Research on Commodity Expectation Based on ABC Classification and Association Rules

Song HaiYan[1, *], Zhang Huan[2]

[1]Hu Zhou Vocational & Technical College, 299 Xuefu Road, Huzhou City, China
[2]North China Electric Power University, No.619, Yonghua North Street, Baoding City, China

Abstract. For some companies, sales means arranging goods and selling them on the shelves. This kind of arrangement of goods regardless of priority brings a lot of internal consumption to the enterprise. With the rapid development of the Internet and industry, many large companies have gradually established and improved their production and sales processes. The traditional ABC classification method can no longer meet this demand. The new method that can compensate for the traditional ABC classification method is particularly important. This article uses association rules to use the ABC classification method to study the expected budget of commodities, uses the Apriori and FP-Growth algorithms in the association rules to collect frequent itemsets of the acquired data, calculates the corresponding association rules, and then classifies the commodities with ABC.

1 Introduction

According to references, when the association rule algorithm generates each candidate item set during the calculation process, it will generate many subsets in the loop, which requires a lot of I/0 load. Although the ABC classification method can circumvent this problem, it also has shortcomings [1][2][3]. ABC analysis method can capture cost-intensive products or processes, which helps to implement key management and improve efficiency. However, due to unreasonable cost allocation in practice, it is easy to miss some important factors. This situation is the focus of research on the expected sales amount budget of inventory products, and there are some professional Internet organizations that conduct specific analysis of these data and publish relevant results for use and reference by enterprises. This article uses association rules in conjunction with ABC classification to complete the demonstration and analysis of the impact on the expected consumption amount of goods.

2 Introduction to ABC Classification Method and Association Rules

2.1 Introduction to Association Rules

In 2001, Agrawal [4] proposed association rules when analyzing information between data. The academic definition is for the implication formula of the form X->Y. The characteristic of this implication formula is that in the representative form of X and Y, they represent the predecessor and successor respectively. Where X represents Antecedent or Left-hand-side, and Y represents Consequent or Right-hand-side. RHS not only has associations, but also carries attributes for support and trust.

2.2 Introduction to Apriori Algorithm

Apriori is a data frequent itemset algorithm based on mining association rules. The core of this algorithm is to mine frequent itemsets through two stages: candidate set generation and plot downward closed detection. At present, the evaluation of Apriori is the frequent itemset algorithm of Boolean association rules. The frequent itemset must be greater than or equal to the minimum support, and then a strong association rule is generated from the frequency set, which must meet the minimum support and minimum credibility standards. The Apriori algorithm has important reference value for the marketing management of diversified brand companies. According to references [5][6], the Apriori algorithm will scan the original data multiple times. Its algorithm is inefficient and time-consuming, and it is possible to produce a huge candidate set, which is an obvious disadvantage of this algorithm.

2.3 Introduction to FP-Growth Algorithm

FP-Growth is one of the algorithms used to mine frequent itemsets, compressing the acquired data to form a frequent pattern tree, and processing the associated information of the itemsets in the form of the tree. The special prefix FP-tree is used for the structure of the entire tree to form a structural model of frequent item header table and prefix tree. This method can speed up

* Corresponding author: paperbo@sina.com

the process time of data mining and association. FP-Growth is proposed mainly because the frequent itemsets generated by the Apriori algorithm in the data processing process will scan the data multiple times and generate a large number of candidate frequent itemsets, which will increase the complexity of time and space when the algorithm is scanned. Reduce the performance of scanning.

2.4 Introduction to ABC Classification

The ABC classification method mainly analyzes the storage and management of items. After years of use and verification, its effect is relatively good. The traditional ABC classification method has both advantages and disadvantages in the inventory and merchandise sales management process, although on the surface it seems that the main energy is spent on solving important things, however, in terms of the actual product sales budget, the variety of products does not fully reflect the weight value of the product, which may cause the weight of the transaction amount to be low, but the purchaser considers the high product to be missed or postponed [7][8].

