Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 7, July 2014, pg.41 – 46

RESEARCH ARTICLE

Improving the Effectiveness of Marketing and Sales using Genetic Algorithm

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

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

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

fegadepratima@gmail.com [1]

dineshonly@gmail.com [2]

Abstract: The mention system is useful tool for discovering the customer purchasing pattern. Main aim of this system is to discover the items frequently purchase by the customer. This system is helpful for marketing and improving the sales. This system helps in making the decision of profit, placement, pricing and promotion of the products. Examining product frequently purchase by the customer help in guessing improvement require in products. In this system first apriori algorithm is applied over the item sets after that optimization method is applied over the result of apriori algorithm.

I. INTRODUCTION

Tracking the items frequently purchase by the customers can improve the effectiveness of marketing and sales. Marketing is nothing but the activity of identifying what market wants to buy. It also tracks the sales using the clustering technique, it generally clusters up the data in the timely fashion know as Time Series [1] where it works and merges the revenue thus collected in different seasons. It analyses the combinations of products the customers buy frequently. This system can be used by shopkeepers to identify the item frequently purchase by the customer. There are various algorithms for finding frequent item sets such as apriori algorithm, frequent pattern growth algorithm, éclat algorithm. Association rules are the main technique for finding frequent item sets in data mining. Association rule is a popular method for finding interesting relation between variables in the data.

Definition of Association Rule: Let $\{i_1, i_2 \ldots\ldots i_n\}$ be a set of n items. Let $\{t_1, t_2 \ldots t_m\}$ be a set of transaction called database. Each transaction in the database has unique id and contain subset of items in I. A rule is defined as X => Y where X, Y $\subseteq$ I and $X \cap Y = \emptyset$. The set X and Y called antecedent and consequent of the rule respectively.
Support: Support(X) is defined as the proportion of transaction in the data set which contains the item sets.
Confidence: Confidence of a rule conf(X=>Y) is defined as Supp(X U Y) / Supp(X). For example the rule {onion, potatoes} => {chilly} has confidence of 0.2/0.2= 1.0 in the database, which means for 100% of transaction containing onion and potatoes the rule is correct.

For example the rule {onion, potatoes} => {chilly} found in the sales data would indicate that if customer buys onion and potatoes, he or she is likely to also buy chilly. Such information is very helpful for product placement and promotional pricing. Each association rule satisfies the user defined minimum support and a user specified minimum confidence at the same time.
Generation of association rule is two step processes:

1. Minimum support is applied to find all frequent item sets in the database.
2. In the second phase these frequent item sets and minimum confidence constraint are used to form rules

The first step is easy to perform as compare to second step. Second require more attention because since it requires searching all the possible item sets.
Many algorithms for generating association rule were presented over time:

Some well-known algorithms are apriori, FP growth and éclat

Apriori: Apriori algorithm is a best known algorithm in data mining to mine association rules. Apriori algorithm uses breadth-first-strategy to count the support of item sets and also uses candidate generation function. In apriori algorithm candidate items are generated.

FP Growth: FP growth is a frequent pattern growth algorithm. In the FP growth algorithm compact data structure called FP tree is constructed. This requires two pass over the data sets. After that extracts frequent items directly from FP tree. One advantage of FP growth algorithm is completeness. It preserves the complete information for frequent pattern mining. Never break the long pattern of any transaction. Frequent pattern growth algorithm reduces the irrelevant info of infrequent items.

Eclat Algorithm: Eclat algorithm is used to perform item set mining. Item set mining finds frequent pattern in the data if consumer buys milk, he also buys bread. This type of rule is nothing but the association rule and is use in many application domains. Eclat algorithm works recursively. Each recursive call compare item transaction id with transaction id of all other items or pair of items to generate candidate item. If newly generated item is frequent it is added to frequent item sets.

II. RELATED WORK

Jaishree Singh, Hari Ram, and Dr. J.S. Sodhi in [1] improved the efficiency of apriori algorithm by transaction reduction. Apriori algorithm is a classical algorithm of association rule mining [1]. Their improved algorithm reduces the scanning time by cutting down unnecessary transaction record. Their algorithm also reduces candidate item sets. Main advantage of this algorithm is to reduce candidate item sets and gives optimized solution. Main idea behind this improved algorithm is to reduce transaction. Classical apriori algorithm generates large number candidate item sets. Due to the generation of candidate item set I/O cost of classical apriori algorithm is more.

