© 2018 IJSRSET | Volume 4 | Issue 4 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Frequent Pattern Mining using Genetic Algorithm in Data Mining

Swati Koshti*[1] , Aniket Patel[2]
[1]Computer Department, Silver Oak College of Engineering, Ahmedabad, Gujarat, India
[2]Assistant Professor. I.T. Department, Silver Oak College of Engineering, Ahmedabad, Gujarat, India

ABSTRACT

In this paper, we discussed about the frequent pattern mining in association rule mining (ARM). An association rule mining have been many approaches like as AIS, SETM, FP-Growth, A priori, Genetic Algorithm, Particle Swarm Optimization. But we use in the research work genetic algorithm (GA). The major advantage of using GAs in the discovery of prediction rules is that they perform global search and its complexity is less compared to other algorithms as the genetic algorithm is based on the greedy approach. The main aim of this research work, fitness function has been improvised more generated rules with pattern enhancement using Genetic Algorithm.

Keywords: Association Rule Mining (ARM), Genetic Algorithm (GA), Support, Confidence, Comprehensibility, Interestingness.

I. INTRODUCTION

Data mining is the process of discovering patterns in large amount of dataset. It is also known as the knowledge Discovery of Database or KDD. Many other terms have similar to data mining like as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. The data mining have two tasks:-Predictive and Descriptive. In the Predictive included Classification, Regression, Time Series analysis, Prediction. In the Descriptive task included the Clustering, Summarization, Association rules, Sequence discovery.

Here, we focus to the association rules of the descriptive task in data mining. In the data mining, an association rule is co-occurrence of the item set. An Association rule mining are finding frequent patterns, associations, correlations, or causal structures among sets of items or objects in transactional database, relational databases, and other information repositories. An association rule many applications like as Market Basket Analysis, Cross-marketing, Catalog design, loss-leader analysis, Clustering, Classification etc.

An association rule deals with categorical data, if one intends to proceed with numerical data, performing data discretization will improve the accuracy of the rules. An association rule implication of the form:-
$X \rightarrow Y$, where X, $Y \subset I$, and $X \cap Y = \emptyset$

An association rule mining is mainly two measures:-Support and Confidence. The Support defined as measure of what fraction of the population satisfies both the antecedent
and the consequent of the rule. It represented as:-
SUP (X) =SUP (XUY)/N
The Confidence defined as measure of how often the consequent is true when the antecedent is true. It represented as:-
Confidence=SUP (XUY)/SUP (X)

The problem of mining associations rule can be decompose into two sub-problems:- Finding all frequent items sets using minimum support and Finding association rules from frequent item sets using minimum confidence value.

The rest of the paper is organized as follows. In section II describes Methods & Material and section III comparative analysis of experimental result. In section IV describe the conclusion & future work.

II. METHODS AND MATERIAL

A. Genetic Algorithm

Genetic algorithm is a search heuristic that mimics the process of natural evaluation. The heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithm belong to the larger class of evolutionary algorithms, which generate solutions to optimization problems using techniques inspired by natural evolution , such as inheritance, mutation, selection and crossover.

The genetic algorithms are important when discovering association rules because they work with global search to discover the set of items frequency and they are less complex than other algorithms often used in data mining. The genetic algorithms for discovery of association rules have been put into practice in real problems such as commercial database, biology and fraud detection event sequential analysis.

Genetic operators:-

(1) Selection:-Selection deals with the probabilistic survival of the fittest, in that, more fit chromosomes are chosen to survive. Where fitness is a comparable measure of how well a chromosome solves the problem at hand.

There are different techniques to implement selection in genetic algorithms. They are tournament selection, roulette wheel selection, proportionate selection, rank selection, and steady state selection etc.

(2) **Crossover:** This operation is performed by selecting a random gene along the length of the chromosomes and swapping all the genes after that point.

The most popular crossover selects any two solutions strings randomly from the mating pool and some portion of the strings is exchanged between the strings. The selection point selected randomly. A probability of crossover is also introduced in order to give freedom to an individual solution string to determine whether the solution would go for crossover or not.

(3) **Mutation:** Mutation is the occasional introduction of new features in to the solution string of the population pool to maintain diversity in the population.

Though crossover has the main responsibility to search for the optimal solution, mutation is also used for this purpose. Mutation operator changes a 1 to 0 or vise versa, with a mutation probability of. The mutation probability is generally kept low for steady convergence.

B. Why Pattern Is Required?

In the Association rule mining, frequent pattern mining is defined as the subsequent item set of the database. There are many data available on the online retail shop. The transactions of the market basket analysis many items required to the users or customers. There are many transaction of the items frequently used to database and create the one sequence of the items sub sequent part for the databases. The all transactions are do successfully is not required. Some fuzzy, non relevant, noisy, blank data should be process the pattern. So, the arranged all data in one format and eliminate the fuzzy, non-relevant, noisy, blank data of the database and create the sequential new patterns of the dataset.

