International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT2174138

Comparative Analysis of Association Rule Mining Based on Genetic Algorithm

Bushra Sheikh[1], Prof. Amita Suke[2], Prof. Khemutai Tighare[2]
[1]Wainganga College of Engineering and Management, Nagpur, Maharashtra, India
[2]Professor, Wainganga College of Engineering and Management, Nagpur, Maharashtra, India

Article Info
Volume 7, Issue 4
Page Number : 641-654

Publication Issue :
July-August-2021

Article History
Accepted : 23 Aug 2021
Published : 30 Aug 2021

ABSTRACT

Association rule mining play an important role in various data mining process. The diversity of association rule mining spread in various field such as market bucket analysis, medical diagnose and share market prediction. Now a days various authors and researcher focus on validation of association rule mining. For the validation of association rule mining used various optimization algorithm are used such as genetic algorithm, Ant Colony Optimization and particle of swarm optimization also used. For the mining of rule mining a variety of algorithm are used such as Apriori algorithm and tree-based algorithm. Some algorithm is wonder performance but generate negative association rule and also suffered from multi-scan problem. In this paper proposed multi-level minimum supports (MLMS-GA) association rule mining based on min-max algorithm and MLMS formula. In this method we used a multi-level minimum supports of data table as 0 and 1. The divided process reduces the scanning time of database. The proposed algorithm is a combination of MLMS and min-max algorithm. Support length key is a vector value given by the transaction data set. The process of rule optimization we used min-max algorithm and for evaluate algorithm conducted the real-world dataset such as heart disease data and some standard data used from UCI machine learning repository.

Keywords : Association Rule Mining, Multi-Level Minimum Supports, Genetic Algorithm, Ant Colony Optimization, Particle Swarm Optimization

I. INTRODUCTION

Association rule mining concept has been applied to market domain and specific problem has been studied, the management of some aspects of a shopping mall, and an architecture that makes it possible to construct agents capable of adapting the association rules has been used. Data mining refers to extracting knowledge from large quantity of data. Interesting association can be discovered among a large set of data items by association rule mining. The finding of interesting relationship among large amount of business transaction records can help in many business decisions making process. Association rules mining is an important task in the field of data mining, and frequent item set mining is a key step of many algorithms for association rules mining. There had been lots of work done for mining of association rules. When the dataset is large, the rules generated may be very large, but some of them are not interesting to the users, so, it is common to set some parameters to reduce the numbers of rules generated, support and confidence are two common parameters. An association rule R is of the form A → B, where A, B are disjoint subsets of the attribute set I. The support for the rule R is the number of database records which contain A ∪ B (often expressed as a proportion of the total number of records). The confidence in the rule R is the ratio:

$$\frac{\text{Support for R}}{\text{Support for A}}$$

These 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.[1] In practical investigations, it is usual to regard these rules as "interesting" only if the support and confidence exceed some threshold values. Hence the problem may be formulated as a search for all association rules within the database for which the required support and confidence levels are attained. 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.

II. RELATED WORK

Tasnia Rahman et. al.[1] adapted by Evolutionary Genetic Algorithm which is a global search heuristic technique that optimizes rule generation. In this paper, the performance of Fuzzy association rule mining algorithms is compared. For performance analysis, classical Apriori algorithm, Fuzzy Apriori algorithm, and Evolutionary GeneticFuzzyAprioriDC algorithm are being used.

In [2] presented the form of the directed item sets graph to store the information of frequent item sets of transaction databases, and give the trifurcate linked list storage structure of directed item sets graph. Furthermore, we develop the mining algorithm of maximal frequent item sets based on this structure. As a result, one realizes scanning a database only once, and improves storage efficiency of data structure and time efficiency of mining algorithm. We introduce a directed item sets graph to store the information of frequent item sets of transaction databases. Next we create the trifurcate linked list storage structure of directed item sets graph, and finally develop the mining algorithm of maximal frequent item sets based on directed item sets graph.

In [3] propose a new algorithm called U2P-Miner for mining frequent U2 patterns from univariate uncertain data, where each attribute in a transaction is associated with a quantitative interval and a probability density function. The algorithm is implemented in two phases. First, we construct a U2P-tree that compresses the information in the target database. Then, we use the U2P-tree to discover frequent U2 patterns. Potential frequent U2 patterns are derived by combining base intervals and verified by traversing the U2P-tree. We also develop two techniques to speed up the mining process. Since the proposed method is based on a tree-traversing strategy, it is both efficient and scalable. Our experimental results demonstrate that the U2P-Miner algorithm outperforms three widely used algorithms, namely, the modified Apriori, modified H-mine, and modified depth-first backtracking algorithms. We evaluate the performance of the U2P-Miner algorithm on synthetic and real datasets.

In [4] presented an algorithm for mining association rules with multiple 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, we exploit the properties of the given constraints to prune search space or save constraint checking in the conditional databases. Third, for each item set possible to satisfy the constraint, we generate its conditional database and perform the three phases in the conditional database recursively. Experimental results show that the proposed method outperform the revised FP-growth algorithm

III. PROPOSED METHODOLOGY

In this section discuss proposed 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 is a great challenge for large dataset. In the generation of valid rules association existing algorithm or method generate a series of negative rules, which generated rule affected a performance of association rule mining. In the process of rule generation various multi objective associations rule mining algorithm is proposed but all these are not solved. In this Paper we proposed MLMS-GA of association rule mining with min-max algorithm. In this algorithm we used MLMS used for multi-level minimum support for constraints validation. The scanning of database divided into multiple levels as frequent level and infrequent level of data according to MLMS. The frequent data logically assigned 1 and infrequent data logically assigned 0 for MLMS process. The divided process reduces the uninteresting item in given database. The proposed algorithm is a combination of MLMS and min-max algorithm along this used level weight for the separation of frequent and infrequent item. The weight value act as Support length key is a vector value given by the transaction data set. The support value passes as a vector for finding a near level between MLMS candidates key. After finding a MLMS candidate key the nearest level divide into two levels, one level takes a higher odder value and another level gain infrequent minimum support value for rule generation process. The process of selection of level also reduces the passes of data set. After finding a level of lower and higher of given support value, compare the value of level weight vector. Here level length vector work as a fitness function for selection process of min-max algorithm. Here we present steps of process of algorithm step by step and finally draw a flow chart of complete process.