In [2] Michael J. Shaw, Chandrasekar Subramaniam, Gek Woo Tan, Michael E. Welge have proposed a technique to manage the marketing knowledge and support marketing decisions. In this paper they have presented the model for customer retention which accounts for the dynamics of the today's market. Their methodology helps for enhancing the customer relationship management. Their methodology introduces an efficient database encoding technique a novel tree structure called PC-Tree and PC-Miner algorithm. This paper shows how data mining can be integrated into a marketing knowledge management framework [2]. The database encoding technique utilizes Prime number characteristics and transforms each transaction into a positive integer that has all properties of its items. The PC_Tree is a simple tree structure but yet powerful to capture whole of transactions by one database scan. The PC_Miner algorithm traverses the PC_Tree and builds the gcd (greatest common divisor) set of its nodes to mine maximal frequent itemsets. Experiments verify the efficiency and advantages of the proposed method [2].

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 [3].

III. EXISTING SYSTEM

There are many algorithms for generating frequent items like apriori, frequent pattern growth and éclat. The base paper which I have used for developing this system includes best combination method of apriori, frequent pattern growth and ecalt algorithm.

1. Scan the database to find all frequent items. Store all frequent items in the table. All the items are stored in the in table in the frequency of their ascending orders after that those items are mapped to new integers.

2. From the step 1 all frequent 1 items are extracted mapped them to new identifier and the transaction are inserted into the reduced tree. A pointer is maintained.

3. In the mining process all the frequent item sets of two or three items are mined.

A. Limitations of Existing System

In the existing system multiple scanning of database is required and also required to construct a tree. Construction of tree occupies more memory. In the propose methodology construction of a tree is not required so save memory and also application of genetic algorithm optimizes the solution.

IV. PROPOSED METHODOLOGY

In the existing system Saurabh Malgaonkar, Sakshi Surve and Tejas Hirave have developed their own best combination method. Three different algorithms apriori, frequent pattern growth and eclat algorithms are given. Authors have proposed the combination method of these three algorithms. But in this system first apriori algorithm is applied on the database to find out frequent item sets after that for optimizing the solution genetic algorithm applied over the results of apriori algorithm. In genetic algorithm candidate item sets are not generated also it optimizes the solution.

1. Start

2. Load sample of records from the database that fits in the memory.

3. Scan the transaction database to get the support S of each item.

4. If S> = min_support

- Add S to frequent 1-itemsets L1

5. Use LK-1 join Lk-1 to generate a set of candidate K-itemsets

6. Scan the transaction database to get the support S of each candidate K-itemsets

7. If S>= min_suppport

- Add to K-frequent itemsets

8. If generated set= NULL then Stop

- Else go to step 5

9. For each frequent item set L, generate all nonempty subset of L

10.For each nonempty subset of L find confidence C of

11. if C>= min_confidence add to Strong Rule.

12. Input the termination condition of genetic algorithm

- A solution is found that satisfies minimum criteria
- Fixed number of generations reached

13. Represent each frequent itemsets as a binary string.

14. Apply crossover and mutation on the selected member to generate te association rules.

15. Find the fitness function of each rule

Fitness function = W1*AC(F) + W2 * CS(F)/ W1+ W2

- Where AC(F)= Support(F) / max(Support(z element of F))
- max(Support(z element of F)) is the support of the item with the highest support in F
- and CS(F)= (1-v(F)) / (1- E[v(F)]) * E[v(F)] / v(F)
- where v(F) is the violation rate
- Violation rate is defined as the fraction of transaction which contains some of the item in an itemsets but not all.
- E[] is the expected value for independent item. W1 and W2 are weight assign by user

16. if (fitness function > min confidence)

17. If the desired number of generations is not completed, Apply Apriori algorithm to find the frequent item sets with the minimum support again.

V. SYSTEM IMPLEMENTATION

A. *System Requirements*

Hardware Requirements

- Processor Core 2 Duo.
- Ram 1 GB.
- Keyboard Standard 102 keys.

Software Requirements

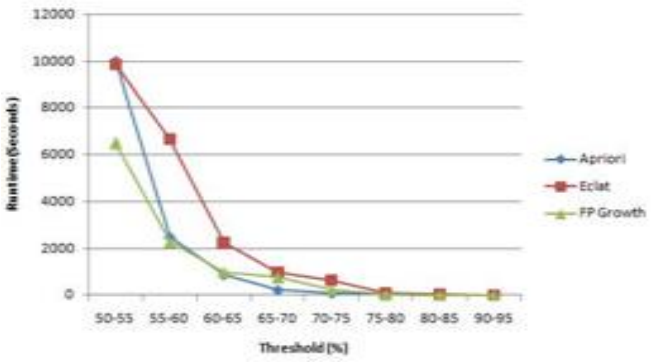
- Front end : java
- Operating system : windows xp

In the base paper they have check the performance of their system by applying apriori, frequent pattern growth and éclat algorithm and best combination method. Support range is from 50 to 95. They prove the best combination method is better as compare to apriori, frequent pattern growth and éclat algorithm.

