DESIGN OF DYNAMIC ALGORITHM BY HYBRIDIZATION OF GA-PSO AND MULTI OBJECTIVE OPTIMIZATION THROUGH ASSOCIATION RULE MINING ON BAKERY DATASETS

Shreya Patel[1] & Dr. Preeti Baser[2]
[1]M.E., Student, Computer Engineering Department, Kalol institute of technology and research center, Kalol, India.
[2]Associate Professor, Student, Computer Engineering Department, Kalol institute of technology and research center, Kalol, India.

Received: January 25, 2019 Accepted: March 10, 2019

ABSTRACT: *Association Rule Mining (ARM), a Data Mining process, extracts hidden strong relationships among a large set of the correlated data. With the burgeoning advancement and application of Association Rule Mining in diverse fields ranging from the web usage mining to medical diagnosis and business intelligence to geographical information systems, the decision making in ARM involves a multiobjective perspective to obtain an interesting and accurate rule set. By considering the Pareto optimality, an optimal trade-off is established between the conflicting and incommensurate performance parameters comprehensibility, interestingness and confidence of the mined rules. Both, Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), being population-based stochastic search method, have found their strong base in mining association rules. We propose an association rule mining scheme using our proposed multi-objective hybridization of GA-PSO algorithm. The primary advantage of the proposed algorithm is that the hybridization of multiple objective-GA with multi objective-PSO balances the exploration and exploitation tasks, resulting in valuable extraction of accurate and interpretable mined rules.*

Key Words: : *Pareto dominance,Geneticalgorithm,PSO*

I. Introduction

A Genetic algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms 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. [17]
Discovering association rules is at the heart of data mining. Mining for association rules between items in large database of sales transactions has been recognized as an important area of database research. These rules can be effectively used to uncover unknown relationships, producing results that can provide a basis for forecasting and decision making. The original problem addressed by association rule mining was to find a correlation among sales of different products from the analysis of a large set of data. Genetic Algorithms are a family of computational models inspired by evolution. These algorithms encode a potential solution to a specific problem on a simple chromosome-like data structure and apply recombination operators to these structures as to preserve critical information. An implementation of genetic algorithm begins with a population of chromosomes. [17] One then evaluates these structures and allocated reproductive opportunities in such a way that these chromosomes which represent a better solution to the target problem are given more chances to reproduce than those chromosomes which are poorer solutions. The goodness of a solution is typically defined with respect to the current population. [11]

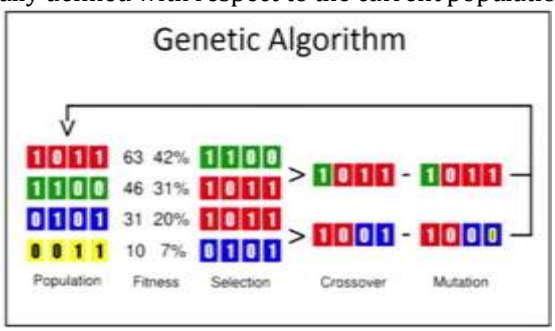

Figure 1: Genetic algorithm process

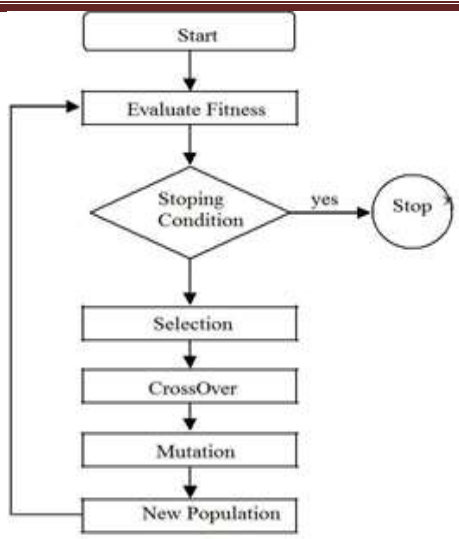

Figure 2: Genetic algorithm flow

There are two association rules mentioned in Example, The first one states that when peanut butter is purchased, bread is purchased 30% of the time.[18] The second one states that 40% of the time when peanut butter is purchased so is jelly. Association rules are often used by retail stores to analyze market basket transactions. [18] The task of mining association rules over market basket data is considered a core knowledge discovery activity. Association rule mining provides a useful mechanism for discovering correlations among items belonging to customer transactions in a market basket database.
Association rules are also used for other applications such as prediction of failure in telecommunications networks by identifying what events occur before a failure. [18]