Steps of algorithm (MLMS-GA)

1. 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; L^{j}i≠ φ, j = 1,2….,n; i++)

   $Ci^{G} = \cup j = 1,2,\ldots nL_{i}^{j}$

   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 vector value for selection process

   for all transaction LG do

   generate count table TC

L_1 =(frequent 1-itemsets);
C_2 =L_1 ∞ L_1;
L_2 ={cEC_2 | sup(c)≥MinSupNum};
For(k=3;L_{k-1} ≠Ø ;k++)do begin
For (j=k;j≤m;j++)do
Generate CIV_{ij}^{k-1};
C_k=candidate_gen(L_{k-1})
L_k ={cEC_k| sup(c)≥MinSupNum};\
End

3. Set of rule is generated
Return L = U' L_k;
Candidate_gen(frequent itemset L_{k-1})
a. for all(K-1)-itemsetlE L_{k-1} do
b. for all i_jE L_{k-1} do
c. //S is the result of the formula(2)
If for every r(1≤r≤k) such that S[r]≥k-1 then
L_1 = (frequent 1-itemsets);
C_2 =L_1 ∞ L_1;
L_2 = {cEC_2 | sup(c)≥MinSupNum};
For (k=3;L_{k-1} ≠Ø ;k++)do begin
For (j=k;j≤m;j++)do
Generate CIV_{ij}^{k-1};
C_k=candidate_gen(L_{k-1})

4. Check MLMS value of table

5. If rule is not MLMS go to selection process

6. Else optimized rule is generated.

7. Exit

a.) Data Encoding

The process of data in min-max algorithm needs some data encoding technique for representation of data. In this technique used binary encoding technique.

b.) Fitness function

The population selection of Min-max Algorithm is a design of Fitness Function:

$$m(S) = \frac{Ai}{wi} + \frac{Bi}{L \times (1 - wi)}$$

Ai = {frequent item support}
Wi_1= {level of Wight value of MLMS}
Bi = {those value or Data infrequent}

The selection strategy based on the basis of individual fitness and concentration pi is the probably of selection of individual whose fitness value is greater than one and m(s) is a those value whose fitness is less than one but near to the value of 1.The Min-max operators determine the search capability and convergence of the algorithm. Min-max operators hold the selection crossover and mutation on the population and generate the new population. In this algorithm it restores each chromosome in 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 divides 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 groups of two individuals by a certain rate. Any bit in the chromosomes is mutated by a certain rate, that is, changing “0”to”1”,”1”to”0”. Now we explain complete process of algorithm shows block diagram of proposed algorithm using min-max algorithm.

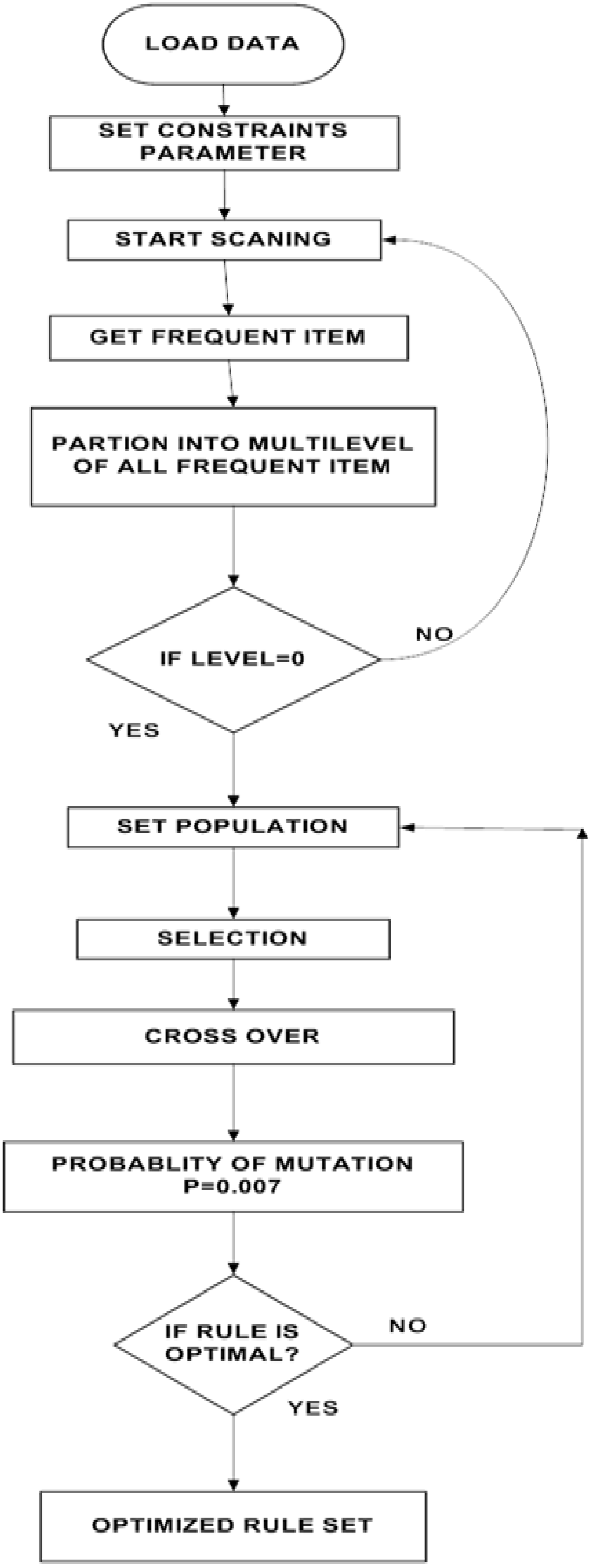


Figure 1: workflow of proposed system

IV. EXPERIMENTAL RESULT AND PROCESS

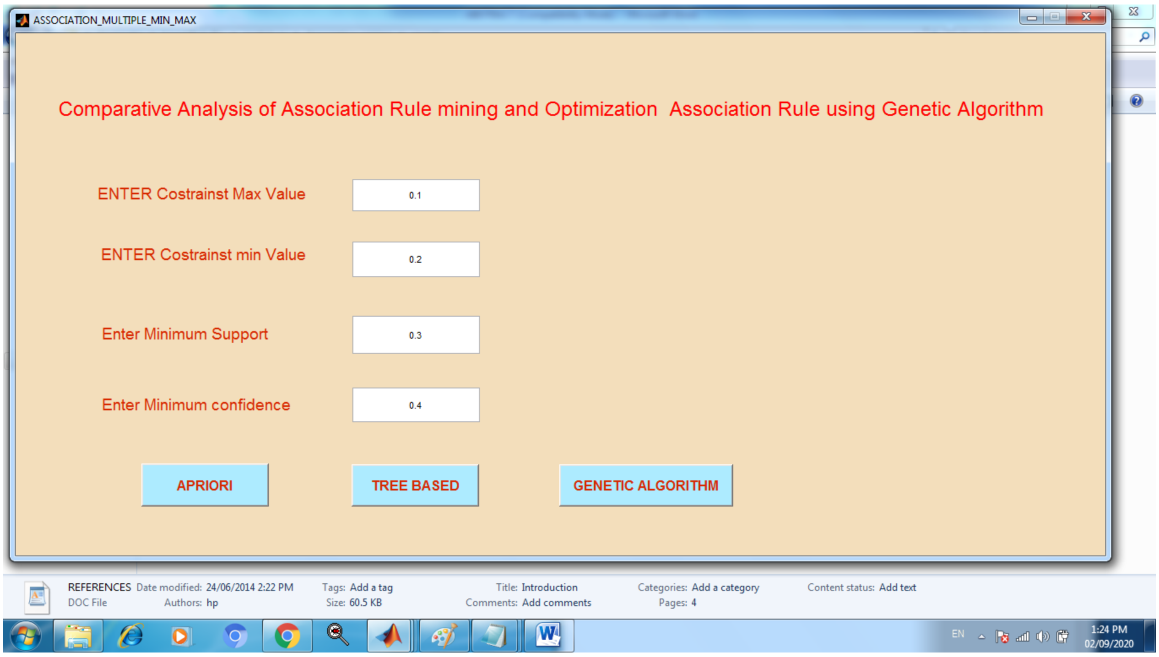


Figure 2: shows that the main window initially empty for experimental results.

