ISSN:0975-9646

Mining Frequent Itemsets for Improving the Effectiveness of Marketing and Sales

Pratima O. Fegade[#1], Prof. Dinesh D. Patil[#2]

[#1]*M.E. Student, Shri Sant Gadge Baba College of Engineering & Technology, Bhusawal, North Maharashtra University, India*

[#2]*H.O.D. Computer Science & Engineering Dept., Shri Sant Gadge Baba College of Engineering & Technology, Bhusawal, North Maharashtra University, India*

Abstract: Mention system is for finding most frequent combination of items. Main aim of this system is developing an efficient algorithm for finding frequent patterns. Initially apriori algorithm is applied on the data sets after that optimization method is applied over the rule fetched from apriori association rule mining. This system is useful in marketing and sales. Analyzing the frequently purchase items by the customer is helpful to the marketers. For the purpose of selling the product it is necessary to communicate the value of a product to the customers. Marketing is beneficial for the organization. Marketing is nothing but set of processes for creating, delivering and communicating value to the customers.

Keywords: Apriori, best combination, éclat, frequent pattern growth, genetic.

I INTRODUCTION

Analyzing customer behavior is important for improving the effectiveness of marketing and sales. Analysis of products frequently purchase by the customers is helpful for the retailers or any other company. Direct marketers could take the advantage of this information to determine what new products to offer to the customers. Main objective of this system is to find out which products should be cross-sold , which products frequently purchase by the customer, helps in learning customer behavior also helps in taking decision about product placement, pricing , promotion and profitability. For learning customer behavior it is necessary to identify which product should be frequently purchase by the customer. There are various algorithms for finding frequent item sets from large database. Apriori algorithm, frequent pattern growth algorithm and éclat algorithm are the algorithms for finding frequent item sets.

Apriori algorithm is two step processes. In the first step items which have support factor greater than or equal to minimum support are generated. In the second step all the rules having confidence factor greater than or equal to the user specified minimum confidence are generated.

In frequent pattern growth algorithm compresses the database and frequent pattern tree is generated after that compress database is divided into a set of conditional databases, each associated with one frequent item and mine each such database.

Eclat algorithm is an algorithm for mining item sets. Item sets mining allows us to find frequent pattern in the data like if a consumer buys milk, he also buys bread.

In this system first apriori algorithm is applied over a set of transactions for finding frequent item sets, after that for optimizing the result optimization method is applied on the frequent item sets which are obtain through apriori algorithm. Use of optimization method over frequent item sets optimizes the result and also time consuming method for finding frequent item sets. Finding frequent item sets is important for identification of items frequently purchase by the customer, identifying customer behavior.

II RELATED WORK

Jaishree Singh, Hari Ram, Dr. J.S. Sodhi proposed a method for improving the efficiency of apriori algorithm. They proposed an improved apriori algorithm which reduces the scanning time also reduces the redundant generation of sub-items during pruning the candidate item sets. They proposed optimized method which reduces the size of database. They used an attribute Size_Of_Transaction (SOT). SOT contains number of items in single transaction in database. Their method not only reduces the size of candidate item but also reduces the I/O spending by cutting down transaction records in the database[1].

Shruti Aggarwal, Ranveer Kaur have done a comparative study on various improved version of apriori algorithm. This paper describes methods for frequent item set mining and various improvements in the classical algorithm “Apriori” for frequent item set generation[2]. They have studied AIS, Apriori, Direct Hashing and Pruning and Partitioning algorithm. They have discuss the various parameter which are require for these algorithm. Then methods to improve the Apriori Algorithm are mentioned and improved approaches have been discussed here also[2].

Alexander Thayer and Thérèse E. Dugan suggest specific methodologies that can inform project team’s decisions at the outset of product design projects. Their method help practitioners understand how to establish testable metrics for the product experiences they are creating, quantify user’s feeling about a given product experience using those metrics and use those information to change the user experience of a product that is still being developed[3].

Xiaohui Yu, Yang Liu, Jimmy Xiangji Huang, and Aijun done online reviews for predicting sales performance. They had done review on movie domain and tackle the problem of mining review for predicting sales and performance. They proposed sentiment PLSA in which review is considered as document generated by number of hidden sentiment factor. On the basis of S-PLSA, they propose ARSA an Autoregressive Sentiment-Aware model for sales prediction[4].

