SCIREA Journal of Computer

http://www.scirea.org/journal/Computer

September 23, 2020

Volume 5, Issue 3, June 2020

Association analysis of dishes in Catering Enterprise

Xiaoyang Zheng, Yue Yu, Ziming Fang, Ling Huang

College of Artificial Intelligence, Chongqing University of Technology, Chongqing, China

Email: zhengxiaoyang@cqut.edu.cn

Abstract

This article implement Python language to realize the association rule mining of a restaurant company's dish data by using Apriori algorithm. Then the association rules of the dishes obtained are visualized by the R language expansion package. The frequent itemset no less than the minimum support and minimum confidence is found out and it is applied to mining the value information between dishes. The mining results show that homemade preserved pork, celery kidneys and braised pork are of strong association. The excavation results provide a basis for the recommendation of catering company and the formulation of setting meal plans, thereby increasing the corporate income.

Keywords: Association rules; Apriori algorithm; dish association analysis; R language

Introduction

With the development of informatization applications, various industries have produced

massive amounts of data. We are eager to analyze the massive amounts of data in-depth and extract the hidden information in order to make better use of these data. However, only the functions of the database system cannot discover the relationships and rules that exist in the data and cannot predict the future development trend based on the existing data, and lack the means of mining the hidden knowledge behind the data. It is under this condition that data mining technology came into being [1]. In this paper, the association rules regarding data of dishes of real catering company are explored with Apriori algorithm and visualized analysis for the exploration results is carried out by using the R language extension package.

Basic idea of association rules

Exploration of association rules can be described as [2] that the set of all the items in dish data is $I=\{i_1,\cdots,i_m\}$, the database transaction is $D=\{t_1,\cdots,t_m\}$ composed of a series of transactions with the unique identification. The itemset included in each transaction is one subset on the I.

1) An itemset refers to a set of items. The support of the itemset A_0 is the percentage ratio of A_0 to D. Frequent itemset is defined as the itemset whose support is no less than the minimum support set.

2) The association rule [3] is an implicit formula of the form $A \to B$, which means that B can be deduced from A. $A \subset I$ and $B \subset I$ are called as the antecedent and consequent of association rules, respectively [3]. The Rules are established in the transaction set D and are of support and confidence. In this paper, the support and confidence are used to quantize the association rules between things. In addition, the rule reached the minimum support value and the minimum confidence value is called strong rules [4]. The lift value represents the lift of used rules compared with unused rules.

The probability of simultaneous occurrence of itemsets A and B is called the support of association rules:

$$\text{Support}(A \to B) = \text{P}(A \cup B). \tag{1}$$

The probability of occurrence of event B on the basis of occurrence of itemset A is called the confidence of association rules:

$$\text{Confidence}(A \to B) = \text{P}(B|A). \tag{2}$$

Lift formula:

$$\text{Lift}(A \to B) = \text{Confidence}(A \to B) / \text{support}(B). \tag{3}$$

Apriori algorithm

Apriori algorithm belongs to Boolean association rules exploration algorithm [5]. The frequent itemset is generated by using the candidate set, and then the association rules are realized by using the frequent itemset [6]. The Apriori algorithm uses an iterative method of layer-by-layer search to scan the database multiple times to find all frequent itemsets. The scanning steps are described as follows:

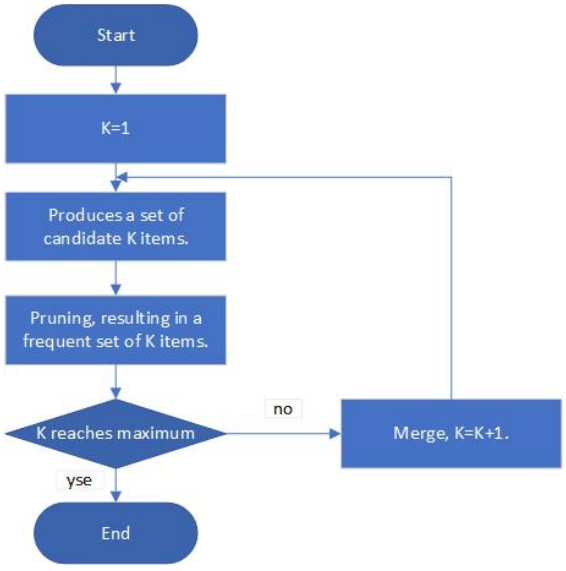

Figure 1. Apriori flow chart

Dish association analysis by Apriori algorithm

Data preprocessing

The raw data are converted into a user order table. Each row of the user order table represents all the dishes. Removing the idle value and invalid symbol, the rest data are saved. Because Apriori algorithm on exploration analysis is only applicable to Boolean data, the name of unstructured dishes should be converted into structured data which can be identified by computers. Dish information is converted into a 0-1 matrix, that is, ordered dishes are represented by 1 and unordered dishes are represented by 0.