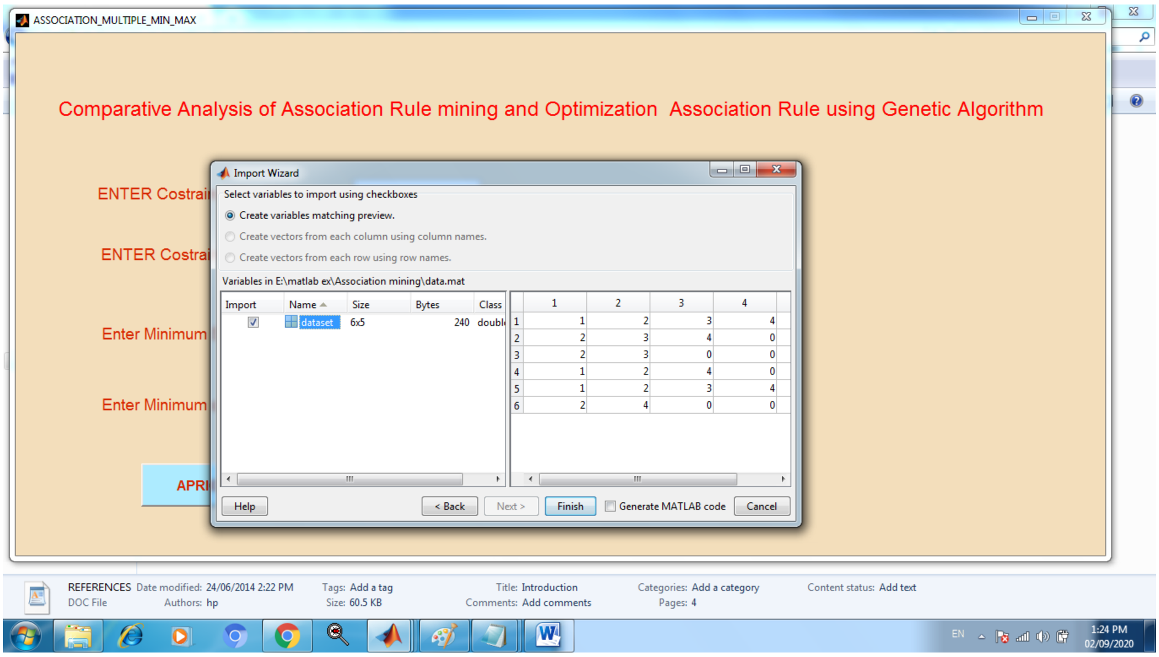


Figure 3: shows that the main window with load the data for experimental results.

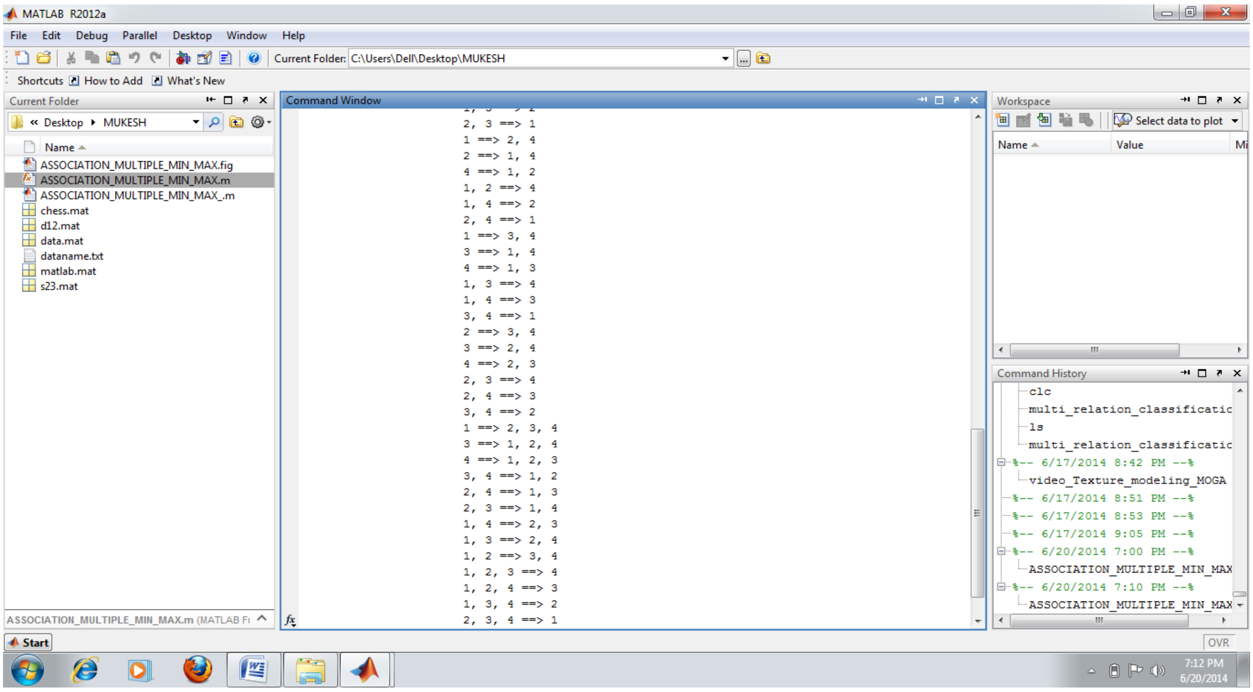


Figure 4: Shows that the Rule generation by Apriori method with the no. of generated rule is 48.

