|| Volume 5 || Issue 1 || January 2020 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

MINING OF INTERESTING POSITIVE AND NEGATIVE ITEMSETS USING OPTIMIZED ASSOCIATION RULE

Shubhangi S. Mandwale

Department of Computer Science and Engineering, CSMSS,Chh. Shahu College of Engineering, Aurangabad, India
shubhangimandwale@gmail.com

Abstract- Association rule mining plays a very important role in various data mining process. The fundamental idea of association rules is to mine the interesting (positive) frequent patterns from a transaction database. The point of this study is to build up the new model for mining interesting negative and positive association rules out of a transactional data set. For the mining of rule mining a variety of algorithm are used such as Apriori algorithm and multiple level minimum support based algorithm. Some algorithm is wonder performance but produces negative association rule and also has the problem of multi-scan database. MLMS-GA association rule mining based algorithm is proposed. In this mthod a multi-level multiple support of data table as 0 and 1 is used. The divided process reduces the scanning time of database and get reduced set of association rules.

Keywords: Data Mining, Association Rule Mining, Positive Rule Mining, Negative Rule Mining, Genetic Algorithm, Apriori Algorithm**.**

I INTRODUCTION

Association rule mining approach has been utilized in to market domain and particular problem has been studied, the management of some conditions of a shopping mall, and an architecture that makes it possible to produce agents capable of adapting the association rules has been used [1]. Data mining introduced to extracting knowledge e from large quantity of data. Interesting association rules that can be discovered among various large set of data items by association rule mining. The searching of interesting relationship among huge amount of business transaction records can help in many business decisions making process. Association rules mining is an important task in the area of data mining, and frequent item set mining is a key step of many algorithms for association rules mining. There had been **i**mmeasurable work done for mining of association rules. When the dataset are large, the rules generated may be very large, but some of them are not interesting to the users mean unnecessary for the user, so, it is universal to set some parameters to reduce the numbers of rules generated in the process of mining, support and confidence are two common parameters. An association rule R is referred to as in the form A → B, where A, B are disjoint subsets of the attribute set I. The support for the R is the number of database records that contain A∪B (often expressed as a proportion of the total number of records).

Two properties support and confidence provide the empirical basis for derivation of the inference expressed in the rule, and a measure of the interest in the rule. The support for a rule expresses the number of records within which the association may be observed, while the confidence expresses this as a proportion of the instances of the antecedent of the rule. In practical analysis, it is usual to regard these rules as “interesting” only if the support and confidence exceed some threshold values. Hence this matter is developed as a research for all association rules within the database for which the required support and confidence levels are obtained. Note that the confidence in a rule can be determined immediately once the relevant support values for the rule and its antecedent are computed. Thus the problem essentially resolves to a search for all subsets of I for which the support exceeds the required threshold. Such subsets are referred to as “large”, “frequent” or “interesting” sets. [2]

II RELATED WORK

This section describes some related work for mining of negative and positive association rule mining.An author proposes an algorithm PNAR_MLMS for mining interesting negative and positive association rules out of a transactional data set. This model propose a replaced approach (PNAR_MLMS) for mining both negative and positive association rules from the interesting frequent and infrequent item sets mined by the MLMS model. The experimental results show that the PNAR_MLMS model provides automatically better results than the previous model. The purpose of association rule mining is to search out certain associations between groups of items in a database [3].

In this paper author has present an algorithm for mining association rules with multiple level constraints, the proposed algorithm simultaneously copes with two different kinds of constraints, it consists of three phases, first, the frequent 1-itemset are generated, second, they exploit the properties of the given constraints to prune search space or save constraint checking in the conditional databases. Third, for every item set potentially possible to satisfy the constraint, authors generate its conditional database and perform the three phases in the conditional database recursively.[4]

In this paper authors have proposed Mining Optimized Association Rules Algorithm (MOAR) which maintains two populations: the internal population, and a Pareto−store. This algorithm consolidates partial ordering in the search mechanism for rules. The algorithm iteratively evaluates a solution by first decoding the bit string into a rule which is compared to the data in the database. For all records that cover a given consequent the values are compared against the consequent to calculate the supports of the antecedent and the consequent. After the dataset scan, the measures of strength used as the objectives are calculated for each rule. This algorithm adopts the Pareto dominance approach pro-posed in for maintaining multiple stable niches. This ensures that the individuals in the Pareto store are uniformly distributed near the Pareto−optimal front. To determine the dominance of an individual all individuals in the internal population and the Pareto store are evaluated [5].

As stated in this paper to correlation and dual confidence measures association rules are classified in to positive and negative association rules ,but one drawback of dual confidence, is if less confidence would be a lot of rules even produce large number of contradict rules (¬C→¬D), if greater confidence may missed useful positive association rules[6].