C. Flowchart of proposed work

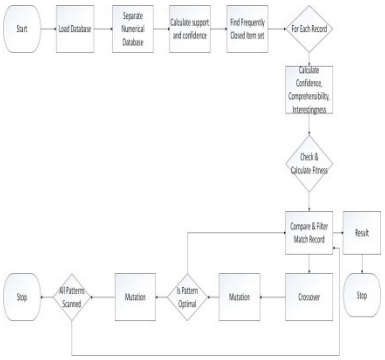

Figure 1. Flowchart of proposed work

D. Algorithm

Step:-1 Load Dataset
Step:-2 Separate Numerical Dataset
Step:-3 Calculate Support & Confidence
Step:-4 Find Frequently Closed Item Set
Step:-5 Confidence=SUP (AUC)/SUP (A)
Step:-6 Comprehensibility=log (1+|C|)/log (1+|AUC|)
Step:-7 Interestingness= [SUP (AUC)/ SUP (A)]*[SUP (AUC)/SUP (C)]*[1-SUP (AUC)/ |D|]
Step:-8 Calculate Fitness
Step:-9 Calculate & Match Records
Step:-10 loop
Step:-11 If match record→result else stop
Step:-12 Otherwise Continue step 13
Step:-13 Crossover
Step:-14 Mutation
Step:-15 loop
Step:-16 If patterns optimal→Mutation
Step:-17 Otherwise go to step 9
Step:-18 loop
Step:-19 If all patterns are scanned→stop
Step:-20 Otherwise go to step 9

E. Description

In above proposed algorithms, three measures, Confidence, Comprehensibility, and Interestingness have been used as different objectives for our multi-objective optimization which is amplified with genetic algorithms approach. First of all load the dataset and separate the numerical data of the population. Calculate the support and confidence of the data and find frequently closed item set. For each record calculate the confidence, comprehensibility, and interestingness of the dataset. The confidence is transaction of both antecedent part and consequent part of the dataset and divide to the number of transaction data of antecedent part. The comprehensibility is Check out the fitness function of the all records and compare all records to the original dataset. The compare and filter for match records dataset then, get the result and stop the process. Otherwise, apply the crossover operation of current solution. Apply the mutation operation and get new solution & also Scanned all the patterns get the solution and stop the process. If patterns no scanned then the go to compare and filter records of the dataset and get accurate results.

III. RESULTS AND DISCUSSION

The experiment uses online retail dataset obtained from UCI machine learning repository. The data set has 541909 samples. The experiment was executed software was Net Beans IDE and back end XAMPP server. The below graphs are show the analysis of result.

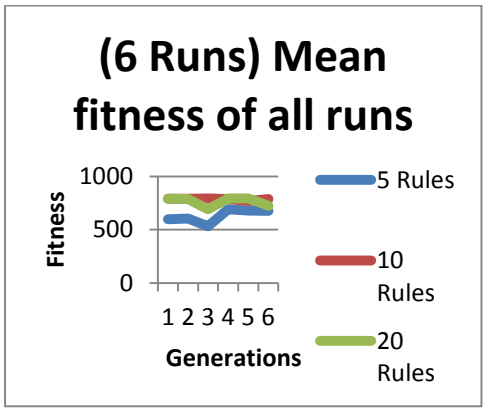

Figure 2. Fitness of Generation Graph

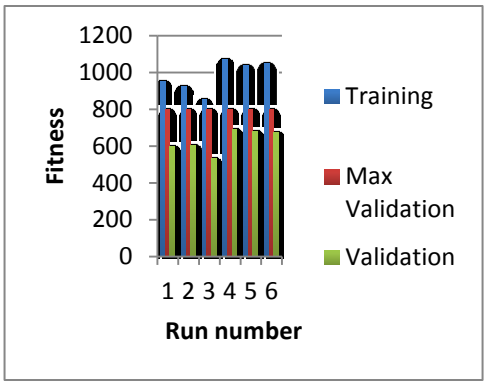

Figure 3. Fitness graph

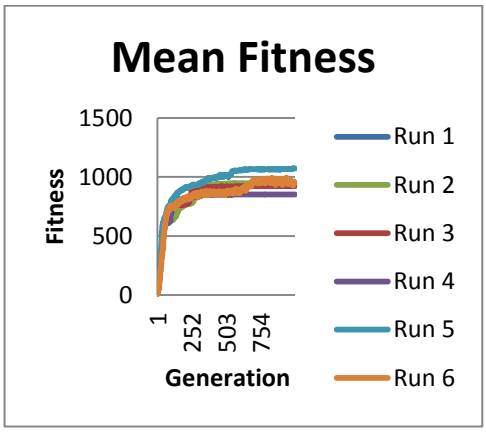

Figure 4. Mean Fitness value graph

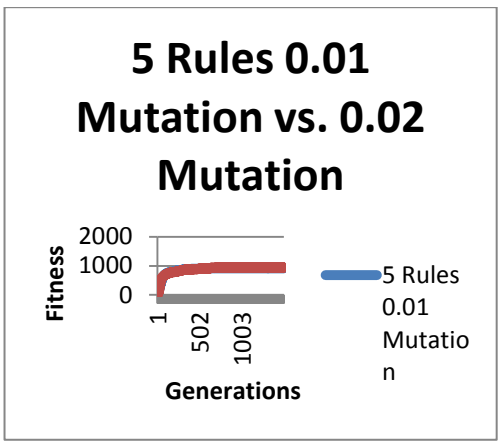

Figure 5. Mutation generation graph

IV. CONCLUSION

The genetic algorithm based on association rule mining algorithm for the frequent pattern mining. In this research work, implementation work is get the better results for more generated rules with pattern enhancement.