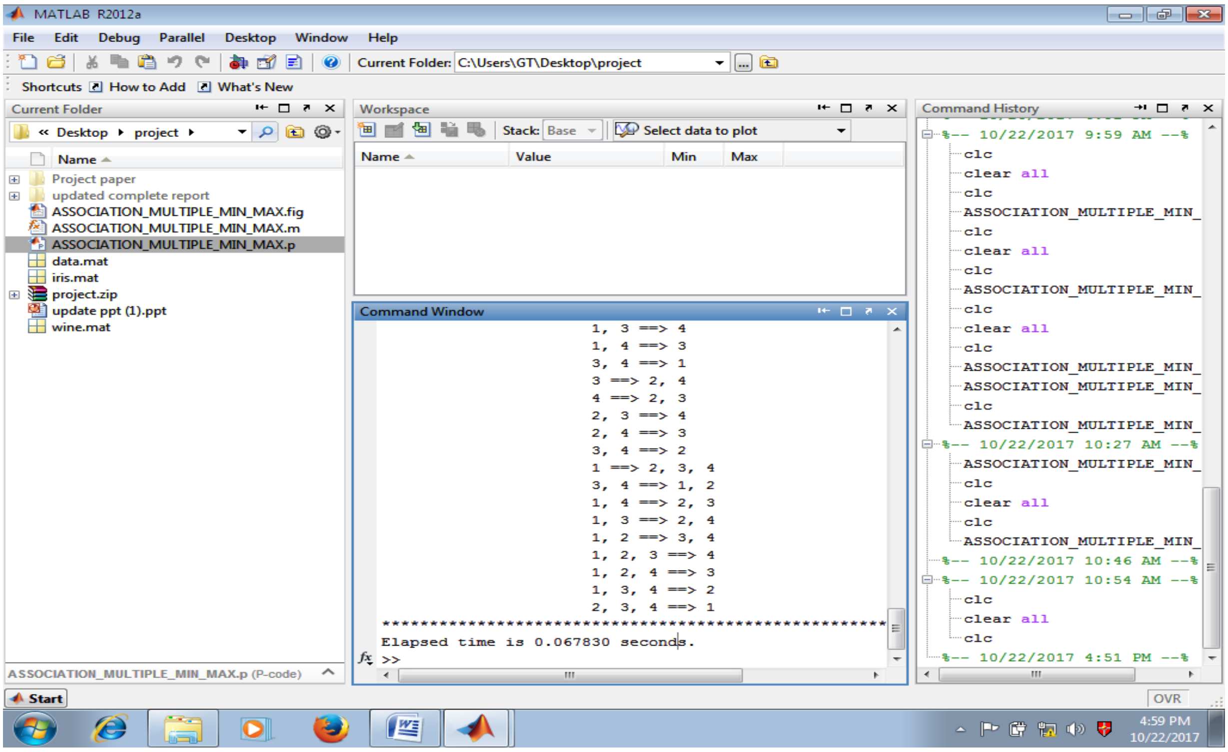


Figure 5: Shows that the Rule generation by Tree method with the no. of generated rule is 36.

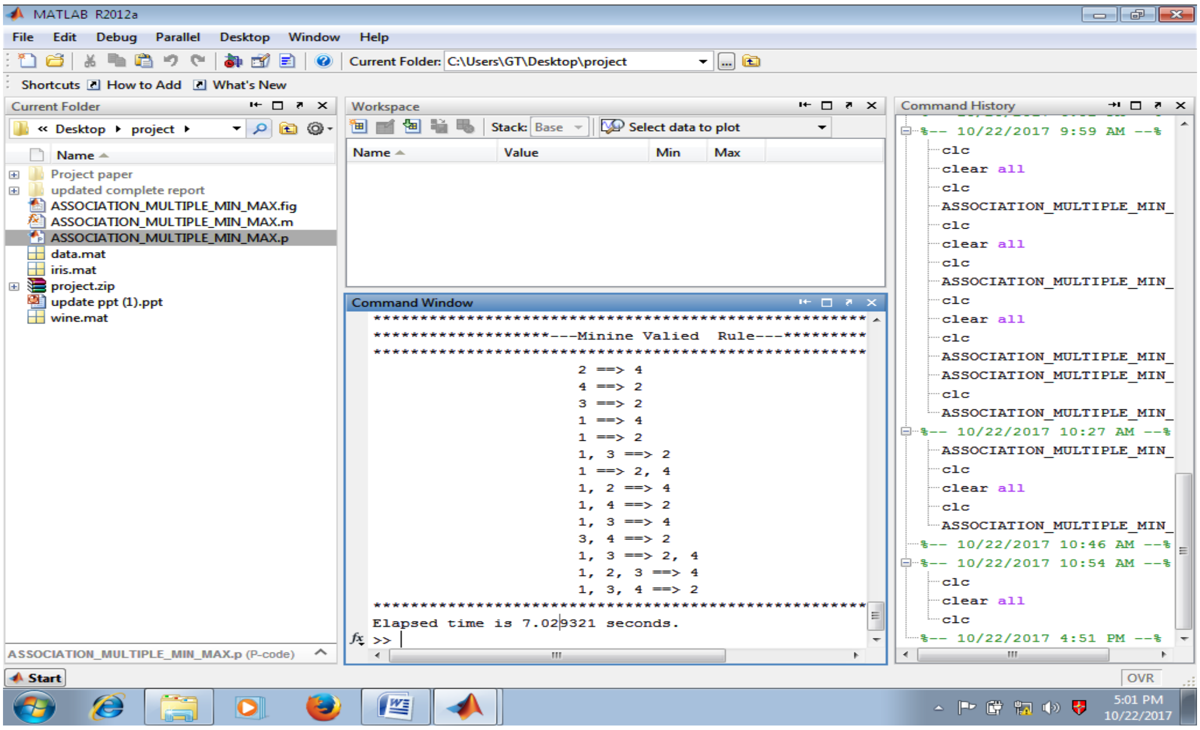


Figure 6: Shows that the Rule generation by Genetic Algorithm with the no. of generated rule is 14.

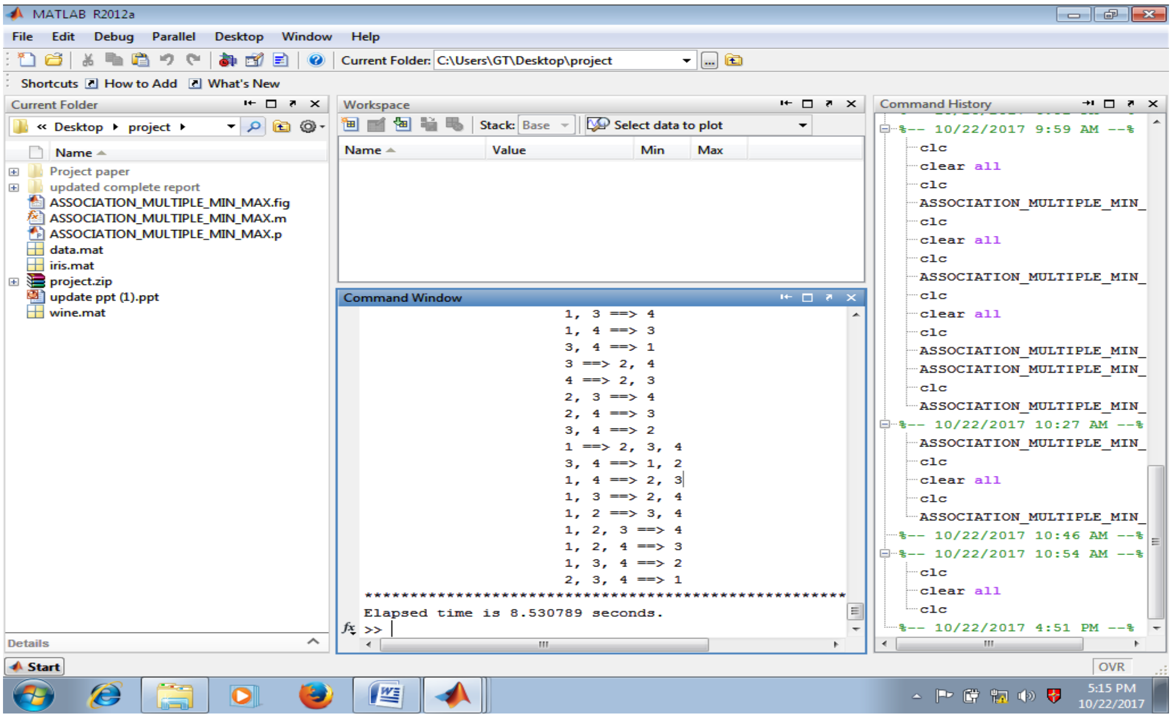


Figure 7: Shows that the Rule generation by Apriori method with the no. of generated rule is 36.

