

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Application of Apriori Algorithm to Customer Analysis

Mu jiankang

School of Management, Henan University of Technology,
Zhengzhou, Henan 45001, China

Abstract: How to maximize the information is a very important problem for the decision-makers. Apriori algorithm based on association rules of data mining technology has been employed as the research instrument. This study analyzes the definition of association rules and apriori algorithm, studies the process of apriori algorithm, make an empirical analysis on the consumer's purchase behavior. The experimental results show that apriori algorithm is an important technology to help managers make accurate decision. In this study, author also built a model to identify frequency of the customer behavior.

Key words: Data mining, apriori algorithm, customer relationship management

INTRODUCTION

Market competition of the domestic industry is increasingly fierce, with more emphasis on customer relationship management and on that the explosive growth of customer data needs the support of data mining (Gu and He, 2010). Customer Relationship Management (CRM) is a customer-oriented business strategy, it is the means of information technology, customer data as the basis, through data analysis technology to strengthen the understanding of customers and the CRM process redesign. How to use effective tools to find the valuable information from massive has become a hot topic at present. As the most effective means, data mining is an essential part of customer relationship management.

Data Mining can help CRM to analyze large data streams and gain insight into customer behavior, needs and preferences. Such knowledge facilitates the design of customer-centric business processes as well as personalized marketing and service activities, which, in turn, help to leverage customer loyalty and maintain competitiveness in globalized and saturated consumer markets. And, it enables them to determine the impact on sales, customer satisfaction and corporate profits. Finally, it enables them to “drill down” into summary information to view detail transactional data.

The goal of customer analysis is to target high-value customers with product and service offerings tailored to their specific needs. To achieve this targeted capability, companies segment their customers into groups based on demographic information that provides insight into their purchasing behavior. Various studies on customer behaviors have been presented and used in real problems. Data mining techniques are expected to be a more effective tool for analyzing consumer behaviors. However,

the data mining method has disadvantages as well as advantages. Therefore, it is very important to select effective appropriate algorithm to mine databases. Apriori algorithm is one of the most effective mining Boolean association rules algorithms. On the above situation, this paper mainly analysis customer behavior by apriori algorithm of association rules and get the customer relationship. So that it can provide scientific basis for the marketing decision-making.

LITERATURE REVIEW

The most important theme of data mining is the association rules. Association rules can be found that the relationship among the different commodities in the transaction database, these rules can find the mode of customer purchase behavior, such as the purchase of a commodity to buy other goods. These rules can be applied to the goods shelf design and goods storage arrangements and so on.

Association rules are first proposed by Agrawal and Srikant (1994). It is one of the most basic problems in data mining field. Association rules is necessary to Customer Relationship Management (Gu and He, 2010). Association rules needs to be explore valuable information on transactional data to assess customer's preferences for products sold (Sulianta *et al.*, 2013). It can be used to establish customer classification model, which may be used for facilitating cross-sales in many industries, such as mobile telecom market and so on (Ahn *et al.*, 2011). It also can be used for customer segmentation. It extracts the related data fro a sample survey data to be a basis of the analytic process, then implements the operation of rule discovery and customer segmentation. The study on VIP customer by Zhang validates this idea (Zhang, 2010).

Recommender systems are usually regard as another important use for association rules in the e-commerce (Kim, 2011). Furthermoe, association rules can be used to discovery potential customers.

Apriori algorithm plays a distinct role in the field of information technology. With the unceasing development of information technology, enterprise's demand of data mining is getting bigger and bigger. Based on the enterprise information mining situation, many researchers study on the issue of mining association rules. Their study includes the original algorithm for optimization, such as the introduction of random sampling, parallel ideas, application of association rules for promotion and so on. Apriori algorithm is an algorithm for mining frequent item sets of Boolean association rules. Apriori algorithm is an important algorithm for customer analysis. It can be used to bank customer segmentation (Yang, 2013), market basket analysis (Balaji and Rao, 2013) and so on. Apriori algorithm is a process of analyzing a huge data from variable perspectives and summarizing it into information and knowledge. It can be help researcher to establish a model to identify frequency of customer behavior in particular geographical area with the aid of association rules (Ilayaraja and Meyyappan, 2013). Further study of apriori algorithm suggests that improved apriori algorithm can reduce large number of useless candidate item sets. Experimental conclusions show that the improved algorithm has more effective than traditional apriori algorithm (Liao *et al.*, 2013).

This study develop a method to analyze customer behavior based on apriori algorithm and use this model to discovery the rules of relevance of species which can affecting individual's decision to purchase.