Let D be the database of transactions and J = {J1...Jn} be the set of items. A transaction T includes one or more items in J (i.e., T⊆ J). An association rule has the form X⇒Y, where X and Y are non-empty sets of items (i.e. X ⊆J, Y⊆ J) such that X ∩ Y = Ø.

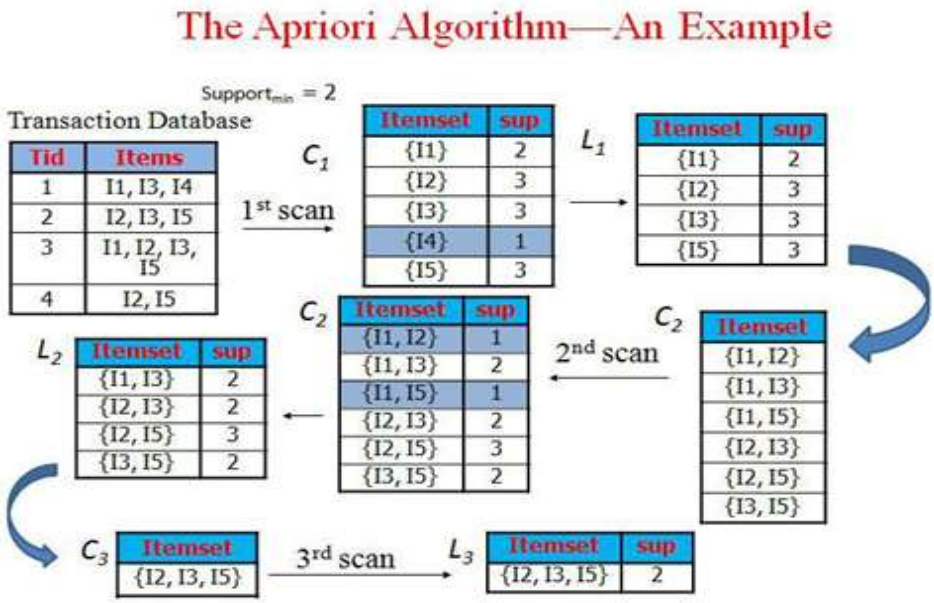

Figure 3: Apriori algorithm process with sample data

Particle swarm optimization (PSO) is a population based stochastic optimization technique inspired by social behavior of bird flocking or fish schooling.
PSO shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA). The system is initialized with a population of random solutions and searches for optima by updating

generations. However, unlike GA, PSO has no evolution operators such as crossover and mutation. In PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles. The detailed information will be given in following sections.[17]

Compared to GA, the advantages of PSO are that PSO is easy to implement and there are few parameters to adjust. PSO has been successfully applied in many areas: function optimization, artificial neural network training, fuzzy system control, and other areas where GA can be applied.

II. Related Work

'Support' and 'Confidence' are measures for Association Rule Mining aimed as extracting rule that identify Co-occurrence of Item sets. This Mining technique contain two phases: First phase one includes extracting 'Frequent-Item-sets' (FISs): Identify total Frequent Item Sets. Second phase includes rules that do not satisfy minimum confidence are removed.

Researchers have characterized this One as an "Optimization problem" and tried to unearth Association Rules using global optimization algorithms. Genetic algorithms are one of the best global optimization algorithms and as our Problem nature – having multiple objectives in ARM; a multi-objective genetic algorithm can be a useful approach.

The main motivation of this project has emerged with the deliberate amount of work still in progress and remaining in order to optimized rule that predict the co-occurrence of item sets but issue regarding rule quality and rule quantity target toward the new combined approach of multi-objective genetic algorithm.