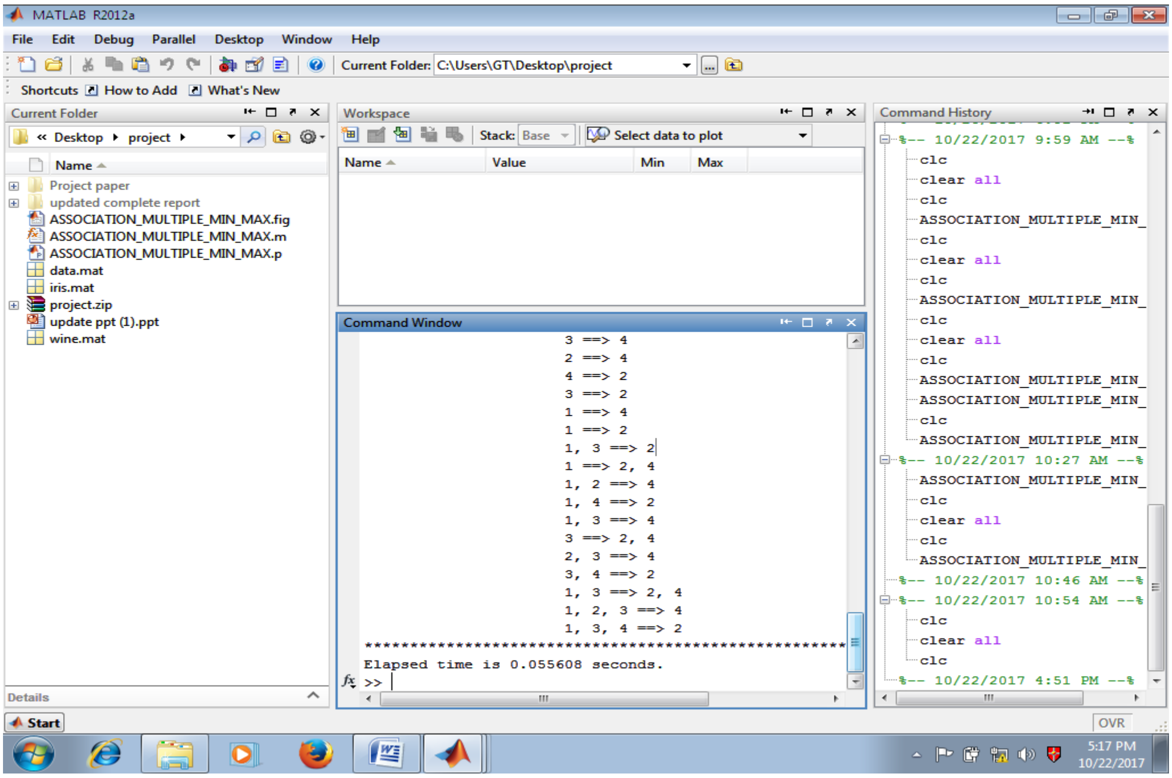


Figure 8: Shows that the Rule generation by Tree method with the no. of generated rule is 17.

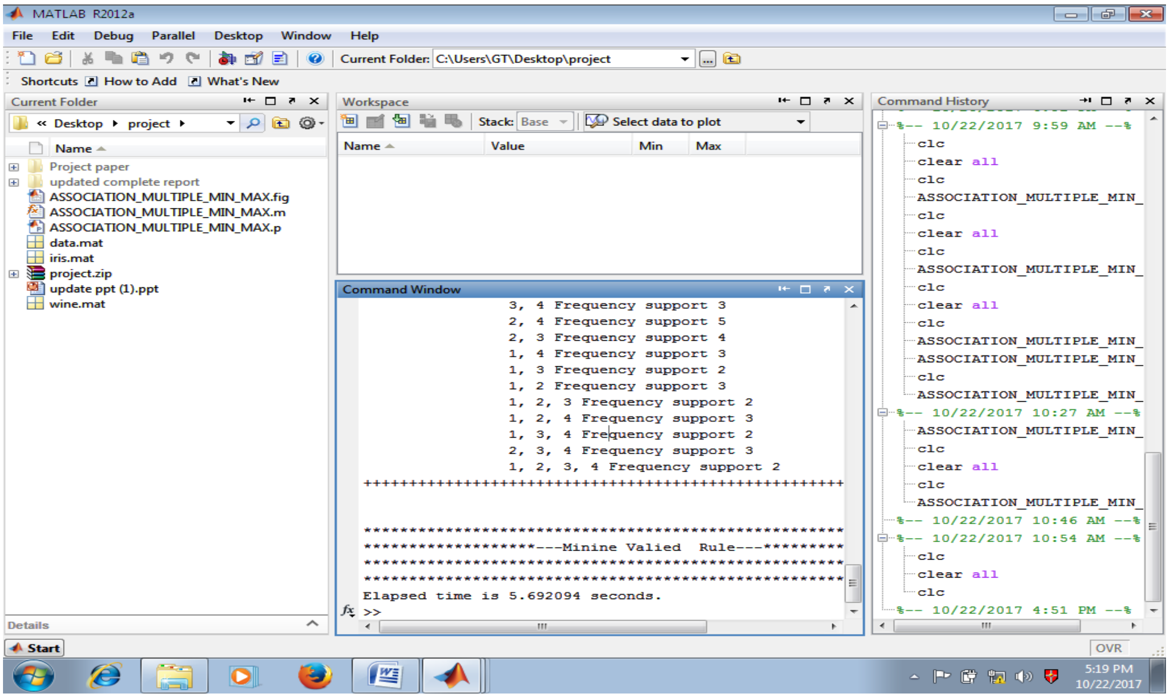


Figure 9: Shows that the Rule generation by Sin Cosine method with the no. of generated rule is 0.