3 Algorithm Selection and Feasibility Analysis

3.1 Algorithm Selection

There are two main types of association rules. The first one uses the Apriori algorithm, and the second uses the FP-Growth algorithm. These two algorithms are currently the most used algorithms in the calculation of association rules. The Apriori algorithm takes a long time to scan all the data in the database multiple times, thereby generating a large number of candidate frequent sets, and processing data multiple times in all non-empty subsets, which will lead to its performance degradation. The FP-Growth algorithm is an algorithm improvement based on the original Apriori algorithm to fix this drawback. The FP-Growth processing scan method only needs to scan twice. The first scan is to build the FP tree, and the second scan is to mine frequent itemsets from FP. This method can obtain item set information more quickly.

3.2 Feasibility Analysis

This article is divided into two parts in terms of technology. The first part uses the Apriori algorithm to obtain frequent itemsets of the data and calculates the results of association rules, uses the Apriori algorithm to collect the frequent itemsets of the data, filters the frequent itemsets based on the support, and finally measures the credibility through the original frequent itemsets to obtain association rules. The second part is the ABC classification of the original data. The ABC classification method arranges the target code and the cumulative ratio of the sales amount and income as the basis for dividing the ABC product category. The production process of each coded product The amount is calculated and made into a table, and then it is cumulatively compared, and finally the structure of the classification is completed. Matlab is used for data processing when using the Apriori algorithm.

4 Data Acquisition and Analysis

4.1 Data Collection

Data acquisition is mainly collected through the sales records of the MYSQL database. Afterwards, code conversion is performed to facilitate subsequent data processing. When exporting data, the header name of the data needs to be defined and corresponding to its specific database field to ensure the correctness of the data. Data acquisition and conversion process, and a graphical representation of results is shown in figure 1.

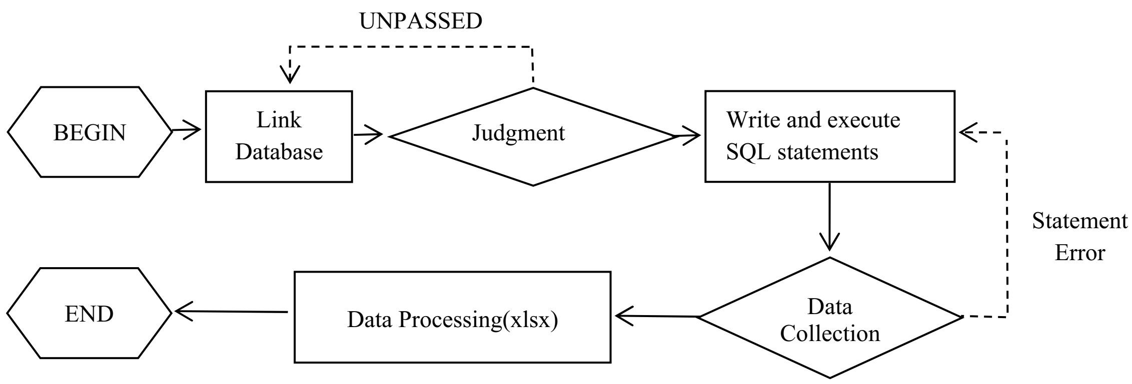


Figure 1 Data acquisition design process

4.2 Data Design and Analysis

Obtain target data by analyzing frequent itemsets, complete the screening of commodity frequent itemsets, and generate candidate sets. Eliminate the itemsets in the candidate set that do not meet the minimum support to obtain the system candidate set selection. Then merge the remaining data to form an item set containing two

https://doi.org/10.1051/e3sconf/202129202026

elements, and remove the item set that does not meet the minimum support. This process will be repeated.

4.2.1 The Design of Candidate Set for Expected Sales Amount

First tran each transaction record in the data, then can judge each candidate set. If all the subsets of 'can' and 'tran' meet the requirements, increase the count of can, and then judge each candidate set. If it is determined that it is greater than the minimum support, then the current candidate item set is reserved, otherwise a list of all frequent itemsets is returned. The setting process for generating candidate item sets is shown in Figure 2.

