, vol. 2017, Article ID 6978146, 2017.
- [21] Z. Wang and L. Cao, "Novel Apriori-based multi-label learning algorithm by exploiting coupled label relationship," *Journal of Beijing University of Technology*, vol. 26, no. 002, pp. 206–214, 2017.
- [22] H. Xie, "Research and case analysis of Apriori algorithm based on mining frequent item-sets," *Open Journal of Social Sciences*, vol. 09, no. 04, pp. 458–468, 2021.
- [23] H. B. Wang and Y. J. Gao, "Research on parallelization of Apriori algorithm in association rule mining," *Procedia Computer Science*, vol. 183, pp. 641–647, 2021.
- [24] J. Silva, N. Varela, L. A. Borrero López, and R. H. Rojas Millan, "Association rules extraction for customer segmentation in the SMEs sector using the Apriori algorithm," *Procedia Computer Science*, vol. 151, pp. 1207–1212, 2019.
- [25] O. Liansyah and H. Destiana, "The use of Apriori algorithm in the formation of association rule at lotteria cibubur," *SinkrOn*, vol. 4, no. 2, 76 pages, 2020.
- [26] L. J. Huang, G. Xie, W. Zhao, Z. Dou, and Y. Wang, "Empirical analysis for e-integrated logistics pattern between urban and rural area: from economic and geographic perspective," *Boletín Técnico*, vol. 55, no. 1, pp. 65–76, 2017.
- [27] Z. G. Wei, M. Y. Deng, and J. H. Liu, "Virtual and optimized design of mechanical arm of forestry felling & cultivation machine," *Applied Mechanics and Materials*, vol. 192, pp. 144–148, 2012.
- [28] C. M. Hall and Y. Ram, "Measuring the relationship between tourism and walkability? Walk Score and English tourist attractions[J]," *Journal of Sustainable Tourism*, vol. 27, no. 2, pp. 223–240, 2019.
- [29] J. Zhu and M. Zhu, "A study on the driving factors of tourism destination brand equity based on Internet sharing content: a case study of Shenzhen happy valley," *Journal of Service Science and Management*, vol. 11, no. 06, pp. 578–589, 2018.