Support Level	Apriori	Eclat	FP Growth
50-55	10000	9880	6520
55-60	2490	6670	2280
60-65	870	2230	990
65-70	220	970	760
70-75	82	650	220
75-80	24	98	22
80-85	9	46	10
90-95	6	8	9

Runtime evaluation of apriori, eclat & fp growth algorithm

Above table shows support range from 50 to 95 and runtime taken by the apriori, frequent pattern growth and éclat algorithm.

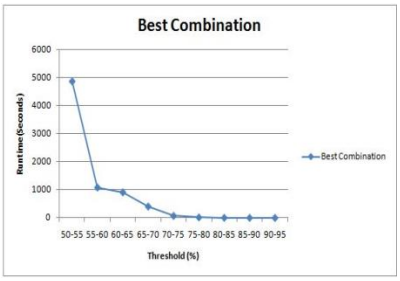


Performance evaluation of apriori, frequent pattern growth and éclat algorithm

From the above graph, apriori and éclat algorithm takes 10,000 seconds and frequent pattern growth algorithm takes 6500 seconds for finding frequent item sets.

Support Level	Best Combination
50-55	4890
55-60	1090
60-65	412
65-70	76
70-75	23
75-80	8
80-85	6
90-95	4

Runtime evaluation of best combination method

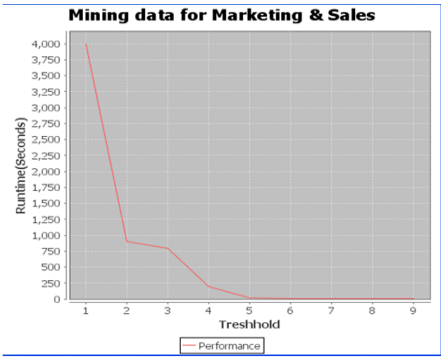


Performance evaluation of Best combination method

From the above graph it is clear that best combination method takes 5000 seconds for finding frequent item set from the same database. By taking the same support ranges I have check the performance of Propose methodology which is as follows.

Support Range	Propose methodology
50-55	20
55-60	50
60-65	100
65-70	150
70-75	200
75-80	800
80-85	900
85-90	4000

Runtime evaluation of best combination method



Performance evaluation of Propose methodology

From the above graph propose methodology takes 4000 seconds for finding frequent item set from the database. Thus propose methodology is time saving as compare to apriori, frequent pattern growth, éclat and best combination method.

CONCLUSION

Advantage of this system is that it is time saving method for finding frequent item sets as compare to apriori, frequent pattern growth and éclat algorithm. Use of genetic algorithm optimizes the solution also finding the frequent item sets by using genetic algorithm do not generate candidate item sets. This system helps to find out which items are frequently purchase by the customers also helps marketers for taking appropriate decision, helps shopkeepers for making appropriate decision about products placement, pricing and promotion.

REFERENCES

[1] Jaishree Singh, Hari Ram, Dr. J.S. Sodhi, " Improving Efficiency of Apriori Algorithm Using Transaction Reduction , International Journal of Scientific and Research Publication"s, Volume 3, Issue 1, January 2013.
[2]Nadimi Shahraki M.H., N. Mustapha, M. N. B. Sulaiman, and A. B. Mamat, " A New Method for Mining Maximal Frequent Item set"s, International IEEE Symposium on Information Technology, Malaysia, pp.309-312,2008.
[3] Xiaohui Yu, Yang Liu, Jimmy Xiangji Huang, and Aijun, "Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain", ieee transactions on knowledge and data engineering, vol. 24, no. 4, april 2012.
[4]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.
[5]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.
[6]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.
[7] Tao Ji, ChongQing," Effective Integration of Onsite Customer and Sales and Marketing Management by Standardized Coding and Centralized Platform", 20th international conference and exhibition on electricity distribution, June 2009, pp 1-4.
[8] 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.
[9]Jian Liu, Yanqing Wang, "Research on Internet Marketing Relationship Mode"l, 9th international conference on Fuzzy System and Knowledge Discovery, May 2012, pp. 826-30.
[10]Yu, Lin A," The Design and Implementation of a Search Engine Marketing Management System Based on Service Oriented Architecture Platform", IEEE international Conference on e-Business Engineering" October 2007, pp. 513-19.