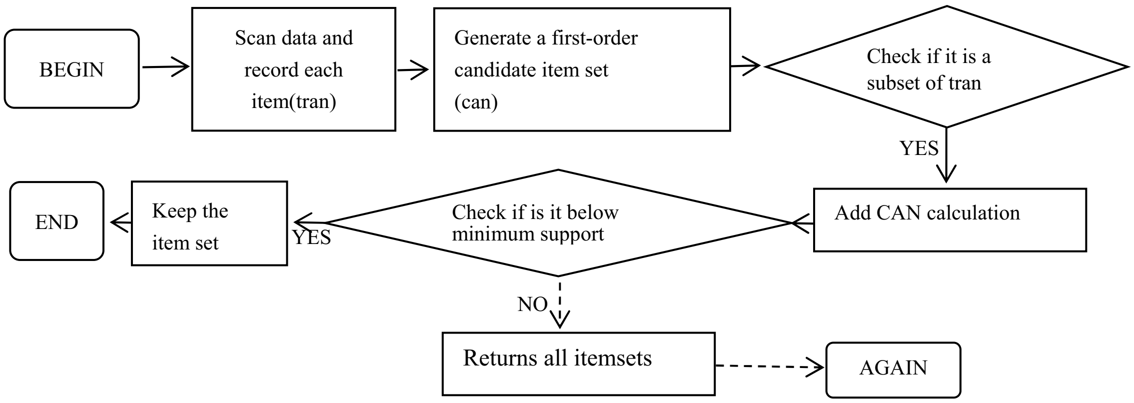


Figure 2 Generate the Candidate Item Set Design Flow Chart

4.2.2 Expected sale item amount frequent item set acquisition design

The acquisition design of frequent itemsets is mainly carried out through Aproiro's complete algorithm. The entire frequent itemset acquisition design is to first construct a candidate set list consisting of multiple items, and temporarily set it to K. Then use the algorithm to determine the support for its candidate item set. If the minimum support is not met, it is an infrequent item set and delete it. And Construct a list of candidate item sets composed of K+1. In the process of generating, first sort each item set according to the obtained elements, and compare the two item sets. Through comparison, it is found that only when the K+1 items are the same and the last item is not the same, the two obtained item sets are merged to generate the required new candidate item set. In this way, it can be determined that a certain item set is frequent, and then all its subsets are also frequent. Similarly, if the items set is not a frequent item set, then its subset is also an infrequent item set. It can be excluded directly or directly after calculating its support. The overall process is shown in Figure 3.

4.2.3 Frequent Item Set Evaluation Standard Design

Association rule design is based on the evaluation criteria of frequent item sets, which are divided into support, confidence and lift. There are corresponding construction algorithms for the three standards. Support is the proportion of the overall data through the number of occurrences in the data set, and can also be understood as the probability of multiple data associated. If the two data X and Y that need to be analyzed are used as references, then the corresponding support algorithm is as follows.

Support (X, Y) = P(XY)= number (XY)/ num (AllSamples)

By analogy in this way, if three data are used, the corresponding support calculation method.

Support (X, Y, Z) = P(XYZ)= number (XYZ) /num (AllSamples)

In this case, the support degree of the construction does not necessarily represent its corresponding frequent itemsets, but the support degree is low, it must be infrequent item sets. In addition to the degree of support, the degree of confidence represents the probability of the appearance of another data after the appearance of one data is completed. Similarly, the two data X and data Y that need to be analyzed for relevance are compared. The calculation method of the confidence of X versus Y is as follows.

Confidence (X, Y)=P(X|Y)=P(XY)/P(Y)

In the same way, the correlation confidence between multiple data can be completed. For example, the correlation confidence calculation method using XYZ data is as follows.

Confidence (X, Y, Z)=P(X|YZ)=P(XYZ)/P(YZ)

This kind of confidence and support is decisive after calculation. It can be regarded as when selling goods, users who buy goods X have a percentage of support to buy goods Y. In this way, after purchasing X product, there is a corresponding support to purchase Y product, and a percentage of users who have purchased X product at the same time corresponds to the confidence that users buy Y product. This situation needs to be expressed by lift. Under the condition of buying goods Y, it also contains the percentage probability of being able to buy goods X and the ratio of the probability of simultaneous purchase of goods X. The corresponding algorithm is as follows.

Lift (X, Y) =P(X|Y)/P(X)=Confidence (X, Y)/P(X)

This calculation method improves the relationship between X commodity and Y commodity. When the corresponding increase is greater than 1, it means that X<=Y are two commodities with strong association rules.

