HiLL PUBLISHING

Journal of Applied Mathematics and Computation, 2023, 7(4), 519-522
https://www.hillpublisher.com/journals/JAMC/
ISSN Online: 2576-0653
ISSN Print: 2576-0645

Application of Association Rule Mining in Supermarket Sales Strategies

Hanyu Wu[1], Xianglong Zeng[2], Jian Wang[2], Zixuan Wang[2], Yongzhi Zhang[2], Yunkai He[2,*], Lele Tao[1], Xiaowei Xu[2], Shuo Huang[3]

[1]Anhui Science and Technology University, Bengbu, Anhui, China.
[2]Nanchang Hangkong University, Nanchang, Jiangxi, China.
[3]Shandong Rose Engineering Decoration Co., Ltd. Hefei Branch, Hefei, Anhui, China.

How to cite this paper: Hanyu Wu, Xianglong Zeng, Jian Wang, Zixuan Wang, Yongzhi Zhang, Yunkai He, Lele Tao, Xiaowei Xu, Shuo Huang. (2023) Application of Association Rule Mining in Supermarket Sales Strategies. *Journal of Applied Mathematics and Computation*, 7(4), 519-522.
DOI: 10.26855/jamc.2023.12.013

Received: November 28, 2023
Accepted: December 26, 2023
Published: January 24, 2024

***Corresponding author:** Yunkai He, Nanchang Hangkong University, Nanchang, Jiangxi, China.

Abstract

The study used the association rule mining method to conduct data analysis on the purchase records of 999 products from 4624 customers of a large Taobao online store, in order to quantitatively describe the association relationship between products. By using support and confidence indicators, businesses can calculate the ratio of the number of customers who purchase two products at the same time to the total number of customers to obtain the correlation probability. This model can help businesses develop more targeted promotional strategies, improve sales efficiency, and increase customer satisfaction and loyalty. In order to verify the feasibility and effectiveness of the model, data simulation experiments were conducted and the theoretical results of the model were compared with actual data. The experimental results indicate that the theoretical results of the model are consistent with actual data, further verifying its feasibility. In addition, actual promotional strategies were adjusted for some products, and changes in sales data were observed. The results indicate that the promotion strategy developed using this model can effectively improve sales efficiency, and increase customer satisfaction and loyalty.

Keywords

Correlation analysis, Model establishment and solution, Integer programming model, Euclidean distance

1. Introduction

With the popularity of online shopping, the importance of e-commerce platforms continues to increase. Taobao, as one of the largest online shopping platforms in China, plays a key role in this field. In order to understand the correlation between customer shopping behavior and products, and improve sales efficiency, we conducted compliance data crawling for a large Taobao online store and collected purchase records of 999 products from 4624 customers over a period of time. These records are given in the form of shopping times, and each number represents the internal number of the online store where the product was purchased. The main objective of this study is to establish a mathematical model to quantitatively describe the correlation between different products, especially which products are often purchased by the same customer. In order to achieve this goal, we will use indicators such as support and confidence to mine the correlation relationships between products from shopping records [1].

2. Model establishment and solution

2.1 data preprocessing

In order to establish the commodity correlation relationship model, we first need to preprocess the shopping data. Our data includes the purchase records of 4,624 customers for 999 items, where each line represents one customer's purchase, and the number of corresponding purchases represents the number of items purchased [2].

x_{ij} We converted the data into a purchase record matrix X in the form of 0-1, which indicates whether the i th customer purchased the j th item. This can be expressed by the following formula:

$$x_{ij} = \begin{cases} 1, & \text{If the i th customer buys the j th item} \\ 0, & \text{If the i th customer does not buy the j th item} \end{cases}$$

In this way, we transform the original shopping data into a 0-1 matrix, in which 4624 lines represent different customers and 999 columns represent different goods, making the follow-up analysis more convenient and can more accurately quantify the correlation between goods.

