Agricultural & Forestry Economics and Management (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/agrfem.2024.070110
ISSN 2616-2202 Vol. 7 Num. 1

Exploration of the Function Mechanism of "Village Super" and "Village BA" in Hunan County Economy Based on Collaborative Filtering and Apriori Algorithm

Yuan Shao

School of Computer and Electrical Engineering, Hunan University of Arts and Science, Changde, Hunan, 415000, China
1433884927@qq.com

Keywords: Collaborative Filtering, Apriori Algorithm, Coverage Rate, The Herfindahl Index

Abstract: In the rural economy of Hunan Province, the operation efficiency and market strategy of "Village Supermarket" and "Village BA" are of great significance for promoting local economic development. In order to explore the mechanism of collaborative filtering and Apriori algorithm in analyzing the economic activities of "Village Supermarket" and "Village BA", this study first introduces the position of "Village Supermarket" and "Village BA" in the county economy of Hunan Province, and the application value of data analysis technology in business decision-making. The method part describes in detail the steps of data collection, preprocessing, implementation of collaborative filtering algorithm and application of Apriori algorithm. By comparing and analyzing the execution time, diversity and coverage, we get the high efficiency of Apriori algorithm in dealing with specific data sets and its advantages in recommending diversity and coverage. Among them, the coverage of Apriori algorithm can reach up to 96%, the lowest value of Herfindal index is 0.07, and the highest execution time is only 3.2s. These findings provide important insights for understanding the mechanism of"Village Supermarket" and "Village BA" in the county economy of Hunan Province.

1. Introduction

Not only is it a bridge connecting farmers and the market, but the "Village Supermarket" and "Village BA" in Hunan Province play a unique role in promoting the circulation of agricultural products and economic growth.

For the first time, this paper tries to apply collaborative filtering and Apriori algorithm to the economic activity analysis of "Village Supermarketmarket" and "Village BA". Through the combination of these two algorithms, the pattern of user purchasing behavior and the association rules between commodities are revealed, which provides a new perspective for optimizing commodity layout and marketing strategy. In addition, this paper also compares and analyzes the performance of different algorithms in execution efficiency, recommendation diversity and coverage, which provides practical guidance for algorithm selection and system design.

This paper first introduces the research background, including the importance of "Village Supermarketmarket" and "Village BA" in county economy and the application status of data analysis technology, and then elaborates the research purpose and main contributions of this paper. Then this paper describes the research methods in detail, including data collection, preprocessing, algorithm implementation and performance evaluation. Then the experimental results are presented and the performance of different algorithms is compared and analyzed. Finally, the research findings are summarized, and the future research direction is proposed.

2. Related Work

Many researchers have studied collaborative filtering and Apriori algorithm, and Qiu Agen proposed a collaborative filtering recommendation method for government services combining user characteristics. In order to solve the problem that collaborative filtering does not consider user attributes, this method combines user portrait technology with it [1]. In order to understand the problem that the existing collaborative filtering technology of graph product makes the expression of embedded vectors of users and commodities unreasonable, Zhu Jinxia proposed a collaborative filtering recommendation model that integrated the attention mechanism of graph product [2]. Collaborative filtering technology based on deep learning neural networks has attracted much attention, Liu Hao proposed an automatic encoder for noise reduction, and finally verified it on the data set. The results showed that the noise reduction autoencoder had better generalization ability [3]. Zhao Wei statistically analyzed the repetitiveness of teaching syllabus knowledge points of 28 courses of a school's safety engineering major, built relevant data sets, and then used Apriori algorithm to analyze the correlation data of courses, and then proposed the possible causes and corresponding impacts based on the calculation results [4]. In order to effectively enhance the operation and maintenance level and management control level of power distribution terminal equipment, Ge Yonggao proposed an automatic diagnosis method for intelligent distribution terminal equipment defects based on Apriori algorithm [5].