Scope of this research work is, Technique of rule extraction and greedy approach of genetic algorithm make the rule optimal. To predict the co-occurrence of subset : "How Subset influences the presence of another subset "Combine Approach of Genetic Algorithm over rules that states by association rule mining can also predict the rules which contain negative occurrence of attributes. Generate the set of multiple non-dominated solution- As Optimal High level prediction rules.

III. BACKGROUND

The Interesting Multiple Level Minimum Supports(IMLMS) approach aims to find interesting FISs and in FIS. The approach was designed by adopting the original MLMS model with a pruning strategy, which is modified to prune uninteresting item sets in a manner that is suitable to the model.A novel method for optimization of association rule mining. Our proposed algorithm is combination of Interesting Multiple Level Minimum Support (IMLMS) and Genetic Algorithm (GA). We have observed that when we modify the distance weight new rules in large numbers are found. This implies that when weight 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 proved a relation between locally large and globally large patterns that are 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. In MLMS, large no. of rules is generated in comparison of IMLMS. The empirical evaluation of modified algorithm is better in comparison of MLMS and IMLMS in terms of no. of rules are reduced. In the process of IMLMS-GA calculation small number of rule set generated in comparison of MLMS and IMLMS algorithm.

A. Association Rule Mining as Multi-Objective problem

The implication A→C denotes the typical form of an Association Rule, where A and C represent the item-sets, including the possible combination of items {i1, i2, . . , im} in a transactional database, such that A ∩ C = ∅. A is called antecedent, while C is known as consequent, and the rule A→C is interpreted as "If there exists an item -set A in a transaction, then item-set C will co-exist in the same transaction" [15].

Association Rule Mining is inherently a multi-objective problem. One basic measurement in ARM is Support count. Support count of an item-set A, i.e. Support (A), is defined as the ratio of the transactions-count that contain A to the total number of transactions in the database. In the present work, we use the following measures for Association Rule Mining.

1) Confidence: Confidence measures how strong a rule is. Also known as predictive accuracy, Confidence is said to be the conditional probability P(C|A), denoting the fraction of given transactions that contain A, also contain C.

Confidence (A→C) = Support (A∪C) / Support (A) (1)

2) Comprehensibility: One of the key points of Data Mining is Comprehensibility; if the generated rule contains a large number of conditions then the rule gets incomprehensible for the user. Hence for better interpretability, Comprehensibility is measured, which makes sure that the attribute count in the Antecedent part is less than that of the Consequent part [1], [16].

Comprehensibility $(A \rightarrow C) = \log(1 + |C|) / \log(1 + |A \cup C|)$ (2)

Where, $|C|$ and $|A \cup C|$ are the attribute counts in the consequent and in the entire rule, respectively.

3) Interestingness: In literature, [16] and [17], are few of the many works that mention that all high-confidence rules may not be interesting. For example, considering the case of Market Basket Analysis, the rule $A \rightarrow C$ may have high confidence, because C is just purchased very often (independent of A), resulting into although high confidence but less interestingness rule. Some rules are so obvious, that they can be easily predicted by the user, but to employ the 'unexpected and hidden' features of the Data Mining task, the parameter of interestingness is employed.

Interest $(A \rightarrow C) = |(\text{Support}(A \cup C) / \text{Support}(A)) - \text{Support}(C)|$ (3)

Where, first part gives confidence of the given rule, and second part calculates the probability of consequent.

Algorithm Multi-Objective PSO based on

Input: N = Population size, M = Maximum Generation

```
Initialise Swarm Population
Generate random Swarm Particles
Determine Domination
Initialise External Repository
for each Generation do
for each Particle do
Select Leader from External Repository
Roulette Wheel Selection
Update particle-position and velocity
Evaluate Objective Functions
Update Pbest
end for
Determine Domination
Update External Repository
end for
return External Repository
```