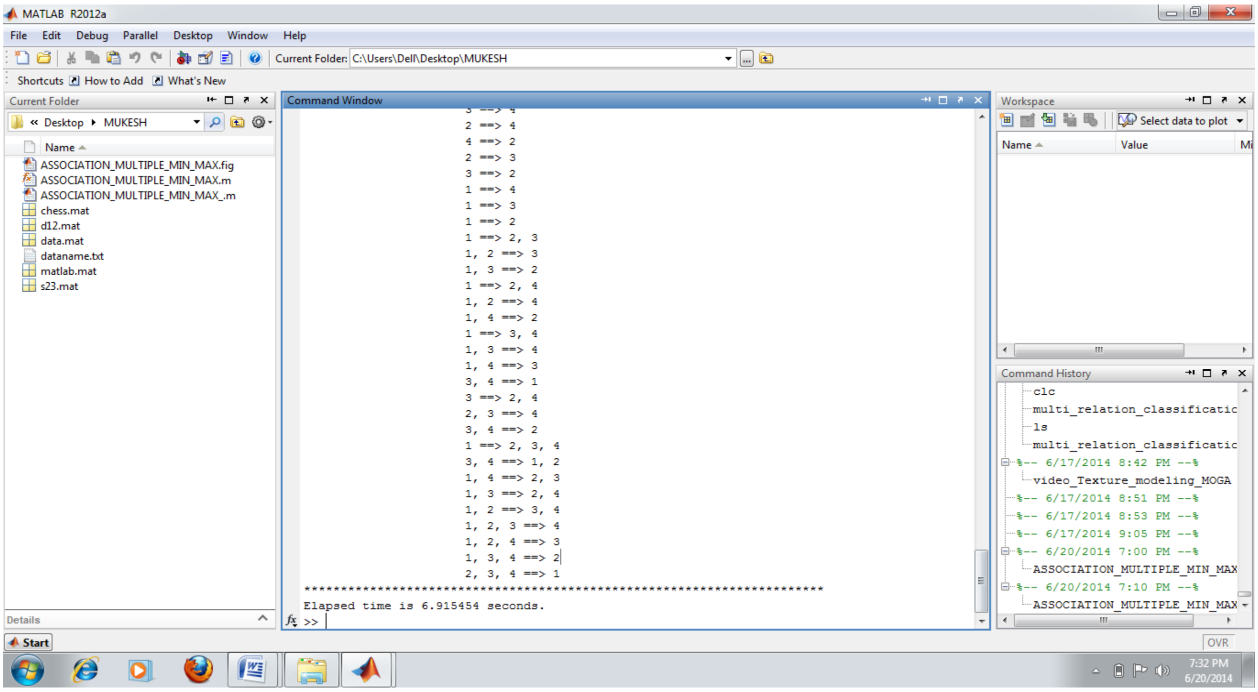


Figure 10: Shows that the Rule generation by Apriori method with the no. of generated rule is 30.

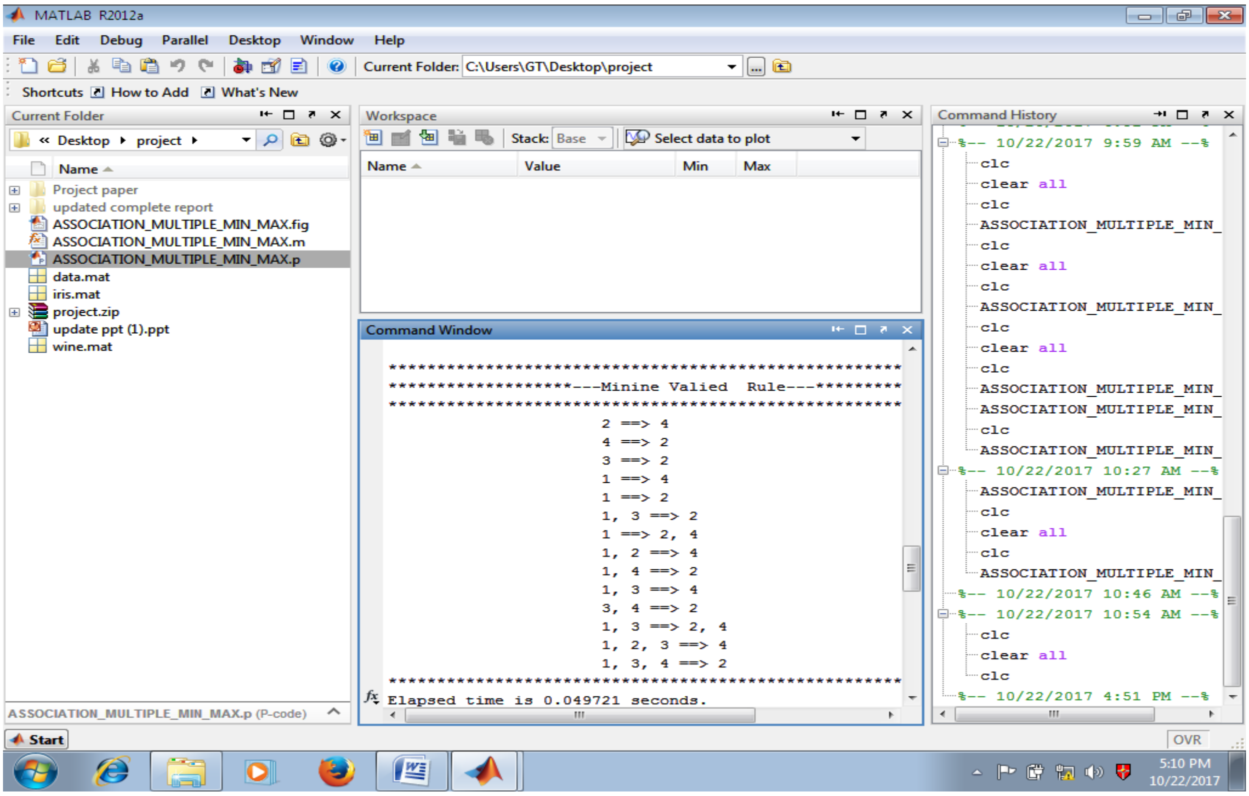


Figure 11: Shows that the Rule generation by Tree method with the no. of generated rule is 15.