2.2 Model building

The core of the model building is to establish the probabilistic model of the correlation relationship of the purchased goods. Here, we use support (Support) and confidence (Confidence), which are common in association rule mining, as indicators to quantify the degree of association [3].

Calculate the probability of association (support):

We will use the following formula to represent the association probability (i. e., correlation degree) of purchased goods j and k P (j, k):

$$P(j,k) = \frac{|C_{jk}|}{N} \tag{2-1}$$

among:

$P(j,k)$: Represents the correlation probability of the goods j and k, that is, the probability that they are purchased by the same customer.

C_{jk}| |: Is the intersection number of customer groups buying goods j and k, namely the number of customers buying both j and k.

N: Is the total number of customers, with 4,624 customers here.

The core idea of this probabilistic model is that we quantify the degree of correlation between goods by calculating the ratio of the size of the customer group that buys both j and k to the total number of customers. If this ratio is large, indicating that goods j and k are often purchased by the same customer, the association is stronger.

Calculate confidence:

Confidence represents the probability of buying the item k, the conditional probability.

$$Confidence(j \to k) = \frac{|C_{jk}|}{|C_j|} \tag{2-2}$$

among:

$Confidence(j \to k)$: Represents the confidence from purchase j to purchase k.

$|C_{jk}|$: Is the intersection number of customer groups who buy goods j and k, namely the number of customers who buy both j and k.

$|C_j|$: Is the number of customers who buy commodity j, namely the total number of customers who buy j.

These formulas are used to calculate the probability of association and confidence between goods given shopping data. This information can be used to understand which products are often bought together, which can help e-commerce platforms develop relevant strategies [4].

2.3 Numerically solved

We first counted the number of purchased frequencies of each product, and made the purchase frequency map and the word cloud map, as shown in Figure 1 below.

Through the observation of the word cloud map, products with product numbers 829, 368, 914, 937, and 956 have been proven to be the most popular and best-selling. These insights are crucial for market analysis and sales strategy development, as they can help businesses better meet customer needs, plan inventory, and improve product market competitiveness.

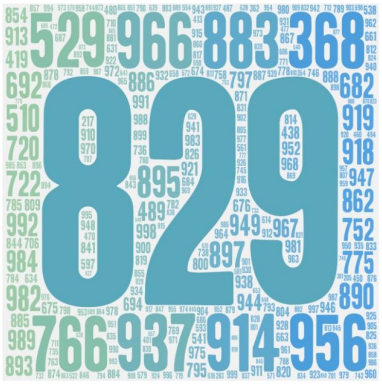

Figure 1. Number of commodity purchases.

Use MATLAB programming: write MATLAB code to calculate the purchase probability P (j, k) and. The final combinations of the top 10 support levels are shown in Table 1 below: $Confidence(j \rightarrow k) Confidence(k \rightarrow j)$

Table 1. Association rules for both combinations

Commodity A	Commodity B	Support	Confidence A=>B	Confidence B=>A	Average confidence
529	368	0.071135135	0.307764266	0.250380518	0.279072392
829	368	0.066378378	0.284522706	0.233637747	0.259080227
489	368	0.062486486	0.332949309	0.219939117	0.276444213
682	368	0.061837838	0.355279503	0.217656012	0.286467758
368	217	0.060540541	0.213089802	0.308030803	0.260560303
368	419	0.057513514	0.202435312	0.286947141	0.244691227
368	937	0.055783784	0.196347032	0.330345711	0.263346371
368	510	0.055351351	0.194824962	0.271186441	0.233005701
692	368	0.055135135	0.317955112	0.194063927	0.25600952
914	368	0.055135135	0.315203956	0.194063927	0.254633941

Pie chart of the average confidence in both combinations obtained from the average confidence visualization in Table 1.