Association rule mining of the dishes

The dish data processed are analyzed by using Apriori algorithm. The minimum support is set as 0.03 and the minimum confidence is set as 0.5 to ensure that the number of association rules exported is reasonable. Functions are explored by using association rules to export the frequent itemsets whose support and confidence are greater than the minimum support and the minimum confidence and to export the support and the confidence of the frequent itemset. The exportation partial results are as shown in Table 1:

Table 1. Dish association rules

	support	confidence
Braised pork – Celery kidneys – Spicy fired squid – Homemade preserved pork	0.0307856	0.87878788
Homemade preserved pork- Celery kidneys – Spicy fired squid – Braised pork	0.0307856	0.85294118
Braised pork – Homemade preserved pork – Spicy fired squid-Celery kidneys	0.0307856	0.85294118
Spinach salad – Braised pork- Homemade preserved pork – Celery kidneys	0.0339703	0.84210526
Braised pork – Mongolian Roasted Lamb Leg – Spinach salad	0.0339703	0.82051282
Spinach salad – Braised pork – Celery kidneys – Homemade preserved pork	0.0339703	0.8
Homemade preserved pork – Celery kidneys – Braised pork	0.0477707	0.78947368
Chopped pepper fish head – Mongolian Roasted Lamb Leg – Spinach salad	0.0318476	0.78947368

Spinach salad – Homemade preserved pork – Celery kidneys-Braised pork	0.0339703	0.7804878

To intuitively display the association of the dishes, this paper applies the extension package to draw Association rules with the top 20 support in Figure 2. The dishes are marked as the top point and the association rules are represented by the arrows connecting to the top point of the dishes. The size of the graph represents the value of the support. Similarly, the color degree of the circle is darker, the lift of the rule is larger. This indicates that the dish set is more universal and the probability of the dish combination is larger.

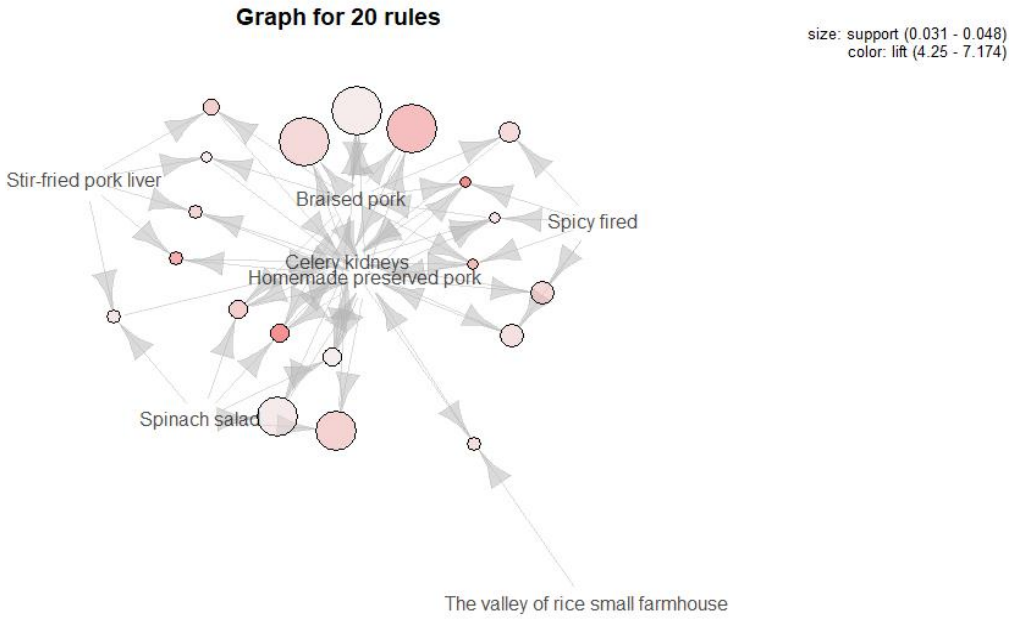

Figure 2. Network diagram of the association rules

According to Figure 2, the homemade preserved pork, celery kidneys, braised pork, spicy fried squid, spinach salad and stir-fried pork liver are of strong association. Consequently, the sales volume may be improved if the celery kidneys, spinach salad and homemade preserved pork are recommended to customers who order braised the pork, and then the operating revenue will be improved. Accordingly, it can consider a price reduction promotion for one of the dishes. Although the price of the related dishes remains unchanged, the customers will increase their desire to buy because of the dish discounts. Of course, the profits lost because of the discounts can get more profits through the recommendation of the associated dishes.

Conclusion

According to the exploration of the association rules based on the dish data, the strong associations of the homemade preserved pork, celery kidneys and braised pork are obtained in this article. Catering enterprises recommend dishes according to association analysis of dishes to promote dining consumption of the customers and bring the customers good dining experience. According to the sales situation of the dishes, we can comprehensively consider many factors to predict the sales of the dishes so that the catering companies can prepare the raw materials in advance. In addition, the dish analysis of catering companies can provide more reference models for the industry, and the analysis of association rules can help discover new information to facilitate decision-making and formulate more reasonable set marketing strategies.

References

[1] Python data analysis and mining actual combat. Machinery Industry Press, 2016.

[2] A Survey of Association Rules Algorithms. Electronic Testing, 2016(7X):41-42.

[3] Srikant R. Fast algorithms for mining association rules and sequential patterns, University of Wisconsin,1996.

[4] Agrawal R. Fast Discovery of Association Rules. Advances in Knowledge Discovery & Data Mining, 1996.

[5] Research on the method of combining clustering and association rules to improve the quality of catering services. Modern Information Technology, 2019, 3(16): 10-12.

[6] The application of algorithm based on R language in mining commodity transaction data. Computer programming skills and maintenance, 2015(2): 54-56.