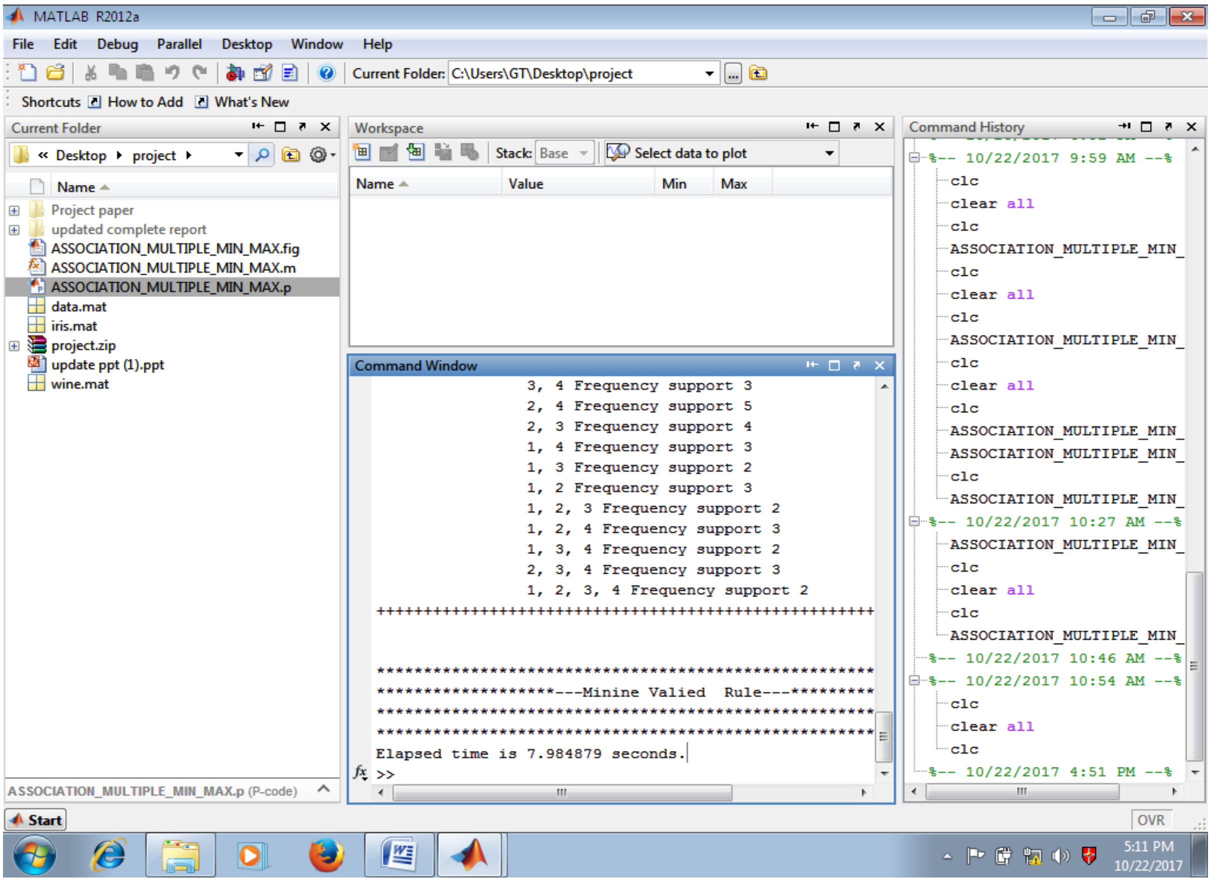


Figure 12: Shows that the Rule generation by Sin Cosine method with the no. of generated rule is 0.

V. COMPARATIVE RESULT ANALYSIS

Method Name	Min value	Max value	Minimum support	Minimum Confidence	No. of Rule generation	Elapsed time
Apriori	0.3	0.4	0.5	0.6	48	7.624
Tree	0.3	0.4	0.5	0.6	36	0.0678
Genetic Algo	0.3	0.4	0.5	0.6	14	7.029

Table 1 : Shows that the Comparative result analysis of different methods and the no. of generated rule are also different, for our proposed experimental method.

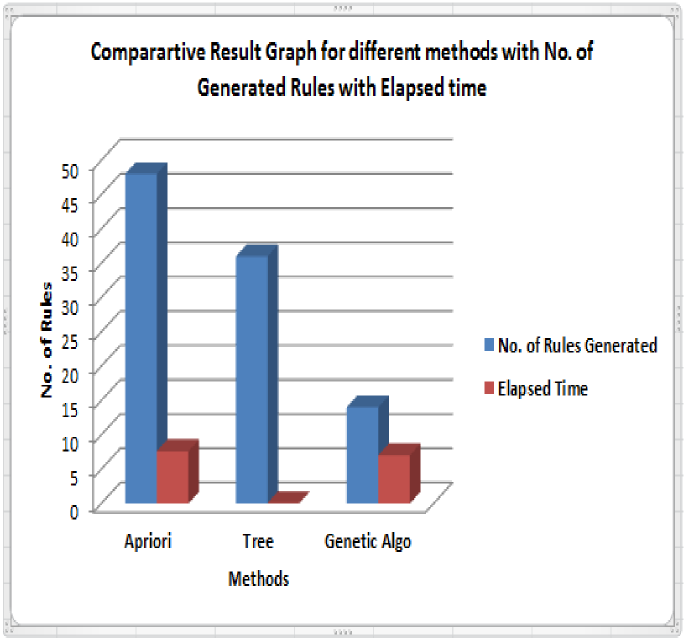
Method Name	Min value	Max value	Minimum support	Minimum Confidence	No. of Rule generation	Elapsed time
Apriori	0.4	0.5	0.6	0.7	36	8.530
Tree	0.3	0.4	0.5	0.6	17	0.055
Genetic Algo	0.3	0.4	0.5	0.6	0	5.692

Table 2 : Shows that the Comparative result analysis of different methods and the no. of generated rule are also different, for our proposed experimental method for different value.

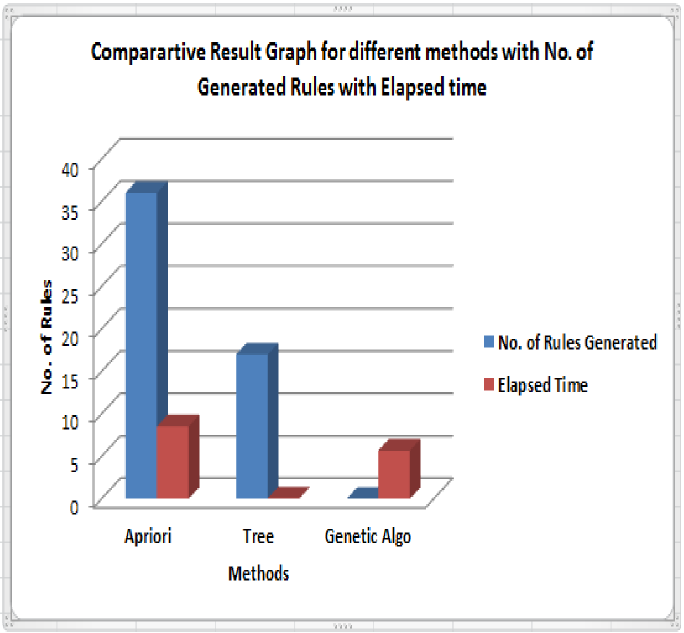
Method Name	Min value	Max value	Minimum support	Minimum Confidence	No. of Rule generation	Elapsed time
Apriori	0.3	0.4	0.6	0.8	36	7.128
Tree	0.3	0.4	0.6	0.8	14	0.049
Genetic Algo	0.3	0.4	0.6	0.8	0	7.984

Table 3 : Shows that the Comparative result analysis of different methods and the no. of generated rule are also different, for our proposed experimental method for different value