In addition, Papadakis H summarized the methods in the whole research field of collaborative filtering recommendation system, and classified each method according to the tools and technologies used for easy understanding [6]. Ajaegbu C proposed an algorithm, which aimed to balance the current three traditional measurement metrics in the direction of cold start. The proposed algorithm not only alleviates the shortcomings of the three traditional algorithms in data sparseness or cold start, but also retains the good features of the existing project-based collaborative filtering algorithm [7]. Wang F first proposed a trust-based collaborative filtering (TBCF) algorithm to perform basic rating prediction with existing collaborative filtering methods, and then proposed a collaborative filtering method with user-project-trust records [8]. Edstamap used Apriori algorithm to obtain information about commodities in a database transaction of an eyeglasses company, so that these results could be used to promote the sales and marketing of eyeglasses [9]. Raj S proposed a new approach, the spark-based Apriori algorithm (SARSO) for reducing shuffle overhead. This algorithm uses partitioning method to limit the movement of key-value pairs between cluster nodes, thus reducing the necessary communication caused by Spark shuffle operation [10]. The research on collaborative filtering and Apriori algorithm has made remarkable progress, but it has not been studied in combination with county economy. In view of this, this paper aims to explore the function mechanism of "Village Supermarket" and "Village BA" in the county economy of Hunan Province by using collaborative filtering and Apriori algorithm. By combining the advantages of these two algorithms, this study is expected to reveal the key factors in rural economy more accurately, and provide data support and strategy suggestions for promoting local economic development.

3. Method

3.1 Implementation of Collaborative Filtering

Collaborative filtering algorithm is used to find the purchasing behavior pattern of users and recommend goods or services based on the exploration of the mechanism of "Village Supermarket" and "Village BA" in the county economy of Hunan Province [11-12]. First of all, it is necessary to collect the transaction data of users in "Village Supermarket" and "Village BA", including user ID, commodity ID and related rating or purchase behavior, as shown in Table 1:

Table 1: Transaction data

Transaction ID	User ID	Product ID	Purchase Date	Quantity	Total Amount	Rating	Review Content	Recommended	Related Product IDs
TXN1345	U01	P1501	2024-04-01	3	450.00	4.8	Great value	Yes	P4632
TXN1346	U02	P0502	2024-04-02	2	300.00	4.2	Good quality	No	P1512
TXN1347	U03	P7769	2024-04-03	1	180.00	5.0	Excellent	Yes	P1103
TXN1348	U04	P2254	2024-04-04	4	720.00	3.5	A bit pricey	No	P5533
TXN1349	U05	P1400	2024-04-05	2	250.00	4.5	Satisfactory	Yes	P1001

The basic contents of Table 1 include transaction ID, user ID, product ID, purchase date, purchase quantity, total amount, rating, evaluation content, whether to recommend and associated product ID. By collecting and analyzing this data, it is possible to gain insight into customer purchasing behavior and preferences and optimize product management and marketing strategies.

Data cleaning is performed first to deal with missing values and outliers to ensure data quality. A user-product matrix is then constructed, with rows representing users, columns representing products, and elements in the matrix representing the number of times users rated or purchased the product. The Pearson correlation coefficient is then used to calculate the similarity between users or between goods:

$$r_{xy}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{1}$$

r_{xy} is the Pearson correlation coefficient between two variables x and y; x_i is the i observation of variable x; y_i is the i observation of variable y; $\bar{x}$ is the average of the variable x; $\bar{y}$ is the average of the variable y; n is the number of observations.

User-based collaborative filtering finds other users with similar interests to the target user, while commodity-based collaborative filtering finds other goods that are similar to the target product. After that, recommendations are generated based on items that similar users like and that target users have not yet purchased:

$$\hat{r}_{ui}=\bar{r}_u+\sum_{v\in N(u)}\frac{s_{uv}}{\sum_{v\in N(u)|s_{uv}|}}(r_{vi}-\bar{r}_u) \tag{2}$$

$\hat{r}_{ui}$ is the predicted score of user u on item i; $\bar{r}_u$ is the average rating of user u; s_{uv} is the similarity between user u and user v; r_{vi} is the actual rating of user v on item i; $\bar{r}_u$ is the average rating of user v; $\sum_{v\in N(u)}$ is the sum of all users v in the similar user set $N(u)$.

