Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 5, May 2016, pg.268 – 272

MOGA FOR MULTILEVEL FUZZY ASSOCIATION RULE WITH MSFM APPROACH

Jinal J. Shah[1], Lokesh P. Gagnani[2]

[1]I.T. & Gujarat Technological University, India
[2]I.T. & Gujarat Technological University, India
[1] jincy.1111@gmail.com; [2] gagnani.lokesh@gmail.com

Abstract— Association rule mining is the most popular technique in the area of data mining. The main task of this technique is to find the frequent patterns by using minimum support thresholds decided by the user. The Apriori algorithm is a classical algorithm among association rule mining techniques. This algorithm is inefficient because it scans the database many times. Second, if the database is large, it takes too much time to scan the database. For many cases, it is difficult to discover association rules among the objects at low levels of abstraction. Association rules among various item sets of databases can be found at various levels of abstraction. Apriori algorithm does not mine the data on multiple levels of abstraction. In this paper we introduce detail description about multi level association rule, Fuzzy ARM, Rule based, Multi objective and genetic algorithm.

Keywords— Multilevel ARM, Fuzzy ARM, Fuzzy Logic, Rule based, Multi objectives, Genetic Algorithm

I. INTRODUCTION

MLAR

For many cases, it is difficult to discover association rules among the objects at low levels of abstraction. Association rules among various item sets of databases can be found at various levels of abstraction. The rules which are generated by mining the data at multiple levels of abstraction are called Multi-level association rules. Association rules discovered at high levels represent common sense knowledge. Multi-level association rules can be discovered efficiently using concept hierarchies which satisfy minimum support-confidence framework.

In general, top-down progressive deepening method is used for mining Multi-level association rules, where support is counted to generate frequent item sets at each level starting from the level 1 and moving downward in the concept hierarchy, until no more concept items can be found. Thus one might be interested in discovering frequent item sets composed of items which themselves form taxonomy. Many algorithms in literature discussed this problem. Multilevel databases use encoded transaction table driven by using concept hierarchy information instead of the original transaction table.

© 2016, IJCSMC All Rights Reserved

FUZZY LOGIC

"Fuzzy logic may be viewed as an extension of multivalued logic. Its uses and objectives, however, are quite different. Thus, the fact that fuzzy logic deals with approximate rather than precise modes of reasoning implies that, in general, the chains of reasoning in fuzzy logic are short in length and rigor does not play as important a role as it does in classical logical systems. In a nutshell, in fuzzy logic everything, including truth, is a matter of degree."

Fuzzy logic deduces its greater expressive power from including probability theory and probabilistic logic. The main differences between traditional logic and fuzzy logic are the following:

- Speaking of two-valued logic, a proposition is either true or false. Fuzzy logic allows truth values to range over fuzzy subsets. Therefore, the fuzzy truth value could be viewed as an imprecise characterization of a numerical truth value.
- Fuzzy logic allows crisp predicates, as in two-valued logic, but also fuzzy ones, for example "big", "tall" or "beautiful".
- Two-valued logic allows only the two quantifiers "all" and "some" whereas fuzzy logic allows the use of quantifiers like "most", "many", "several", "few" and so on. These fuzzy quantifiers can be viewed as a second order fuzzy predicate.
- Both fuzzy and and non-fuzzy predicate-modifiers can be represented by fuzzy logic. This leads to a system which enables computing with linguistic variables, i.e. variables whose values are words and expressions from a natural or synthetic language.
- In two-valued logic, a proposition can be qualified by associating it with a truth value ("true" or "false"), a modal operator (such as "possible" or "necessary") "or an intentional operator (such as "know" or "believe").

FUZZY ARM

Based on classical association rule mining, a new approach has been developed expanding it by using fuzzy sets. The new fuzzy association rule mining approach emerged out of the necessity to mine quantitative data frequently present in databases efficiently. Algorithms for mining quantitative association rules have already been proposed. We are confronted with the sharp boundary problem.