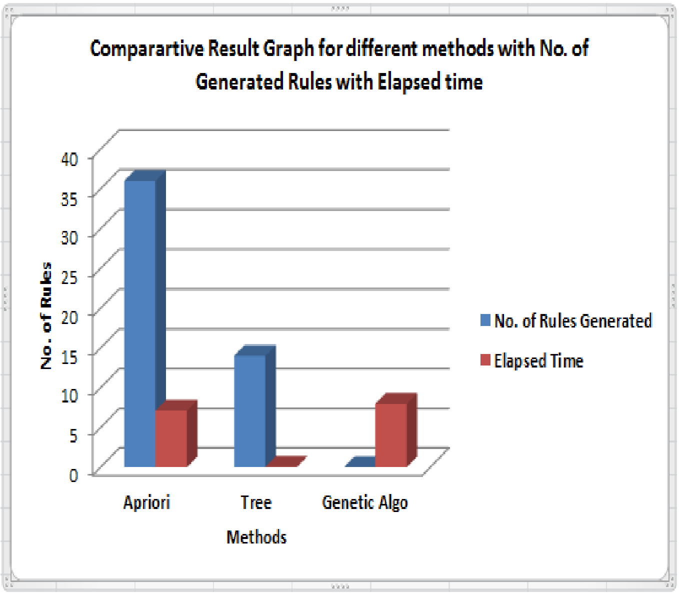
VI. COMPARATIVE RESULT GRAPH AND ANALYSIS



Graph 1: Shows that the comparative result for different methods and shows that our proposed method generated the no. of rule are less than other methods.



Graph 2: Shows that the comparative result for different methods and shows that our proposed method generated the no. of rule are less than other methods.



Graph 3: Shows that the comparative result for different methods and shows that our proposed method generated the no. of rule are less than other methods.

VII. CONCLUSIONS

In this dissertation we proposed a novel method for optimization of association rule mining. Our propped algorithm is combination of min-max function and genetic algorithm. The min-max function and genetic algorithm work together and perform condition-based rule generation process. The min-max condition-based function operate in sine and cosine based trigonometric function for the processing of genetic fitness function. We have observed that when we modify the condition new rules in large numbers are found. This implies that when min-max is solely determined through support and confidence, there is a high chance of eliminating interesting rules. With more rules emerging it implies there should be a mechanism for managing their large numbers. The large generated rule is optimized with genetic algorithm.

We theoretically proofed a relation between locally large and globally large patterns that is used for local pruning at each site to reduce the searched candidates. We derived a locally large threshold using a globally set minimum recall threshold. Local pruning achieves a reduction in the number of searched candidates and this reduction has a proportional impact on their duction of exchanged messages. Our proposed algorithm is combination of MLMS and min-max algorithm. We have observed that when we modify the scan process of transaction generation of rule is fast. With more rules emerging it implies there should be a mechanism for managing their large numbers. The large generated rule is optimized with min-max algorithm. We theoretically proofed a relation between locally large and globally large patterns that is used for local pruning at each site to reduce the searched candidates. We derived a locally large threshold using a globally set minimum recall threshold. Local pruning achieves a reduction in the number of searched candidates and this reduction has a proportional impact on the reduction of exchanged messages.

Our proposed algorithm is performed better optimization in comparison of MLML-GA and monotonic condition-based association rule mining. Condition based association rule mining is great advantage over conventional rule generation technique. The conventional rule generation technique used some standard algorithm. The proposed algorithm is very promising in the field of association rule mining. The proposed algorithm has multiple constraints such as genetic algorithm and sine and cosine function. The value of sine and cosine increases the process of algorithm work as normal apriori and tree-based technique.

VIII. REFERENCES

[1]. Tasnia Rahman, Mir Md. Jahangir Kabir, Monika Kabir, "Performance Evaluation of Fuzzy Association Rule Mining Algorithms", 4th International Conference on Electrical

Information and Communication Technology (EICT), 20-22 December 2019
[2]. Ying-Ho Liu "Mining frequent patterns from univariate uncertain data" Data & Knowledge Engineering, 2012. Pp 47-68.
[3]. Li Guang-yuan, Cao Dan yang, Guo Jian-wei "Association Rules Mining with Multiple Constraints" Elsevier ltd. 2011, Pp 1678-1683.
[4]. 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, 2012. Pp 650-655.
[5]. Huang QingLan, Duan LongZhen "Multi-level association rule mining based on clustering partition" IEEE Third International Conference on Intelligent System Design and Engineering Applications, 2013. Pp 982-986.
[6]. Vidhu Singhal, Gopal Pandey "A Web Based Recommendation Using Association Rule and Clustering" International Journal of Computer & Communication Engineering Research, 2013. Pp 1-5.
[7]. Poundekar, Mukesh & Manekar, Amitkumar & Baghel, Mukesh & Gupta, Hitesh. (2014). Mining Strong Valid Association Rule from Frequent Pattern and Infrequent Pattern Based on Min-Max Sinc Constraints. 450-453. 10.1109/CSNT.2014.95..
[8]. Noha Negm , Passent Elkafrawy , Mohamed Amin, Abdel Badeeh M. Salem "Investigate the Performance of Document Clustering Approach Based on Association Rules Mining" International Journal of Advanced Computer Science and Applications, vol-4, 2013. Pp 142-151.
[9]. Huang QingLan, Duan LongZhen "Multi-level association rule mining based on clustering partition" Third International Conference on Intelligent System Design and Engineering Applications, IEEE 2013. Pp 982-985.
[10]. Aritra Roy, Rajdeep Chatterjee "Introducing New Hybrid Rough Fuzzy Association Rule Mining Algorithm" Proc. of Int. Conf. on Recent Trends in Information, Telecommunication and Computing, ITC, 2014. Pp 167-175.
[11]. Deepak A. Vidhate, Dr. Parag Kulkarni "improvement in association rule mining by multilevel relationship algorithm" International Journal of Research in Advent Technology, Vol-2, 2014. Pp 366-373.
[12]. Yin-Fu Huang, San-Des Lin "Applying Multidimensional Association Rule Mining to Feedback-based Recommendation Systems" International Conference on Advances in Social Networks Analysis and Mining, 2011. Pp 412-417.
[13]. XING Xue, CHEN Yao. WANG Yan-en "Study on Mining Theories of Association Rules and Its Application" IEEE, 2010. Pp 2-10.
[14]. Yingqin Gu, Hongyan Liu, Jun He, Bo Hu and Xiaoyong Du "A Multi-relational Classification Algorithm based on Association Rules" IEEE, 2009. Pp 235-242.
[15]. W. Li, J. Han, and J. Pei "CMAR: Accurate and efficient Classification Based on Multiple Class-Association Rules" Proceedings of the ICDM, IEEE Computer Society, San Jose California, 2001, Pp 369-376.

Cite this article as :

Bushra Sheikh, Prof. Amita Suke, Prof. Khemutai Tighare, "Comparative Analysis of Association Rule Mining Based on Genetic Algorithm", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 4, pp. 641-654, July-August 2021. Available at doi : https://doi.org/10.32628/CSEIT2174138
Journal URL : https://ijsrcseit.com/CSEIT2174138