METHODOLOGY

The definition of association rules: Association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. . The prototypical example of utilizing association rules mining is determining what products are found together in a basket at a checkout line at the supermarket; hence the often-used term: market basket analysis. It is intended to identify strong rules discovered in databases using different measures of interestingness. Based on the concept of strong rules, Agrawal and Srikant (1994) introduced association rules for discovering regularities between products in large-scale transaction data recorded by point systems in supermarkets. For example, the rule is: the purchases of coca cola and Hamburg are likely to buy French fires. This rule found in the sal es data of a supermarket

would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy hamburger meat.

The basic thought of apriori algorithm: The Apriori Algorithmis an influential algorithm for mining frequent item sets for boolean association rules. Apriori is a classic algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis.

In the Apriori algorithm, the basic thought of finding maximal item sets is: the algorithm needs multi step processing. The first step, statistics all elements of the project containing a set of frequency of occurrence and find out who is not less than the minimum support degree program, i.e., the one-dimensional maximum item set. The second step is starting the cycle process until no maximal item sets generated. Cyclic process: step k, according to the step k-1 generation (k-1) dimensional maximal item sets from candidate item sets dimension k, then the database search, get a candidate item set support degree, compared with the minimum support degree, in order to find K dimensional maximal item sets.

The description of apriori algorithm: Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers or details of a website frequentation). Other algorithms are designed for finding association rules in data having no transactions or having no timestamps. Each transaction is seen as a set of items.

The Apriori algorithm is described as follows:

```

C1= {candidate1-itemsets}
L1={c∈C1|c.count≥minsupport}
For(k=2,Lk-1*Φ,k++)
Ck=sc_candidate (Lk-1);
for all transactions t∈D
Ct=count_support (Ck, t)
for all candidates c∈Ct
c.count=c.count+1;
next
Lk={c∈Ck|c.count≥minsupport};
next
result set=result set∪Lk
    
```

Among them, D database; minsupport represents a given minimum support degree; result set represents all the maximal item sets.

Case study: Examples of Apriori algorithm for mining maximum frequent item sets: Assume that a large supermarket tracks sales data by Stock-keeping Unit (SKU) for each item: each item, such as "butter" or "bread", is identified by a numerical SKU. The supermarket has a database of transactions where each transaction is a set of SKUs that were bought together.

Let the database of transactions consist of the sets {1, 2, 3, 4}, {1, 2}, {2, 3, 4}, {2, 3}, {1, 2, 4}, {3, 4} and {2, 4}. We will use Apriori to determine the frequent item sets of this database. To do so, we will say that an item set is frequent if it appears in at least 3 transactions of the database.

The first step of Apriori is to count up the number of occurrences, called the support, of each member item separately, by scanning the database a first time. We obtain the following result, as shown as Table 1.

The next step is to generate a list of all pairs of the frequent items, as shown as Table 2.

The pairs {1, 2}, {2, 3}, {2, 4} and {3, 4} all meet or exceed the minimum support of 3, so they are frequent. The pairs {1, 3} and {1, 4} are not. Now, because {1, 3} and {1, 4} are not frequent, any larger set which contains {1, 3} or {1, 4} cannot be frequent. In this way, we can prune sets: we will now look for frequent triples in the database but we can already exclude all the triples that contain one of these two pairs, as shown as Table 3:

In the example, there are no frequent triplets -- {2, 3, 4} is below the minimal threshold and the other triplets were excluded because they were super sets of pairs that were already below the threshold.

We have thus determined the frequent sets of items in the database and illustrated how some items were not counted because one of their subsets was already known to be below the threshold.

Table 1: The composition of the urban competitiveness

Item	Support
{1}	3
{2}	6
{3}	4
{4}	5

Table 2: All pairs of the frequent items

Item	Support
{1,2}	3
{1,3}	1
{1,4}	2
{2,3}	3
{2,4}	4
{3,4}	3

Table 3: The minimal threshold

Item	Support
{2,3,4}	2

AN APPLICATION OF APRIORI ALGORITHM IN CUSTOMER ANALYTATION

4S stores have accumulated a large amount of customer information data during the sales process. We can find the phenomenon with the follow-up research on customer behavior that some customers to buy the car soon, while others not. Therefore, the customer information analysis is very necessary to car sales. This paper chooses customer data as mining object, mining data to discovery useful rules using apriori algorithm, investigates the relationship between transaction and customer information in order to supply accurate decision.

Data preparation: This study chooses 100 customers information from customer database, transaction database as shown in Table 4.

According to certain standards, we will convert original score into standard score, then make each data attribute into Boolean data. If the standard score is higher than average score, it is set 1, contrary to the 0. The transformed data is shown by Table 5.

Application of apriori algorithm:

- Step 1:** Scan database D, Candidate 1- set C1 and calculating the support of item sets, as shown as Table 6
- Step 2:** Pruning and del the supply less than 4, get a new Table 7
- Step 3:** Self connection operation using apriori Algorithm, attain C2, scan database D and get Candidate2- set, as shown as Table 8.