In [5] Michael J. Shaw, Chandrasekar Subramaniam, Gek Woo Tan, Michael E. Welge have proposed a method to manage the marketing knowledge and support marketing decisions. They have proposed an efficient encoding technique called PC_Tree . PC_Tree is simple tree structure but powerful to capture whole of transaction by one database scan[5].

Gosta Grahne and Jianfei Zhu proposed a fast algorithm for frequent item sets mining using frequent pattern trees. In this paper they proposed a novel frequent pattern array called FP-array technique that reduces traverse of FP-trees. In their paper they have given comparison of results with other existing algorithm [6].

III EXISTING SYSTEM

In the existing system three different algorithms apriori, frequent pattern growth, éclat for frequent item set mining are given. They have developed their own best combination method for finding frequent item sets from large database.

A .Apriori Algorithm

Apriori algorithm is designed to operate on database containing transaction. Database contains set of transaction, items purchase by the customer. Apriori algorithm find out the item frequently purchase by the customer.

1. Count the number of transactions in each item.
2. Item is said to be frequent item if it is bought by the customer at least n times. This is single frequent item.
3. Starts making pair from first frequent item. Now count how many times each pair is bought together.
4. Remove all the pair with number of transaction less than n.

B. Frequent pattern growth algorithm

Frequent pattern growth algorithm allows us to generate frequent item without generation of candidate items.

1 Built a compact data structure called frequent pattern tree. It requires two pass over the data –set.

Compact data structure can be constructed on the basis of following observation:

a. Only frequent item play important role in frequent pattern mining, so perform one scan over database to find frequent one item.
b. Storing the frequent one item in compact data structure avoids repeated scanning over the database.
c. If multiple transaction share single frequent item set, they can be merged into one with number of occurrences registered as count.
d. If two transactions share a common prefix the share part can be merged using prefix structure as long as count is registered properly.

2 Extract frequent item directly from FP-tree.

C. Éclat algorithm

Eclat algorithm is used for item set mining which allows us to find frequent item sets. Eclat algorithm use tidset intersection to calculate the support of a candidate item set avoiding the generation of subsets that does not exists in the prefix tree.

In éclat algorithm initial call use single items with their tidset. In each recursive call this algorithm intersects each itemset-tidset pair with all other pair to generate new candidate. If the new candidate is frequent it is added to frequent set.

In the existing system Saurabh Malgaonkar, Sakshi Surve and Tejas Hirave developed best combination method for finding frequent item set. For the extraction of interesting pattern the analyst has to perform database connectivity. After that mention a specific condition and view the best combination. Much transaction may contain the same set of items. Many transactions that do not contain same set of item may also contain subset of the items that are common in them. Transaction tree is constructed to group such transaction.

The transaction tree is further truncated by using logical reduction scheme. Thus improving performance and efficiency of mining mechanism. At the time of constructing the tree counts at each node are stored in an array.

Their method require to construct ITable and Tlink

ITable: It stores each individual frequent item with support of each item.

TLink: It represent all the transaction of the database containing the frequent item.

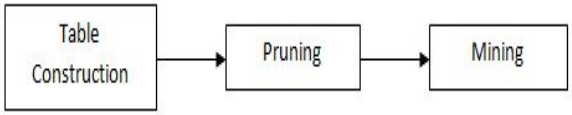

Fig. 1. Basic Processing Cycle.

The data needs to be cleaned before it is Actually to be mined.

1-frequent items are then identified from ITable. Infrequent items are removed from TLink. After this step data is cleaned as require for actual mining procedure.

In the existing system best combination method of mining is designed and implemented which is as follows:

1. Database is scanned to find frequent 1-items which are then stored in ITable. All the entries in ITable are stored in ascending order and then items are mapped to new identifiers that are ascending sequence of integer.
2. 1-Frequent items are extracted from the database. Pointer is maintained to the subset of each item set. This pointer helps for mining the frequent item set corresponding to the path ending at nodes of the corresponding subset.

Mining Process:

All the frequent item sets of two or more items are mined in a recursive manner.

D. Actual Mining Working Procedure:

1 First construct ITable and the reduced tree.

2 Construct newly mapped entries.

3 Scan newly mapped entries.

3.1 For each newly form entries

3.1.1 Scan the item set entry in each of the table entries.

3.1.2 Use the node pointer of the tree.

3.1.3 Traverse the tree depending upon the item position.

3.1.3.1 If itemset/item found -> Record and increment its count by its frequency of occurrence; Reset search for next entry; goto step 3.1.1

3.1.3.2 Else -> Shrink itemset area by 1 item count from left ; Reset search for this new entry; Goto step 3.1.1

3.1.4 If all entries traversed, no entries left to scan; Goto step 4.

4. Stop.

After all the entries have been covered then they are remapped into their original values and the frequency of per item set.