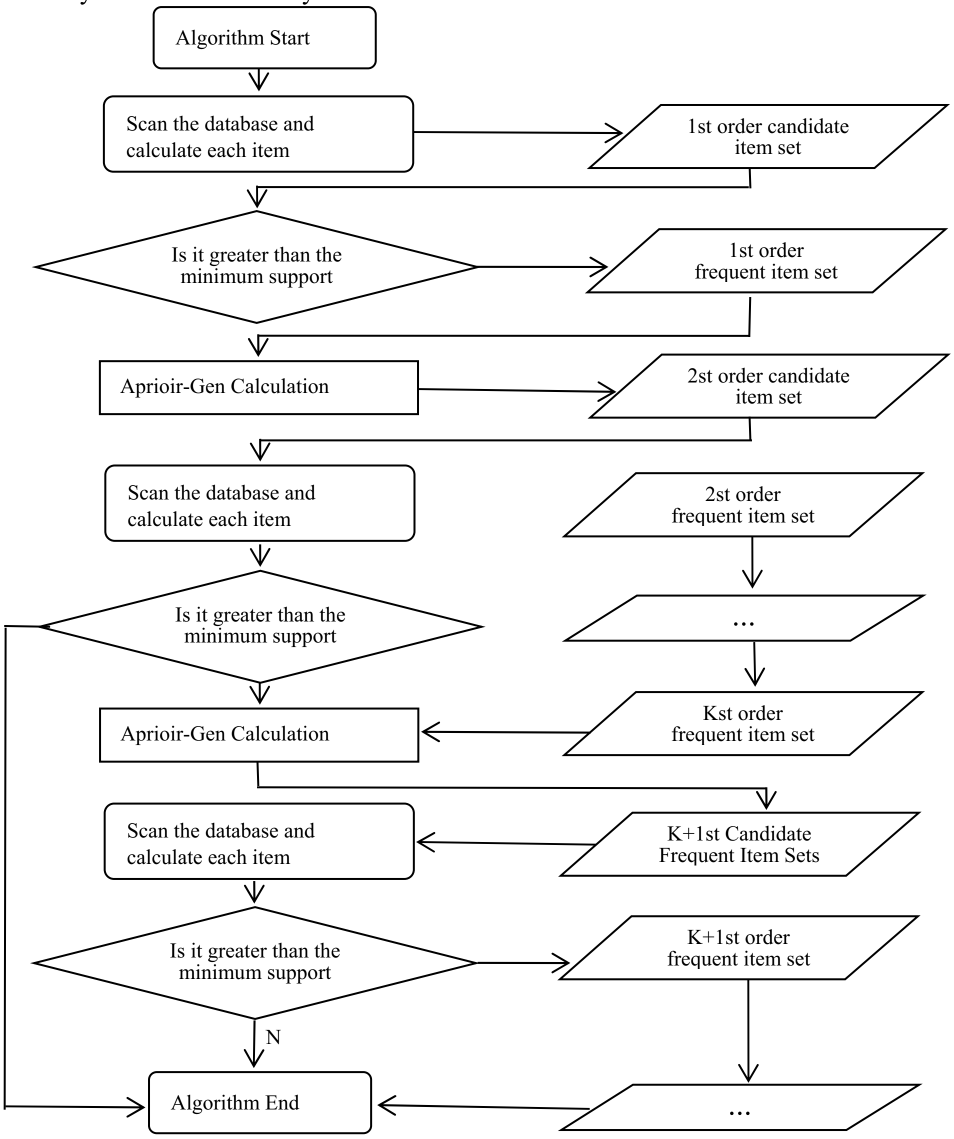


Figure 3 Frequent Item Sets Design Flowchart

4.3 ABC Method Classification Design

Regarding the expected sales amount of the product, the product sales category code, the percentage of product revenue to the total, and the cumulative percentage of sales, the design steps are as follows. First, collect data such as the sales volume and unit price of all single products. Second, calculate the acquired data, such as sales, product categories, cumulative product sales, percentage of total revenue, cumulative percentage, etc. Third, make an ABC classification table, use the enumeration method when there are not many product categories, and use the layered method when there are too many categories. Forth, use the cumulative category percentage as the abscissa, and use the data in the ABC analysis table to make an ABC analysis chart. Fifth, use different strategic decisions on the results of the analysis.

