ISSN No. 0976-5697

Volume 4, No. 3, March 2013 (Special Issue)

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

Extended Realization of Associative Property for Data Mining using Apriori Optimization Technique for Frequency Pattern Generation

Arpita Lodha* and Charu Kavadia
M Tech Scholar, Department of Computer Science
Arya College of Engineering & IT
Jaipur, India
lodhaarpi2803@gmail.com*, charukavadia@gmail.com

Abstract: Mining Association rules in transactional or relational databases have recently attracted a lot of attention in databases communities. Associative Rule Mining as defined in [13] is a popular and well researched method for discovering interesting relationship between various items involved in large databases. Introduced association rules for discovering regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems. Our objective is to find the Frequency Pattern based upon Apriori optimization technique. An Apriori algorithm is the most commonly used Association Rule Mining. This algorithm somehow has limitation and thus, giving the opportunity to do this work. This paper introduces a new way in which the Apriori algorithm can be tested with large number of transactions and item sets.. The modified algorithm introduces factors such as set size and set size frequency which in turn are being used to eliminate non significant candidate keys. With the use of these factors, the modified algorithm introduces a more efficient and effective way of minimizing candidate keys.

Keywords: Data Mining, Apriori algorithm, Frequent items, Set size, Set size frequency, Minimum support.

I. INTRODUCTION

Data Mining methodologies have been extensively adopted in various domains such as database marketing, credit scoring, fraud detection, pattern recognition, machine learning, data visualization, and business success. Industrial process optimization and control are also increasingly using data mining methods. Data mining is the identification or extraction of relationships and patterns from data using computational algorithms to reduce, model, understand or analyze data.

The automated process of converting raw data into useful information by which intelligent computer system sort data with little or no manual interaction, to search patterns or to predict market trends.

One of the major aspects of data mining is the Association Analysis or association Rule mining. It consists of two steps: [10]

a. First, minimum support is applied to find all frequent itemsets in a database.
b. Second, these frequent itemsets and the minimum confidence constraint are used to form rules.

It relates to the association of items wherein for every occurrence of A, there exists an occurrence of B. This mining is more applicable in the market basket analysis [9]. That application is helpful to the customers that buy certain items. That for every item that they bought, what would be the possible item/s coupled with the purchased item. Apriori algorithm is the most widely used association rule mining algorithm [9]. However, several limitations have been discovered in this method [7] such as:

a) Several iterations are needed for data mining.
b) Usually generates item set which are irrelevant.
c) Difficulties in finding unusual event.

II. TERMS RELATED TO ASSOCIATION RULE MINING

A. ***Support:*** An itemset containing i items is called an i-itemset.The percentage of transactions that contain an itemset is called itemset's support.

B. ***Confidence:*** Confidence is the ratio of the number of transaction that include all items in the consequent as well as antecedent (namely,the support)to the number of transactions that include all items in the intecedent.

C. ***Supportcount/Frequency:*** The occurrence frequency of an itemset is the number of transactions that contain the itemset.

D. ***Frequent Sets:*** Let in transaction set T ,α is the user specified minimum support and A={i1,i2,i3……in}is the set of items.An itemset M ⊑ A is said to be frequent set in T with respect to α if s(M)t ≥α

E. ***Maximal Frequent Sets:*** A frequent set is maximal if it is a frequent set and no superset of this is a frequent set.

III. RELATED WORK

The Apriori Algorithm:The apriori algorithm is one of the most widely used tools for association rule mining. The steps followed in apriori algorithm is as follows:

A. In first pass, algorithm simply counts item occurrences to determine frequent itemsets. Therefore all singleton items are the candidates and the items that has the support value less than threshold is eliminated from the candidate item-list,Call this list L1.

B. The singleton items are combined to form two member candidate item-sets (call it L2)and support values of these candidates are determined by

CONFERENCE PAPER II International Conference on
"Advance Computing and Creating Entrepreneurs (ACCE2013)"
On 19-20 Feb 2013
Organized by
2nd SIG-WNs, Div IV & Udaipur Chapter , CSI , IEEE Computer Society Chapter India Council ,
IEEE Student Chapter Geetanjali Institute of Technical Studies, Udaipur, Rajasthan, India

scanning the data sets again. Also,candidates with support value higher than threshold are considered.

C. In the next pass,create three member candidate itemsets(call it L3) and the process is repeated again.When all frequent itemset are accounted then the process stops.

D. The Frequent itemset are then used to generate association rules which have confidence values greater than or equal to the threshod.

E. It first creates the rules for frequent itemsets and then rules for subsets are created recursively.

Based on this work, the probability model [8] was calculated with two factors:

a. Success rate: the probability that the set is a success i.e. it is a maximal frequent pattern set.
b. Failure rate: the probability that the set is a failure i.e. it is a not maximal frequent pattern set.

IV. MODIFIED APRIORI ALGORITHM

The improved algorithm for Apriori takes for the set size which is the number of items per transaction and set size frequency which is the number of transactions that have at least "set size" items.