In the existing method it is necessary to construct tree which take too much time and also occupies more memory. For improving the results in proposed system apriori method is used for finding the frequent item sets after that optimization method will be applied on the result for optimizing and improving the result.

IV. PROPOSED METHOD

In the proposed method apriori algorithm and optimization method is used. Advantages of proposed method over existing method are proposed method require less time, acquire less memory as compare to existing method.

First apriori algorithm is applied over dataset as describe in introduction after that optimization method (genetic algorithm) applied on the frequent items obtained through apriori algorithm.

Genetic algorithm is an evolutionary algorithm. Genetic algorithm generates solution to optimization problems using technique such as inheritance, selection, crossover and mutation. Genetic algorithm is useful to find approximate solutions to optimization and search problems.

The proposed method is as follows:

1. First load sample of records from the database that fits in the memory.
2. Apply apriori algorithm for finding out frequent item set.
3. The output set contains association rule.
4. Input the termination condition of genetic algorithm.
5. Represent each frequent item set as binary string.
6. Apply the crossover and mutation on the selected members to generate the association rules.
7. Find the fitness function for each rule.
8. If (fitness function> min confidence)
9. If the desired number of generations is not completed, Apply Apriori algorithm to find the frequent item sets with the minimum support again.

V. COMPARISON OF EXISTING SYSTEMS RESULT WITH EXPECTED RESULT OF PROPOSED SYSTEM:

The result of existing method contains more frequent items. In proposed method use of optimization method means genetic algorithm will be able to generate more optimize result which will contain less frequent items as compare to existing method results.

V. CONCLUSIONS

In the existing system best combination method is used for finding the items frequently purchase by the customer. In the proposed method first apriori algorithm is applied over the database after that genetic algorithm is applied over the results of apriori algorithm. The proposed method is more beneficial than existing method. Proposed method require less time, occupy less memory for finding frequent item sets and generate optimize results.

REFERENCES

[1]Jaishree Singh, Hari Ram, Dr. J.S. Sodhi, “Improving the Efficiency of Apriori algorithm using transaction reduction”, International Journal of Scientific and Research Publications, Volume 3, Issue 1, January 2013.

[2] Shruti Aggarwal, Ranveer Kaur, “Comparative Study of Various Improved Version of Apriori algorithm”, International Journal of Engineering Trends and Technology (IJETT) - Volume4 Issue4- April 2013.

[3] Alexander Thayer, Thérèse E. Dugan, “Achieving Design Enlightenment: Defining a New User Experience Measurement Framework”, IEEE 2009.

[4] Xiaohui Yu, Yang Liu, Jimmy Xiangji Huang and Aijun An, “Mining Online Review for Predicting sales Performance: A Case Study in Movie Domain”, ieee transactions on knowledge and data engineering, vol. 24, no. 4, april 2012

[5] Nadimi Shahraki M.H., N. Mustapha, M. N. B. Sulaiman, and A. B. Mamat, A New Method for Mining Maximal Frequent Item sets, International IEEE Symposium on Information Technology, Malaysia, pp.309-312,2008.

[6] Go’sta Grahne and Jianfei Zhu, “Fast Algorithms for Frequent Itemset Mining Using FP-Trees”, ieee transactions on knowledge and data engineering, vol. 17, no. 10, october 2005.

[7] Ando T, Bayesian State Space Modeling Approach for Measuring the Effectiveness of Marketing Activities and Baseline Sales from POS Data, 6th international conference on Data Mining, December 2006, pp 21-32.

[8] Mishra N, Whole Sale Price Index, Financial Indicators and its Effect on Market and Different Sector, 5th international conference on Computational intelligence and Communication Networks, September 2013.

[9] Akbarian R, TavaKoli M, Fard, Establishment Framework of Mobile Marketing in Iran, 7th international conference in developing countries”; with focus on e-security, April 2013, pp. 1-2.

[10] Charanjeet Kaur, Association Rule Mining Algorithm using Apriori Algorithm: A Survey, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 2, Issue 6, June 2013, pp-2081-84.