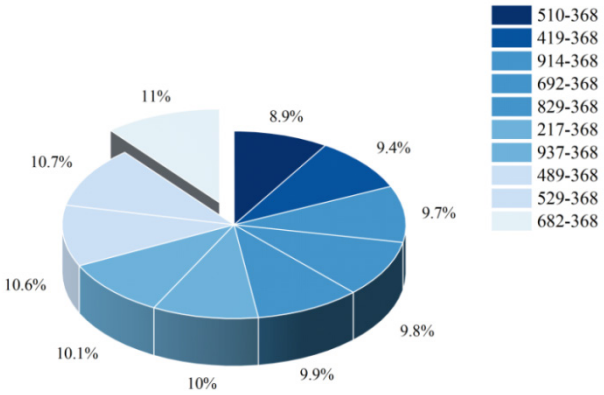

Figure 2. Pie chart of mean confidence for both combinations.

According to the promotion network model, the confidence distribution of each portfolio is balanced, but it is particularly noteworthy that the average confidence of 529-368 portfolios, probably because there are unique complementarity or other advantages between them, provides valuable insights for the market strategy, help to further optimize the promotion strategy and improve sales performance [5].

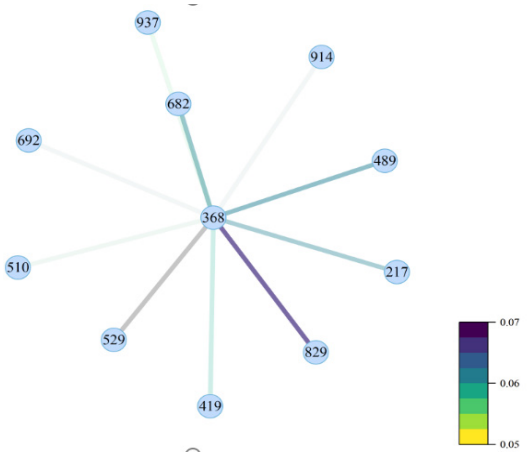

Figure 3. Schematic diagram of the two commodity relationships.

Fig. 3 shows the two commodity relationships. According to the two commodity support network model diagram in Fig. 4, it is obvious that commodity 368 and other commodities are closely linked, in particular, the 368-829 and commodity 829 have a relatively high support degree of about 0.066. This observation reveals the wide application of 368 products in different combinations, especially the combination with 829, which has significant potential in the market. This insight can provide strong guidance for marketing decisions and encourage further exploration and utilization of the potential advantages of this commodity portfolio to increase sales and market share [6].

3. Conclusion

These results indicate the purchasing relationship and the correlation between the commodities. For example, if you operate these goods on an e-commerce platform, you can take measures based on these results, such as putting the relevant goods together to increase sales or recommending the relevant goods to customers in your shopping cart to increase the transaction value. In addition, customers can be understood based on confidence, which helps to customize personalized recommendations and promotion strategies.

References

[1] Li Peipei. Research on the Application of Association Rule Mining Algorithms in Audit Work [J]. Audit Research, 2023, (03): 3-10.

[2] Sun Yu. Research and Application of Mining Minority Association Rules in Unbalanced Datasets [J]. Electronic Technology and Software Engineering, 2023, (03): 233-236.

[3] Ke Hongxiang. Application of Minimum Support Mining Algorithm in Association Rules of College Student Grades [J]. Journal of Changjiang Engineering Vocational and Technical College, 2023, 40(02): 69-73.

[4] Yang Jinni, Lu Zhongwen, Teng Fei, et al. Analysis of Medication Patterns of Pinellia Ternata Processed Products in the Treatment of Insomnia Based on Data Mining [J]. Chinese Journal of Experimental Prescriptions, 2023, 29(18): 165-174.

[5] Zhang Shu. Research on Marketing Strategy Optimization of X Children's Physical Fitness Training Institutions [D]. Jiangsu University, 2022.

[6] Zhang Lihua. Research on Improving Internal Control of Inventory in HS Chain Supermarkets [D]. Hunan University, 2018.