4.4 Data Analysis

4.4.1 Candidate Set Collection

When collecting data, the first step in the data association rule is to collect the candidate item set, and to collect product sales and purchases of different order data. The calculation results show that the original data is not coherent data, so after processing the data. It can be seen that the original data is stored separately, so the first step is to merge the data. The merge rule uses the UPC code of the product, the amount and the order time.

We collect itemsets by encoding the data of the system, using the method of traversing the data. If the code in the traversal exists, increasing its count value. The result of the frequent itemset data obtained for the first time is obtained. At this time, the data table C1 can be obtained by counting the support degree of all the data. At this time, set the minimum support to 1 (minSupport to1), and screen out unsatisfied data, which is divided into code and data value. Get the largest set of items at this time, and the dimension at this time is one-dimensional.

4.4.2 Apriori Gets Frequent Itemsets

We sort all the data in descending order and name the frequent itemset data obtained for the first time as L1. We measure the candidate set C2 through L1, scan the original merged data, and count the support of the data in each candidate set, and by filtering the data by meeting the minimum support degree, the item set L2 at this time can be obtained. At this time, the data dimension is two-dimensional. The corresponding candidate item set data L3 is measured by obtaining the data L2 for the second time, and then the third item set L3 is obtained by comparing the set minimum support minsupport. At this time, the data dimension is three-dimensional. Through the third data acquisition, it can be seen that the data has been basically completed after the three-dimensional calculation, and the frequent itemsets and their subsets of the data have been acquired. After the third data acquisition, the frequent itemsets can be obtained.

4.4.3 ABC Classification Data Processing

The results of the first data processing are summarized with the original data, and the classification uses the product UPC code to correspond to the financial sales, financial income, unit price of the product, and the income that needs to be used in the future as a percentage of the total, and the cumulative percentage of Income. Through these data, the ABC classification table of the later period is made, and the processing result of the data is shown in Table 1.

Table 1.ABC classification method demand data calculation realization

UPC code	Financial sales	Revenue	Unit price	Theratio to revenue	Cumulative ratio
4537053	18464	90988.0257	4.93	0.062418571	0.062419
6082959	8213	25854.5151	3.15	0.17736421	0.080155
5904635	5500	32974.9998	4.36	0.016447057	0.096602
2242220	3489	8544.7624	2.45	0.005861781	0.102464
6082997	2631	135761.7636	51.60	0.093133742	0.195598
5568661	1774	47572.84	26.82	0.032635379	0.228233
6238447	1766	9073.3486	5.14	0.006224395	0.234457
5673630	1173	16058.1999	13.69	0.011016064	0.245473
2079061	1149	6955.4999	6.05	0.004771533	0.250245
2628325	876	11967.61	13.66	0.008209884	0.258455
2518084	805	2704.84	3.36	0.001855544	0.260310
5058703	745	9335.78	12.53	0.006404426	0.266715
5174115	656	9048.2401	13.79	0.006207171	0.272922
5848781	645	11162.17	17.31	0.007657345	0.280579
7314659	619	25522	41.23	0.017508312	0.298088
8391329	593	35276.8219	59.49	0.024200204	0.322288
7873377	487	19534.85	40.11	0.013401076	0.335689
2184959	423	19654.36	46.46	0.013483061	0.349172
7575203	417	23711.65	56.86	0.016266397	0.365438
5848741	378	10416.66	27.56	0.007145919	0.372584
2977424	359	46019.88	128.19	0.031570035	0.404154
5174109	355	6418.91	18.08	0.004403428	0.408558
6602648	345	6061.27	17.57	0.004158084	0.412716
6077100	341	14871.68	43.61	0.010202101	0.422918
5058661	335	16494.34	49.24	0.011315260	0.434233

4.4.4 ABC Classification Results

There are many differences in commodity prices and sales quantities. The total number of commodity data is 522. When using the product ABC to make a classification table, it is difficult to list all the codes of the products one by one. In this case, the data will be listed and displayed in a hierarchical way of sales.