IV. Proposed Flow and Algorithm

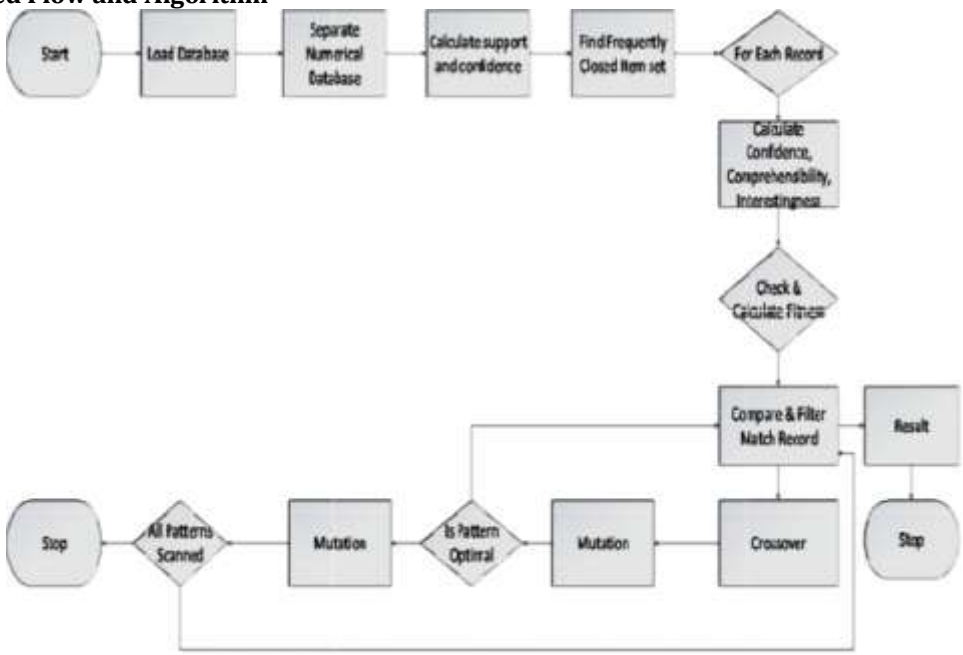

Fig: 4 Flow Chart of Proposed Algorithm

Flow chart presented above shows the process of proposed algorithm.

Proposed Algorithm

1. Load Numerical Dataset D.
2. Compute(SUP)
3. Find frequent item sets for Numerical Dataset using QauntMiner.
4. While (Num. of Iterations is less than X) Do
5. Confidence = SUP(AUC) / SUP(A)
6. Comprehensibility = log (1+ |C|) / log(1+|AUC|)
7. Interestingness = [SUP(AUC) / SUP(A)] x [SUP(AUC) / SUP(C)] x [1-SUP(AUC)/ |D|]
8. Compute(fitness)
9. Add to Non-Dominated solutions
10. Remove from Dominated Solutions
11. New Population = Crossover(Current populations)
12. Mutate(New Population)
13. End While.

V. Result and Analysis

I have used NetBeans and Eclipse software for simulation. For coding used JAVA programming language and SPMF API/JAR files.The comparison analysis of fitness calculated for training , max validation for validation purpose. There are different scenarios to check while experiments for the results of the proposed approach after implementation. I have tested my proposed research over different rules and found the above stated result which are shown promising.

VI. Conclusion

In this approach, we have proposed a new multi objective optimization model as an association rule miner, which is based on hybrid algorithm of the two population based naturally inspired global optimization algorithms which are NSGA-III and MOPSO.In this proposed algorithm multi objective genetic algorithm approach used for mining association rules. Through this proposed algorithm obtained more accurate rules. Through this proposed algorithm generated rules should also be interesting and comprehensible. Provide efficient solution to problems. Through use of rank based fitness calculation gives solution to problem in which fitness value differ very much. Provide powerful and robust optimization technique. Achieve more efficient accuracy in result and maintain a high confidence and a best coverage of the database and providing user with accurate rules.