Given: Minimum Support =4
Set Size =Number of items per transactions
Set Size Frequency= Number of transactions that have at least set size items

Table 1: Initial data set with set size

Trans action ID	Items	Set Size
T1	Computer, Printer, Scanner, Anti-virus, Pendrive,Speaker, Head Phone	7
T2	Computer, Printer , Speaker Head Phone ,Webcam, Cartridges	6
T3	Computer, Printer , Scanner , Anti-virus	4
T4	Computer, Speaker , Head Phone	3
T5	Computer, Printer , Scanner , Speaker ,RAM, Hard-Disk	6
T6	Computer, Printer, Scanner , Speaker ,RAM, Hard-Disk	6
T7	Speaker ,RAM, Hard-Disk ,Data Card, Mouse-Pad	5

Table 2: Set size Frequency less than min support Value

Set Size	Set size Frequency
7	1
6	4
4	6
3	7
5	5

Step 1: Remove Frequency with less than Min Support Value

Table 3: Data set with eliminated frequency

Transactio n ID	Items	Set Size
T1	Computer , Printer, Scanner, Speaker,	4
T2	Computer , Printer , Speaker	3
T3	Computer, Printer, Scanner	3
T4	Speaker	1
T5	Computer, Printer, Scanner, Speaker	4
T6	Computer, Printer, Scanner, Speaker	4
T7	Speaker	1

Step 2: Determine initial set size to build –get the highest set size whose frequency is greater than or equal to minimum support (set size 4)

Table 4: Set Size with meeting min support

Set Size	Set size Frequency
4	3
3	2
1	2

Step 3: Get list of items from transactions with set size is greater than or equal to set size determined in step2

Table 5: Transactions items meeting Set Size

Transact ion ID	Items	Set Size
T1	Computer, Printer, Scanner, Speaker,	4
T5	Computer, Printer, Scanner, Speaker	4
T6	Computer, Printer, Scanner, Speaker	4

Unique Items = {Computer, Printer, Scanner, Speaker,}
Step: 4 Create combinations with size 4 (as determined in step 2) Determine the TID whose set size>=result of #1.Count the combinations frequency in the database { Computer, Printer, Scanner, Speaker,}=4
Step: 5 Remove combinations with frequency less than min support
Step 6 If list of combinations in #5 is grater than 1 use the next available set size as determine in step#2.Go back to step #3
NOTE: if the list of set sizes in step #2 is exhausted, continue moving to the next lower value until you reach set size=3.

V. CONCLUSION

Our intention to improvise the performance of the database in using the Apriori algorithm, also we have optimized our result set in the light of its frequent-Pattern analysis. So this way we also make a predication base analysis on behalf of the database.

VI. FUTURE WORK

We are looking forward for incorporating this Apriori concept for optimization of frequent-Pattern generation with ratio base analysis.

VII. REFERENCES

[1]. D. Kerana Hanirex, and M.A. Dorai Rangaswamy. 2011. Effi-cient Algorithm for Mining Frequent Itemsets using Clustering Techniques. International Journal on Computer Science and Engi-neering (IJCSE) Vol. 3 No. 3 March 2011

[2]. E. Ramaraj and N. Vankatesan, ― Bit Stream Mask Search Algo-rithm in Frequent Itemset Mining, European Journal of Scien-tific Reasearch, Vol. 27 No. 2 (2009), pp. 286-297

[3]. J. Han, M. Kamber, ―Data Mining: Concepts and Techniques, Morgan Kauffman, San Francisco, 2000

[4]. J.5 Han, J. Pei and Y.Yin, ―Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach, In Proceedings ACM-SIGMOD International Conference Man-agement of Data (SIG MOD'04), pages 53-87, 2004

[5]. J. Hossen, A. Rahman, K. Samsudin, F. Rokhani,

CONFERENCE PAPER II International Conference on
"Advance Computing and Creating Entrepreneurs (ACCE2013)"
On 19-20 Feb 2013
Organized by
2nd SIG-WNs, Div IV & Udaipur Chapter , CSI , IEEE Computer Society Chapter India Council ,
IEEE Student Chapter Geetanjali Institute of Technical Studies, Udaipur, Rajasthan, India

S. Sayeed, and R. Hasan, ― A Novel Modified Adaptive Fuzzy Inference Engine and Its Application to Pattern Classification, World Academy of Science, Engineering and Technology 80, 2011

[6]. M. H. Marghny and A.A. Mitwaly, ―Fast Algorithm for Mining Association Rules, In proceedings of the First ICGST Interna-tional Conference on Artificial Intelligence and Machine ELearning AIML05, pages 36-40, Dec. 2005

[7]. Mamta Dhanda, ―An Approach to Extract Efficient Frequent Patterns from transactional database, International Journal of Engineering Science and Technology (IJEST), Vol.3 No.7, July 2011, pp. 5652-5658

[8]. P.Purdon, D. Gucht and D. Groth, Average Case Performance of the Apriori Algorithm, Society for Industrial and Applied Mathematics (2004) Vol. 33 No.5 pp. 1223-1260

[9]. R. Agrawal and R. Srikant, ―Fast algorithms for mining associa- tion rules, in Proceedings of the 20th VLDB Conference, 1994, pp. 487-499

[10]. R. Agrawal, T. Imielinski and A. Swami, ―Mining association rules between sets of items in large databases. SIGMOD'93, 207-216, Washington, D.C.

[11]. S. Murali Krishna and S. Durgs Bhavani, ― An Efficient Ap-proach for Text Clustering Based on Frequent Itemsets, Euro-pean Journal of Scientific Research, Vol. 42 No. 3 (2010), pp.385-396

[12]. S. Prakash and R.M.S. Parvathi, ―An Enhanced Scaling Apriori for Association Rule Mining Efficiency,European Journal of Scientific Research, Vol. 39 No. 2 (2010), pages 257-264

[13]. S Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining association rules between sets of items in large databases". Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93. pp. 207. doi:10.1145/170035.170072. ISBN 0897915925