V. REFERENCES

[1]. Jyotsana Dixit and Abha Choubey, "A Survey of Various Association Rule Mining Approaches", International Journal of Advanced Research In Computer Science and Software Engineering, Volume 4, Issue 3, March 2014, 651-655.

[2]. Mr. Prasanth Kumar, Mr. B. Yakhoob, and Mrs. N. Deepika, "Association Rule Mining Algorithms: A Comparative Survey", International Journal of Advanced Engineering and Global Technology Vol-04, Issue-06, November 2016, 1350-1354.

[3]. Pandya Jalpa P., and Morena Rustom D., "A Survey on Association Rule Mining Algorithms Used in Different Application Areas", International Journal of Advanced Research in Computer Science, Volume 8, No. 5, May-June 2017, 1430-1436.

[4]. Rupali Haldulakar and Prof. Jitendra Agrawal, "Optimization of Association Rule Mining through Genetic Algorithm", International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 3 March 2011, 1252-1259.

[5]. Anandhavali M, Suraj Kumar Sudhanshu, Ayush Kumar and Ghose M.K., "Optimized association rule mining using genetic algorithm", Advances in Information Mining, ISSN: 0975-3265, Volume 1, Issue 2, 2009, 01-04.

[6]. M. Ramesh Kumar and Dr. K. lyakutti, "Genetic Algorithms for the prioritization of Association Rules", IJCA Special Issue on "Artificial Intelligence Techniques-Novel Approaches & Practical Applications" AIT, 2011, 35-38.

[7]. Peter P. Wakabi-Waiswa and Dr. Venansius Baryamureeba and Karunakaran Sarukesi, "Optimized Association Rule Mining with Genetic Algorithms", International Conference on Natural Computation, 2011 IEEE, 1116-1120.

[8]. Deep Hakani and Harshita Kanani, "Association Rule Optimized by Multi Objective Genetic

Algorithm", IJSRD-International Journal for Scientific Research & Development Vol. 2, Issue 01, 2014, 385-389.

[9]. Manish Saggar, Ashish Kumar Agrawal, and Abhimanyu Lad, "Optimization of Association Rule Mining using Improved Genetic Algorithms", International Conference on Systems, Man and Cybernetics, IEEE 2004, 3725-3729.

[10]. Jeetesh Kumar Jain, Nirupama Tiwari, and Manoj Ramaiya, "Mining Positive and Negative Association Rules From Frequent and Infrequent Pattern Using Improved Genetic Algorithm", 5th International Conference on Computational Intelligence and Communication Networks, IEEE 2013, 516-521.

[11]. B. Minaei-Bidgoli, R. Barmaki, and M. Nasiri, "Mining Numerical association rules via multi-objective genetic algorithms", Information Sciences 233, ELSEVIER 2013, 15-24.

[12]. Ghanshyam Dhanore and Setu Kumar Chaturvedi, "An Optimization of Association rule mining for large database using K-map and Genetic Algorithm: A Review", International Conference on Electronics and Communication System (ICECS-2014), 1-5.

[13]. Dimple S. Kanani and Shailendra K. Mishra, "An Optimized Association Rule Mining Using Genetic Algorithm", International Journal of Computer Applications (0975-8887), Volume 119-No. 14, June 2015, 11-15.

[14]. Aashna Agarwal and Prof(Dr.) Nirali Nanavati, "Association rule mining using hybrid GA-PSO for multi-objective optimisation", International Conference on Computational Intelligence and Computing Research, IEEE 2016.

[15]. Kamel Eddine Heraguemi, Nadjet Kamel, and Habiba Drias, "Multi-Objective Bat Algorithm for Mining Interesting Association Rules", Springer International Publishing AG 2017, 13-23.

[16]. Mohit K. Gupta and Geeta Sikka, "Association Rules Extraction using Multi-objective Feature of Genetic Algorithm", World Congress on Engineering and Computer Science 2013, Volume II.

[17]. Sanat Jain and Swati Kabra, "Mining & Optmization of Association Rules Using Effective Algorithm", International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 2, Issue 4, April 2012, 281-285.

[18]. Sanjay Tiwari and Mahainder Kumar, "Optimization in Association Rule Mining Using Distance Weight Vector and Genetic Algorithm", International Journal of Advanced Technology & Engineering Research (IJATER), Volume 4, Issue 1, Jan. 2014, 79-84.