Table 4: Transaction database

TID	I1	I2	I3	I4	I5
T1	89	96	89	91	94
T2	81	91	80	91	91
T3	82	67	86	92	93
T4	89	82	78	69	88
T5	70	66	63	72	88
T6	71	64	77	73	85
T7	82	95	65	70	75
T8	72	61	61	55	70
T9	69	69	69	63	75

Note: I1: NISSAN, I2: BYD, I3: GX7, I4: SPORTAGE, I5: TIGUAN

Table 5: Transaction database (transformed)

TID	I1	I2	I3	I4	I5
T1	1	1	1	1	1
T2	1	1	1	1	1
T3	1	0	1	1	1
T4	1	1	1	0	1
T5	0	0	0	0	1
T6	0	0	1	0	1
T7	1	1	0	0	0
T8	0	0	0	0	0
T9	0	0	0	0	0

Table 6: Candidate 1- set C1

Item	Support
{I1}	5
{I2}	4
{I3}	5
{I4}	3
{I5}	6

Table 7: 1-frequent L1

Item	Support
{I1}	5
{I2}	4
{I3}	5
{I5}	6

Table 8: Candidate2- set C2

Item	Support
{I1,I2}	4
{I1,I3}	4
{I1,I5}	4
{I2,I3}	3
{I2,I5}	3
{I3,I5}	5

Table 9: 2-frequent L2

Item	Support
{I1,I2}	4
{I1,I3}	4
{I1,I5}	4
{I3,I5}	5

Table 10: Candidate3- set C3

Item	Support
{I1,I2,I3}	3
{I1,I2,I5}	3
{I1,I3,I5}	4
{I1,I2,I3,I5}	3

Table 11: 3-frequent L3

Item	Support
{I1,I3,I5}	4

Step 4: Pruning and del the supply less than 4, get a new Table 9

Step 5: In same argument, operating same process about C3, as shown as Table10 and 11

Step 6: End, frequent item L3{I1, I3, I5} is {NISSAN, GX7, TIGUAN}

CONCLUSION

As discussed above, data mining is a new technology of data management and analysis and played a good effect in the industry data analysis and decision making. The purpose of this study is to find ways to study apriori algorithm to discover more latent knowledge. This paper will introduce apriori algorithm to analyze customer behavior, mine out the relevant factors. Then using these association rules for different customers to guide the sales decision and improve the turnover rate.

The contributions of this study are: (1) using apriori algorithm to analysis customer behavior and (2) establishing an implementation model for customer relationship management.

ACKNOWLEDGMENT

This study has been supported by Henan Tobacco Corporation (Grant No. HYKJ200935) and The Education Department Henan Province (Grant No. 2010B630006).

REFERENCES

- Agrawal, R. and R. Srikant, 1994. Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases, September 12-15, 1994, San Francisco, USA., pp: 487-499.
- Ahn, H., J.J. Ahn, K.J. Oh and D.H. Kim, 2011. Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques. Expert Syst. Appl., 38: 5005-5012.
- Balaji, M. and G.S.V. Rao, 2013. An adaptive implementation case study of apriori algorithm for a retail scenario in a cloud environment. Proceedings of the 13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, May 13-16, 2013, Delft, Netherlands, pp: 625-629.
- Gu, G. and Y. He, 2010. Research on the application of data mining to customer relationship management in the mobile communication industry. Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology, July 9-11, 2010, Chengdu, China, pp: 597-599.
- Ilayaraja, M. and T. Meyyappan, 2013. Mining medical data to identify frequent diseases using Apriori algorithm. Proceedings of the International Conference on Pattern Recognition, Informatics and Mobile Engineering, February 21-22, 2013, Karaikudi, India, pp: 194-199.
- Sulianta, F., T.H. Liang and I. Atastina, 2013. Mining food industry's multidimensional data to produce association rules using apriori algorithm as a basis of business strategy. Proceedings of the International Conference of Information and Communication Technology, March 20-22, 2013, Bandung, Indonesia, pp: 176-181.
- Kim, K.J., 2011. Customer need type classification model using data mining techniques for recommender systems. World Acad. Sci. Eng. Technol., 56: 279-284.

- Liao, Z.Y., X.F. Fu and Y.G. Wang, 2013. The research of improved apriori algorithm. Proceedings of the International Conference on Information Technology and Management Innovation, November 10-11, 2012, Guangzhou, China, pp: 2179-2184.
- Yang, G.X., 2013. The research of improved apriori mining algorithm in bank customer segmentation. Proceedings of the 2nd International Conference on Optoelectronics Engineering and Materials Research, October 19-20, 2013, Zhengzhou, China, pp: 2244-2249.
- Zhang, Y.H., 2010. VIP customer segmentation based on data mining in mobile-communications industry. Proceedings of the 5th International Conference on Computer Science and Education, August 24-27, 2010, Hefei, China, pp: 156-159.