The author specifies an effective method for Efficient Mining of the Generalized Negative Association Rules (GNAR) is produced interesting negative rules ,this approach could speed up execution time efficiently through the domain taxonomy tree and extract interesting rules easily, advantage of taxonomy tree is to eliminate large number of useless transaction [7].

The author presents an algorithm positive and negative association rule (PNAR) for mining both positive and negative association rules in databases. The algorithm enhances the traditional association rules to have negative association rules. When mining negative association rules, the author uses the same minimum support threshold to mine frequent negative itemsets. The algorithm can find all valid association rules quickly and overcome some limitations of the previous mining methods [8].

Mining association rules using multiple support confidence values and several studies have been addressed the issue of mining association rules using Multiple Level Minimum Supports [9].

III PROPOSED WORK

In this paper proposed a novel algorithm for optimization of association rule mining, the proposed algorithm resolves the problem of negative rule generation and also optimized the process of rule generation. Negative association rule mining could be a huge challenge extreme dataset. In the generation of valid rules association existing algorithm or technique generate a series of negative rules, which generated rule affected a performance of association rule mining. In the method of rule generation numerous multi objective associations rule mining algorithm is recommended however of these don't seem to be solve.

In this Paper we planned MLMS-GA of association rule mining algorithm. In this specified algorithm we used MLMS used for multi level minimum support for constraints validation. The scanning of database separated into multiple levels as frequent level and infrequent level of data according to MLMS. The frequent data logically assigned 1 and infrequent data logically appointed 0 for MLMS technique. The divided method reduces the uninteresting item in given info. The proposed algorithm could be a combination of MLMS and min-max algorithm on this used level weight for the separation of frequent and infrequent item. The weight value serve as Support length key is a vector value given by the transaction data set. The support value passes as a vector for locating a close to level between MLMS candidates key. After finding a MLMS candidate key the nearest level separated into two levels, one level take a higher odder value and another level gain infrequent minimum support value for rule generation process. The process of choice of level additionally reduces the passes of data set. After finding a level of lower and enhance of given support value, compare the value of level weight vector. Here level length vector work as fitness performs for choice method of min-max algorithm. Here we represent steps of method of algorithm step by step and at last draw a flow chart of complete method.

Steps of algorithm (MLMS-GA)

Scanning of database used flowing steps

Some standard notation of pseudo code of algorithm such as D dataset, K level MLMS, Ls generation candidate

K = MLMS dataset (D)
n = Number of multiple level block
For i = 1 to n loop=
Scan_k (Ki k)
Li = gen__itemsets (ki)
For (i = 2; Lj i ≠ f, j = 1,2....,n; i++)
Ci G =
j = 1,2,...nLij
End;
For i = 1 to n
scan_kmap (ki K)
For all items C
CG generate block (C, ki)
End;
LG = {c CG|}

2. Generate multiple support value for selection process for all transaction LG do
generate count table TC

|| Volume 5 || Issue 1 || January 2020 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

L1 =(frequent 1-itemsets);
C2 =L1 8 L1;
L2 = sup(c)=MinSupNum;
For(k=3;Lk-1 ≠Ø ;k++)do begin
For (j=k;j=m;j++)do
Generate CIVij k-1;
Ck =candidate_gen(Lk-1)
Lk = sup(c)=MinSupNum;\
End
3. Set of rule is generated
Return L = U Lk;
Candidate_gen(frequent itemset Lk-1)
for all(K-1)-itemset lE Lk-1 do
for all ij E Lk-1 do
//S is the result of the formula(2)
If for every r(1=r=k) such that S[r]=k-1 then
L1 = (frequent 1-itemsets);
C2 =L1 8 L1;
L2 = sup(c)=MinSupNum;
For (k=3;Lk-1 ≠Ø ;k++)do begin
For (j=k;j=m;j++)do
Generate CIVij k-1;
Ck =candidate_gen(Lk-1)
4. Check MLMS value of table
5. If rule isn't MLMS move to selection method
6. Else optimized rule is generated.
7. Exit.

Data Encoding

The processing of data in min-max algorithm wants some data encryption technique for illustration of data.
In this technique used binary encoding technique.
Fitness function
The population selection of Min-max algorithm could be a design of Fitness Function:
Ai = {frequent item support}
Wi1= {level of 1 Weight value of MLMS}
Bi = {those value or Data infrequent}
The selection strategy depends on the basis of individual fitness and concentration pi is the probability of selection of individual whose fitness value is greater than one and m(s) is the value which fitness is smaller than one however nearer to the value of 1. The Min-max operators verify the search capability and convergence of the algorithm. Min-max operators contain the selection crossover and mutation on the population and generate the new population. In this algorithm it replace each chromosome within the population to the corresponding rule, and then calculate selection probability pi for each rule based on above formula. In which single point are used. It partitioned into the multiple level domain of each attribute into a group and classifies the cut point of each continuous attributes into one group .And the crossover carried out between the corresponding teams of two individuals by a particular rate. Any bit in the chromosomes is mutated by a specific rate, that is, changing “0”to”1”,”1”to”0”.