For the observation through the data sheet, it can be found that, the main data distribution range is from 10,000 yuan to 100,000 yuan, and the rest is more than 100,000 and below 1,000 is another range, the main situation is multi-point hash distribution. Discovered by observing the data, now we divide the 10,000 yuan to 100,000 yuan into the first level, the 1,000 yuan to 10,000 yuan as a range into the second level, and divide the over 100,000 and below 1,000 into the third level, and the classification after the division is completed The screenshot of the table part is shown in Table 2.

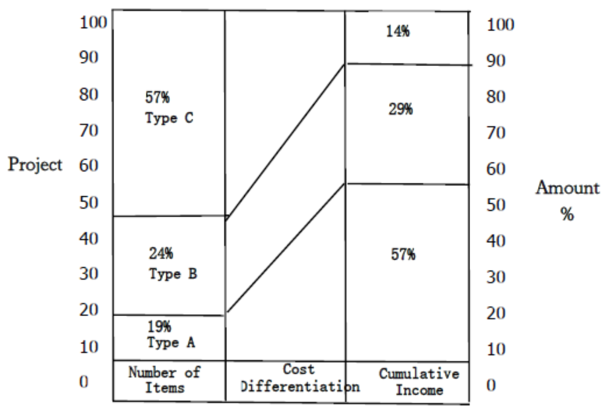
https://doi.org/10.1051/e3sconf/202129202026

Table 2. ABC Classification Table

UPC Code	Percentage of Income	Cumulative Ratio	Classification
4537053	6.24%	6.24%	The First Layer
6082959	1.77%	1.77%	
5673630	1.10%	1.10%	
...	57%	57%	
2518084	0.18%	0.18%	The Second Layer
6602648	0.42%	0.42%	
3919501	0.36%	0.36%	
...	29%	29%	
6082997	9.31%	9.31%	The Third Layer
8231303	0.05%	0.05%	
8764181	0	0	
...	14%	14%	
	1	1	TOTAL

After the product category has been stratified, all products with a cumulative sales of 60%-80% are defined as category A through the layering of the products and the cumulative sales of the products. The product category that maintains the sales degree of the product here at 20%-30% is defined as category B, and the last other categories of data are defined as category C. Through these data, the sales amount and sales accounted for the total percentage and cumulative ratio data. Construct an ABC analysis table, as shown in Table 3.

Table 3. ABC analysis table



5 APPLICATION CONCLUSION

The related design of this article is a solution to the confusion of spare parts inventory management in the production and sales process of enterprises, which is out of touch with sales management. Based on the ABC classification method, using association rules, and calculating the internal logical association rules between inventory and sales behaviors in various key processes through reasonable algorithms. The above results will provide decision makers with more accurate inventory and sales management arrangements, and reduce inventory and sales early-stage costs.

References

1. HU Cong-xu. The Implementation and Improvement of ABC Classification in Stock-managing [J]. Logistics Engineering and Management, 2010, 32(09): 62-64.
2. Silver E A, etal. InventorB management and production planning and scheduling (3rd edition) [M].1988: 30~35.
3. Flo res B E, WhB bar k D C. Implementing multiple criteria ABC ana lBsis [J]. Journa l of Operations Management, 1987, 7 (1& 2): 79~ 85.
4. AGARWAL RC., PRASAD VVV., AGGARWAL CC.. A tree projection algorithm for generation of frequent item sets[J]. Journal of Parallel and Distributed Computing,2001,3(3).
5. RAO Zheng-chan, FAN Nian-bai. A review of associative rule mining Apriori algorithm[J]. Computer Era, 2012(09):11-13.
6. S. Brin, R. Motwani, C. Silverstein. Beyond market baskets: generalizing association rules to correlations. In Proceedings ACM SIGMOD International Conference on Management of Data. ACM Press, Tucson, Arizona, USA, 1997.5: 265-276
7. Dantzing G B, Ramser J H.The truck dispatching problem[J].Management Science, 1959, 6(1): 80~91
8. LUO Lai-gen, YU Min-yan. Analysis on the Application of ABC Classification Method in Inventory Management[J]. Marketing Management Review, 2019(01):130.