Finally, appropriate evaluation indexes are used to evaluate the performance of the recommendation system, and the algorithm parameters are adjusted according to the evaluation results to optimize the recommendation effect. In the implementation process, it is necessary to pay attention to data sparsity and cold start problems, and adopt other techniques such as matrix decomposition or deep learning recommendation models to solve these problems. Through collaborative filtering algorithms, the role of "Village Supermarket" and "Village BA" in county economy can be better understood, and more personalized recommendation services can be provided for users [13-14].

3.2 Association Rules of the Apriori Algorithm

Analyzing transaction data to discover frequent itemsets and association rules is the main purpose of Apriori algorithm in exploring the mechanism of "Village Supermarket" and "Village BA" in the county-level economy of Hunan Province, and its role cannot be ignored [15-16]. Firstly, it is necessary to collect transaction data that includes customer ID, product ID, and purchase quantity, and convert this data into a transactional database format, where each row represents a shopping basket and each column represents a product. This study identifies the item sets of all individual products in the transaction database as candidates for frequent item sets. This study calculates the support for each individual item set and filters out frequent item sets based on a set minimum support threshold. In this study, these frequent item sets were used to generate higher-order candidate item sets, and Apriori attribute was applied for pruning to delete candidate item sets that could not become item sets. This study generates all possible rules for each set of frequent items and calculates their confidence, that is, the probability that one set of items will appear and another will appear. Then, based on the set confidence threshold, strong association rules are filtered out, and the frequent itemsets and strong association rules obtained are analyzed to understand the correlation between products, such as the combination of products purchased simultaneously.

3.3 Collaborative Filtering and Application of Apriori Algorithm

Using user behavior data to discover potential purchase patterns and user preferences is the characteristic of collaborative filtering, while Apriori algorithm extracts frequent and association rules. The combination of these two technologies can reveal the correlation between commodities and the purchase behavior patterns of users. Exploring the mechanism of "Village Supermarket" and "Village BA" in the county economy of Hunan Province needs to combine collaborative filtering and Apriori algorithm, which can provide a comprehensive analysis method [17-18].

Users' purchase history to predict other goods they may be interested in makes collaborative filtering help "Village Supermarket" and "Village BA" better meet customer needs and improve customer satisfaction and loyalty. For example, if the system finds that a user often buys a specific health food, it can recommend other similar health products.

Knowing which goods are often bought together can help "Village Supermarket" and "Village BA" to optimize the commodity placement and inventory management, which is also one of the functions of Apriori algorithm. Combining these two technologies, this study can deeply understand the relationship between customers' buying behavior and goods. Personalized recommendation provided by collaborative filtering can increase sales and customer satisfaction, and the association rules discovered by Apriori algorithm can help formulate more effective marketing strategies and improve inventory turnover [19-20].

4. Results and Discussion

In this study, a series of comparative experiments are set up to compare Apriori algorithm with

FP-Growth algorithm and Eclat algorithm, compare their performance in coverage, diversity and calculation efficiency, and deeply understand the advantages and disadvantages of these algorithms in analyzing the transaction data of "Village Supermarket" and "Village BA".