IV PERFORMANCE ANALYSIS

The proposed method implemented in Matlab R2012 7.14.0.334 and tested with UCI Breast Caner dataset which has Number of Instances: 286 and Number of Attributes: 9.
The research work measured the total number of generated rules and the Elapsed time taken by each method.

Method	Max value	Min value	Minimum support	Minimum Confidence	No. of Rule generation	Elapsed time (in sec)
Apriori	0.5	0.2	0.4	0.3	166	37.72
MLMS	0.5	0.2	0.4	0.3	109	23.73
MLMS -GA	0.5	0.2	0.4	0.3	48	22.09

Table 1.1: Shows that the Comparative result analysis of different methods

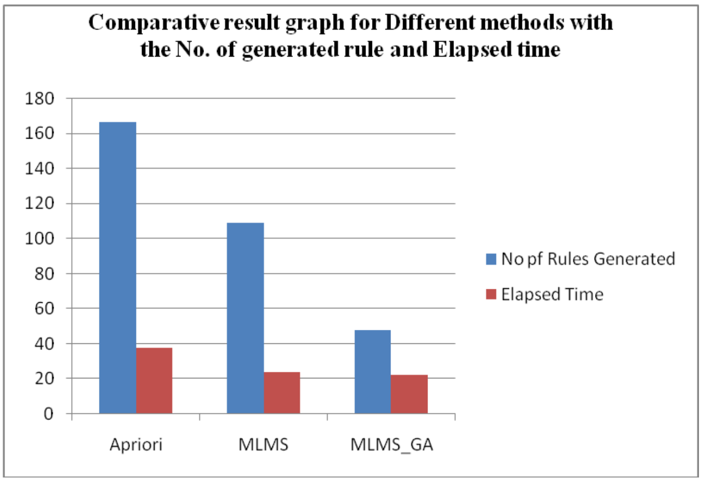

Figure 1.1: Comparative result graph for different methods.

V CONCLUSION

The proposed method MLMS_GA and is suitable for mining positive and negative rule from useful frequent and infrequent pattern sets. The algorithm is based on implemented pruning technique for reducing large number of association rules and improving the performance of algorithm and has used the

Genetic fitness function to check which form positive and negative rule should be mined. The only obstacle with the implemented method is that, this mining algorithm does not work very well for large instances of database. In the future, the study is still going on to modify this approach to achieve two goals:

1. An interesting measure is added to this approach for working large instances of databases.
2. To improve the performance of the algorithm.

REFERENCES

[1]Agrawal R, Imieliński T, Swami A, editors. Mining association rules between sets of items in large databases. ACM SIGMOD Record; 1993: ACM.

[2] D.A. Vidhate, Dr. Parag Kulkarni," Mining Association Rule by Multilevel Relationship Algorithm: An Innovative Approach for Cooperative Learning" International Journal of Emerging Trends and Technology in Computer Science (IJETICS), vol-4, 2013, Pp 130-137.

[3] Idheb Mohamad Ali O. Swesi, Azuraliza Abu Bakar, Anis Suhailis Abdul Kadir "Mining Positive and Negative Association Rules from Interesting Frequent and Infrequent Item sets" IEEE 9th International Conference on Fuzzy Systems and Knowledge Discovery, 2013. Pp 650-655.

[4] Li Guang-yuan, Cao Dan yang, Guo Jian-wei "Association Rules Mining with Multiple Constraints" Elsevier ltd. 2011, Pp 1678-1683.

[5] Peter P. Wakabi–Waiswa, Venansius Baryamureeba "Optimized Association Rule Mining with Genetic Algorithms " Seventh International Conference on Natural Computation, 2011,Pp 1116-1120.

[6] By Xiufend Piao, Zhan long Wang, Gang Liu: "Research on mining positive and negative association rules based on dual confidence" Fifth International Conference on Internet Computing for Science and Engineering. 978-1-4244-9954-0/11 $31 © IEEE (2011).

[7] By Li-Min Tsai, Shu-Jing Lin and Don-Lin Yang: "Efficient Mining of Generalized Negative Association Rules" in International Conference on Granular Computing 978-0-7695-4161-7/10 $ 26 © IEEE (2010).

[8] By CH.Sandeep Kumar, K.Shrinivas, Peddi Kishor T.Bhaskar: "An Alternative Approach to Mine Association Rules" 978-1-4244-8679- 3/11 $26.00 © IEEE (2011).

[9] By Dong, X., Niu, Z., Zhu, D., Zheng, Z., Jia, Q:" Mining Interesting Infrequent and Frequent Itemsets Based on MLMS Model". The Fourth International Conference on advanced Data Mining and Applications, ADMA (2008).