II. MULTI OBJECTIVE

In many real-life problems, objectives under consideration conflict with each other. Hence, optimizing x with respect to a single objective often results in unacceptable results with respect to the other objectives. Therefore, a perfect multi-objective solution that simultaneously optimizes each objective function is almost impossible. A reasonable solution to a multi-objective problem is to investigate a set of solutions, each of which satisfies the objectives at an acceptable level without being dominated by any other solution. If all objective functions are for minimization, a feasible solution x is said to dominate another feasible solution y (x _ y), if and only if, zi(x)pzi(y) for i ¼ 1, y, K and zj(x)ozj(y) for least one objective function j. A solution is said to be Pareto optimal if it is not dominated by any other solution in the solution space.

A Pareto optimal solution cannot be improved with respect to any objective without worsening at least one other objective. The set of all feasible non-dominated solutions in X is referred to as the Pareto optimal set, and for a given Pareto optimal set, the corresponding objective function values in the objective space are called the Pareto front. For many problems, the number of Pareto optimal solutions is enormous (perhaps infinite). The ultimate goal of a multi-objective optimization algorithm is to identify solutions in the Pareto optimal set. However, identifying the entire Pareto optimal set, for many multi-objective problems, is practically impossible due to its size. In addition, for many problems, especially for combinatorial optimization problems, proof of solution optimality is computationally infeasible. Therefore, a practical approach to multi-objective optimization is to investigate a set of solutions (the best-known Pareto set) that represent the Pareto optimal set as well as possible. With these concerns in mind, a multi-objective optimization approach should achieve the following three conflicting goals:

1. The best-known Pareto front should be as close as possible to the true Pareto front. Ideally, the best-known Pareto set should be a subset of the Pareto optimal set.
2. Solutions in the best-known Pareto set should be uniformly distributed and diverse over of the Pareto front in order to provide the decision-maker a true picture of trade-offs.
3. The best-known Pareto front should capture the whole spectrum of the Pareto front. This requires investigating solutions at the extreme ends of the objective function pace.

© 2016, IJCSMC All Rights Reserved

III. GENETIC ALGORITHM

Genetic Algorithm is a search heuristic that mimic the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic Algorithm is based on ideas of evolution theory (Holland, 1975) as key principle is that only the fittest entities survive. 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 databases, biology and fraud detection event sequential analysis.

Genetic Algorithm work in this flow:
The functions of genetic operators are as follows:-

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

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.

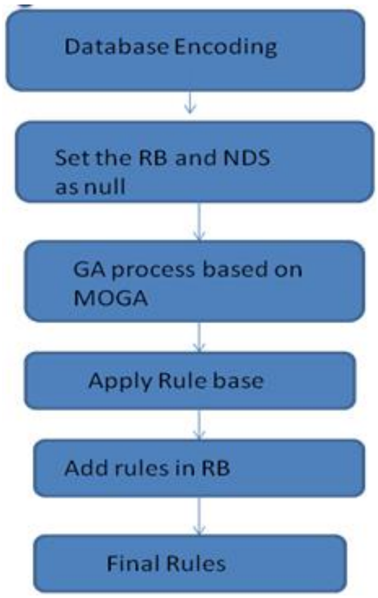
3) Mutation: Mutation changes randomly the new offspring. For binary encoding we can switch a few randomly chosen bits from 1 to 0 or from 0 to 1[9].

IV. RULE BASED APPLICATION

The best rule to be inserted in the rule base in each iteration is selected among the non dominated solutions, using a criterion related to the accuracy of the rule base**.** A recent trend in the scientific community has been to generate the FRBS by means of Multi-objective Genetic Algorithms (MOGA), where the accuracy is measured in terms of error prediction and the interpretability is evaluated by means of indexes related to the complexity and/or the semantic meaning of the Rule Base (RB) or the Data Base (DB). This way, the task of searching for a balance between both objectives is embedded in the genetic process. IRL approach, the rules are generated and inserted in the rule base one at a time, after each execution of the MOGA. Since the MOGA provides a set of non-dominated solutions, a criterion to select the rule that is going to be inserted in the rule base is necessary.
Best rule selection and fitness functions definition

Accuracy= Integrity X Consistency

V. PROPOSED APPROACH



© 2016, IJCSMC All Rights Reserved

Algorithm Steps:

1. Read Transaction Data
2. Encode Items names with *
3. Decide k-Level Representation

   Group first k-Digit Items of each Transactions

   Repeat until k-level Items founded.
4. Decide RB (Rule Base) as NULL
5. Set NDS (non-dominated set) to null.
6. Calculate the First Object Values for Each Chromosome using suitability(Cq) formula.
7. Calculate the Second Objective Values Number of Large 1-ItemSets (numLarLevel(Cq)

   For each Level

   Create Fuzzy Regions

   Calculate cardinality for each Fuzzy Regions

   Check Minimum Support Value against Threshold

   If Large then Thresh hold then place it to Large DataSet

   Set Summation of each large 1-itemset as Second ObjectiveValue of Chromosome.
8. Rank the Chromosome as per two Objectives.
9. Calculate the fitness value of Chromosome as per Rank.
10. Calculate the Average fitness value of chromosome for same fitness value for Selection.
11. Copy the chromosomes with their ranking values equal to one into the non-dominated set NDS and remove the chromosomes which are dominated by other chromosomes in NDS.
12. Execute the crossover operation on the population.
13. Execute the mutation operation on the population
14. Calculate the fitness values of the new chromosomes by STEPs 5 to 9
15. Get the Rules from NDS
16. Add Rules to RB

VI. CONCLUSION

This paper, we have defined fuzzy set concepts, multiple level taxonomy and different minimum supports for each level with multi objective function and rule based mechanism and find association rules in a given transaction data set. The proposed MOGA fuzzy mining algorithm with multiple support value with rule based can thus generate large itemsets level by level and then get final rules in RB from NDS with accurate rules.

REFERENCES

[1] S. Venkata Krishna Kumar, P. Kiruthika, "A Survey on Association Rule Mining"IJARCSSE,2015

[2] Anand V. Saurkar, S. A. Gode," Association Rule Mining with Fuzzy Logic: an Overview",IJSR,2013

[3] Jun TAN," Different Types of Association Rules Mining Review"AMM,2013

[4] Sonia Setia , Dr. Jyoti," Multi-Level Association Rule Mining: A Review",IJCTT,2013

[5] Md. IQBAL," Performance Analysis of Data Mining Algorithms to Generate Frequent itemset at Single and Multiple Levels" ,OJCST,2012

[6] Sandeep Kumar Singh, Mr.Ganesh Wayal, Mr.Niresh Sharma," A Review: Data Mining with Fuzzy Association Rule Mining",IJERT,2012

© 2016, IJCSMC All Rights Reserved

[7] Chun-Hao Chen, Chi-Hsuan Ho, Tzung-Pei Hong, Wei-Tee Lin," MOGA for Multi-Level Fuzzy Data Mining",IEEE,2012

[8] Chun-Hao Chen, Tzung-Pei Hong, Vincent S. Tseng,and Lien-Chin Chen," MULTI-OBJECTIVE GENETIC-FUZZY DATA MINING",ICIC,2012

[9] Shaikh Nikhat Fatma, Jagdish W Bakal, Madhu Nashipudimath," Integrated Genetic-Fuzzy Approach for Mining Quantitative Association Rules"ICACACT,2012

[10] Chun-Hao Chen, Tzung-Pei Hong, Yeong-Chyi Lee," A Multiple-Level Genetic-Fuzzy Mining Algorithm",IEEE,2011

[11] Rajul Anand Abhishek Vaid Pramod Kumar Singh," Association Rule Mining Using Multi-objective Evolutionary Algorithms: Strengths and Challenges",IEEE,2009

[12] Chun-Hao Chen, Tzung-Pei Hong, Vincent S. Tseng, and Chang-Shing Lee,"A Genetic-Fuzzy Mining Approach for Items with Multiple Minimum Supports",IEEE,2007

[13] M. Kaya, R. Alhajj," Genetic algorithm based framework for mining fuzzy association rules" Elsevier,2004

[14] Edward Hinojosa c., Heloisa A. Camargo," Multi-objective Iterative Genetic Approach for Learning Fuzzy Classification Rules with Semantic-based Selection of the Best Rule",IEEE,2013

© 2016, IJCSMC All Rights Reserved