At the beginning of the experiment, it is necessary to collect and clean the transaction data from "Village Supermarket" and "Village BA" to ensure the data quality. Usually, according to the proportion of 70% training set and 30% test set, the Apriori algorithm is applied to mine frequent itemsets and association rules, and the support and confidence thresholds are set. Similarly, FP-Growth algorithm is applied to the training set to construct FP-tree and mine itemsets, and the support threshold is set, and then Eclat algorithm is run to mine itemsets from the training set data. Then, the coverage, diversity and computational efficiency of each algorithm are calculated by using 9 sets of data sets. Coverage is evaluated by the number of recommended commodities in the total commodities, diversity is expressed by Herfindal index, and calculation efficiency is evaluated by recording the running time of the algorithm and memory usage. Then comparing the performance of three algorithms on these indicators, use statistical tests to determine the significance of performance differences, analyze the performance of each algorithm on "Village Supermarket" and "Village BA" data sets, and discuss the influence of algorithm selection on the performance of recommendation system.

4.1 Coverage Rate

In order to further understand the recommendation effects of different algorithms in practical applications, this study designed a series of coverage comparison experiments to evaluate the comprehensiveness of three frequent itemset mining algorithms, FP-Growth, Eclat and Apriori, when recommending products. The result is shown in Figure 1:

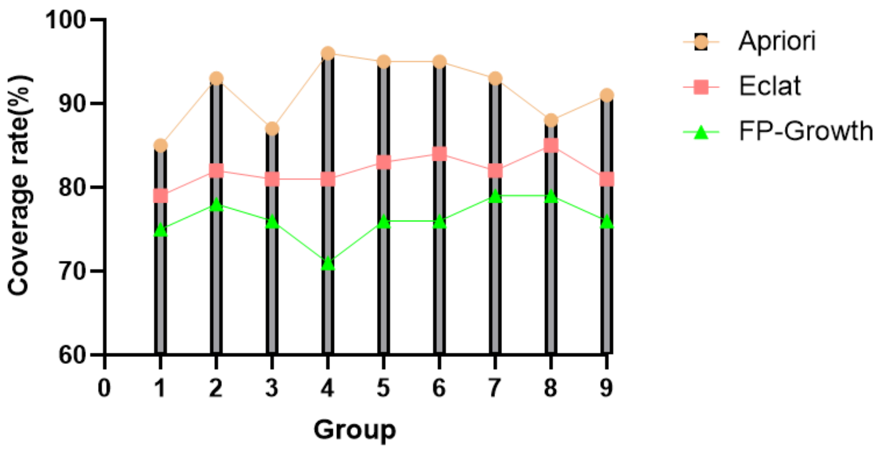

Figure 1: Comparison of coverage

According to the data in Figure 1, Apriori algorithm has the highest coverage rate of 96%, while Eclat algorithm and FP-Growth algorithm only have the highest coverage rate of 85% and 79%, respectively, which indicates that Apriori algorithm can recommend most products and provide users with a wide range of choices. In contrast, the Eclat and FP-Growth algorithms do not work well.

4.2 Diversity

Diversity is a key metric that measures the abundance of different items in the recommended list.

A highly diverse recommendation system can cover a wide range of product categories and provide users with diverse choices, thereby increasing user satisfaction and the attractiveness of the system. In this study, Herfindahl index was used to measure it, and the comparison results are shown in Figure 2:

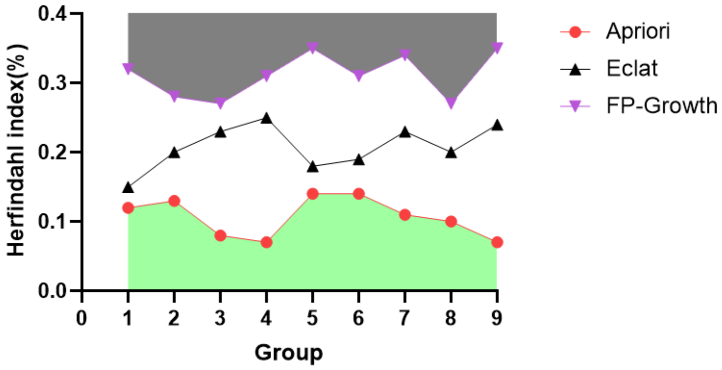

Figure 2: Diversity comparison

Figure 2 shows that the Herfindahl index of Apriori is lower than that of Eclat and FP-Growth algorithms in all data sets. Among them, the minimum value of the Herfindahl index of the Apriori algorithm is 0.07, and in the same situation, the minimum value of Eclat is 0.15, while the minimum value of FP Growth is 0.27. In the compared datasets, the Herfindahl exponent of the Apriori algorithm is generally lower than that of the Eclat and FP Growth algorithms, indicating that the Apriori algorithm provides higher diversity on these datasets. The minimum value of the Herfindahl index of the Apriori algorithm is 0.07, indicating that in some cases, it can evenly distribute recommended products, providing users with a wider range of choices.

4.3 Computing Efficiency

These experiments aim to quantify the time and resource consumption of each algorithm in the process of mining frequent itemsets, so as to reveal which algorithm is more cost-effective and scalable in practical application. This study uses the execution time to measure it, and the result is shown in Figure 3:

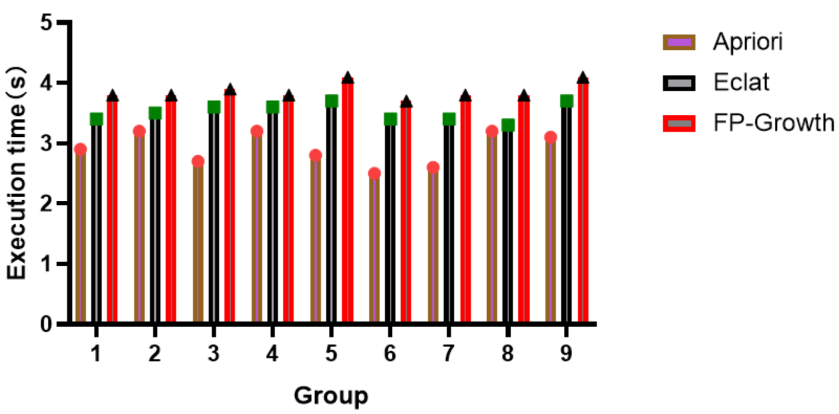

Figure 3: Comparison of computational efficiency

As can be seen from Figure 3, the execution time of Apriori algorithm is kept at a low level, the highest is only 3.2s, but the execution time of Eclat and Apriori is higher than that of Apriori. The execution time of Eclat algorithm is up to 3.7s, and that of Apriori is up to 4.1s, which indicates that under these specific experimental conditions, Apriori algorithm can effectively process data sets and generate results quickly.

5. Conclusion

A series of operations have proven that the Apriori algorithm performs well in terms of execution time and maintains a lower computational cost. This algorithm performs well and can provide a wide range of product recommendations to meet the needs of different customers, which is crucial for county-level economic entities such as "Village Supermarket" and "Village BA" because it helps to increase the range of customer choices and shopping experience. This diversity has a positive effect on improving customer satisfaction and promoting sales. Future research can further explore how to optimize algorithm parameters to improve computational efficiency, and how to combine these algorithms with other technologies such as deep learning to better adapt to constantly changing data characteristics and business needs. In summary, the combination of collaborative filtering and Apriori algorithm provides valuable insights into the mechanisms of "Village Supermarket" and "Village BA" in the county-level economy of Hunan Province. Through these findings, this study can provide more accurate data analysis and recommendation strategies for these county-level economic entities, thereby promoting local economic development and prosperity.

Acknowledgements

The work presented in this paper is financed by Key scientific research projects of Hunan Provincial Department of education in 2023(No.:23A0500), and Doctoral Research Fund Launch Project of Hunan University of Arts and Sciences in 2024(No.:24BSQD05).

References

[1] Qiu Agen, Zhang Yongchuan, Luo Ning, Zheng Yingying, LU Wen. Collaborative filtering recommendation method for government services based on User Characteristics. Integrated Technology, 2023, 12(1):42-55.