References

1. ArvindJaiswal and GauravDubey, “Identifying Best Association Rules and Their Optimization Using Genetic Algorithm”, International Journal of Emerging Science and Engineering (IJESE), ISSN:2319-6378, Volume-1, Issue-7, May 2013, 91-96.
2. Tan, Steinbach, and Kumar, “Introduction to Data Mining”, http://www_users_cs.umn.edu/~kumar/dmbook/dmslides/chap1_intro.pdf.
3. RupaliHaldulakar and Prof. JitendraAgrawal, “Optimization of Association Rule Mining through Genetic Algorithm”, International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 3 Mar2011, 1252-1259.
4. Manish Saggar, Ashish Kumar Agarwal and Abhimunya Lad, “Optimization of Association Rule Mining using Improved Genetic Algorithms”, IEEE 2004.
5. Anandhavalli 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. Farah Hanna AL-Zawaidah, YosefHasanJbara and Marwan AL-Abed Abu-Zanona, “An Improved Algorithm for Mining Association Rules in Large Databases”, World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 1, No. 7, 2011, 311316
7. Peter P. Wakabi-Waiswa and Dr. VenansiusBaryamureeba, “Extraction of Interesting Association Rules Using Genetic Algorithms”, Advances in Systems Modelling and ICT Applications, 101-110.
8. M. Ramesh Kumar and Dr. K. Iyakutti, “Genetic algorithms for the prioritization of Association Rules”, IJCA Special Issue on “Artificial Intelligence Techniques - Novel Approaches & Practical Applications” AIT, 2011, 35-38.

9. S. Dehuri,A. K. Jagadev, A. Ghosh and R. Mall, "Multi-objective Genetic Algorithm for Association Rule Mining Using a Homogeneous Dedicated Cluster of Workstations", American Journal of Applied Sciences 3 (11), 2006, 2086-2095.
10. AnsafSalleb-Aouissi, ChristelVrain and Cyril Nortet, "QuantMiner: A Genetic Algorithm for Mining Quantitative Association Rules", IJCAI2007, 1035-1040.
11. R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules", Proceedings of 20th International Conference on Very Large Databases, Santiago, Chile, 1994.
12. R. Agrawal, I. Imielinski and A. Swami, "Mining Association Rule between Sets of Items in Large Databases", ACM, 1993, 207-216.
13. S. Thomas and S. Sarawagi, "Mining Generalized Association Rules and Sequential Patterns using SQL Queries", 4th International Conference on Knowledge discover and Data Mining, 1998.
14. R. Shrikant and R. Agrawal, "Mining Quantitative Association Rule in Large Relational Tables", SIGMOD, 1996.
15. "Introduction to Genetic Algorithm" http://www.obitko.com/tutorials/genetic_algorithms/crossover_mutation.ph p.
16. Peter P. Wakabi-Waiswa and VenansiusBaryamureeba, "Extraction of Interesting Association Rules using Genetic algorithms", International Journal of Computing and ICT Research, Vol. 2 No. 1, June 2008, 26-33.
17. SourmadipGhosh, SushantaBiswas, DebasreeSarkar and ParthaPratimSarkar, "Mining Frequent Itemsets using Genetic Algorithm", International Journal of Artificial Intelligence and Applications (IJAIA), Vol. 1, No. 4, October 2010, 133-143.
18. R. Agrawal, I. Imielinski and A. Swami, "Mining Association Rule between Sets of Items in Large Databases", ACM, 1993, 207-216.
19. S. Thomas and S. Sarawagi, "Mining Generalized Association Rules and Sequential Patterns using SQL Queries", 4 th International Conference on Knowledge discover and Data Mining, 1998.
20. R. Shrikant and R. Agrawal, "Mining Quantitative Association Rule in Large Relational Tables", SIGMOD, 1996.
21. Peter P. Wakabi-Waiswa and VenansiusBaryamureeba, "Extraction of Interesting Association Rules using Genetic algorithms", International Journal of Computing and ICT Research, Vol. 2 No. 1, June 2008, 26-33.
22. SourmadipGhosh, SushantaBiswas, DebasreeSarkar and ParthaPratimSarkar, "Mining Frequent Itemsets using Genetic Algorithm", International Journal of Artificial Intelligence and Applications (IJAIA), Vol. 1, No. 4, October 2010, 133-143.