[2] Zhu Jinxia, Meng Xiangfu, Xing Changzheng, Zhang Xiaoyan. Collaborative filtering recommendation method for Convolution attention mechanism of Fusion graphs. Journal of Intelligent Systems, 2023, 18(6):1295-1304.

[3] Liu Hao, Guo Xiujuan. Collaborative Filtering Recommendation Algorithm Based on Noise Reduction Autoencoder. Journal of Jilin Jianzhu University,2023,40(2):84-88.

[4] Zhao Wei, WEI Zhongfei, WANG Kai, TONG Rui-Peng, WANG Liang, SONG Yannan. Optimization of safety Undergraduate modular course Content based on APRIORI algorithm. China Safety Science Journal, 2023, 33(6):41-48.

[5] Ge Yonggao, WANG Fuliang, Li Cheng, Lu Yujun. Automatic fault diagnosis method of intelligent distribution terminal Equipment based on Apriori algorithm. Automation Technology and Application, 2019, 42(5):4-7+11

[6] Papadakis H, Papagrigoriou A, Panagiotakis C, et al. Collaborative filtering recommender systems taxonomy. Knowledge and Information Systems, 2022, 64(1): 35-74.

[7] Ajaegbu C. An optimized item-based collaborative filtering algorithm. Journal of ambient intelligence and humanized computing, 2021, 12(12): 10629-10636.

[8] Wang F, Zhu H, Srivastava G, et al. Robust collaborative filtering recommendation with user-item-trust records. IEEE Transactions on Computational Social Systems, 2021, 9(4): 986-996.

[9] Edastama P, Bist A S, Prambudi A. Implementation of data mining on glasses sales using the apriori algorithm. International Journal of Cyber and IT Service Management, 2021, 1(2): 159-172.

[10] Raj S, Ramesh D, Sethi K K. A Spark-based Apriori algorithm with reduced shuffle overhead. The Journal of Supercomputing, 2021, 77(1): 133-151.

[11] Wu L, He X, Wang X, et al. A survey on accuracy-oriented neural recommendation: From collaborative filtering to information-rich recommendation. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(5): 4425-4445.

[12] Xu S, Tan J, Heinecke S, et al. Deconfounded causal collaborative filtering. ACM Transactions on Recommender

Systems, 2023, 1(4): 1-25.
[13] Fkih F. Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison. Journal of King Saud University-Computer and Information Sciences, 2022, 34(9): 7645-7669.
[14] Stinson C. Algorithms are not neutral: Bias in collaborative filtering. AI and Ethics, 2022, 2(4): 763-770.
[15] Santoso M H. Application of association rule method using apriori algorithm to find sales patterns case study of indomaret tanjung anom. Brilliance: Research of Artificial Intelligence, 2021, 1(2): 54-66.
[16] Sornalakshmi M, Balamurali S, Venkatesulu M, et al. An efficient apriori algorithm for frequent pattern mining using mapreduce in healthcare data. Bulletin of Electrical Engineering and Informatics, 2021, 10(1): 390-403.
[17] Liao M, Sundar S S. When e-commerce personalization systems show and tell: Investigating the relative persuasive appeal of content-based versus collaborative filtering. Journal of Advertising, 2022, 51(2): 256-267.
[18] Thakker U, Patel R, Shah M. A comprehensive analysis on movie recommendation system employing collaborative filtering. Multimedia Tools and Applications, 2021, 80(19): 28647-28672.
[19] Zhou X, Sun A, Liu Y, et al. Selfcf: A simple framework for self-supervised collaborative filtering. ACM Transactions on Recommender Systems, 2023, 1(2): 1-25.
[20] Dhinakaran D, Prathap P M J. Protection of data privacy from vulnerability using two-fish technique with Apriori algorithm in data mining. The Journal of Supercomputing, 2022, 78(16): 17559-17593.