Volume 7, Issue 5, May 2017 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Multi-objective Association Rule Mining using Evolutionary Algorithm

Hemant Kumar Soni
Department of Computer Science and Engineering, Amity School of Engineering and Technology,
Amity University, Gwalior Madhya Pradesh, India

Abstract—Generally association rule mining (ARM) algorithms, like the apriori algorithm, initial produce frequent itemsets and afterward, from the frequent itemsets, the association rules that go beyond the minimum confidence threshold. When the data is in large volume, it takes number of scans to generate frequent items.It is a better idea if all the association rules generated directly without generating frequent items and reduce number of scanning of the database. The quality of an association rule cannot only be signified by its support or confidence. There are numerous other metrics existing to determine the quality of an association rule. Then, the concept of ARM can be present as a multi-objective optimization problem in which the objective is to find association rules while optimizing a number of such goodness and quality criteria at the same time. This point of view, evolutionary algorithms have been utilizing extensively for producing association rules. In this paper, in-depth study on various objectives for ARM and some evolutionary algorithms has been done.

Keywords— Association Rule Mining, interestingness, comprehensibility, lift, interest factor.

I. INTRODUCTION

In Data mining significant data dig out from a huge database repository. It is the progression of selection of significant information from big amount of data by using convinced sophisticated algorithms. Data mining is becoming a gradually more vital tool to convert data into valuable information which facilitate in decision making.
The Main objective of data mining is to find out the new, unknown and unpredictable information from the used database, which is useful and helps in decision making. There are a number of techniques used in data mining to identify the frequent pattern and mining rules includes clusters analysis, anomaly detection, association rule mining etc.

II. ASSOCIATION RULE MINING

Data mining's prime techniques include generating association rules. The association rules mining (ARM) first introduced in [1-4]. The objective of data mining is to fetch appealing correlation, regular pattern, and relation amongst items in the transaction database.In association rule mining, only those items, whose minimum support and confidence above the threshold are taken as an association rule. The entire association rule mining process is divided into two subsections. In first subsection, the process identified those items whose repetition is above the pre-decided threshold. These itemsets called frequent or large itemsets. The second subsection of the process identified association rule by large items sets who qualify minimum confidence [5, 6]. There are number of algorithms are available which works on the basis of Tree-based and Apriori based to generate ARM[7]. Data mining techniques use in various domain including time-series data analysis, e-commerce etc[8].

III. BASIC CONCEPTS

Association rule mining can be defined as : Let there are number of items like $I = \{i_1, i_2, \ldots i_m\}$. Let D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. For a transaction T, which is a set of items, such that $T \subseteq I$. Every transaction has an exclusive identifier called TID. In this concept say that a transaction T contains X, a set of some items (called itemset) in I if $X \subseteq T$.

3.1 Association Rules

For a given transaction database T, An association rule is an implication of the form $X \Rightarrow Y$, where $X \subset I, Y \subset I, and\ X \cap Y = \Phi$, i.e. X and Y are two non-empty and non-intersecting itemsets. The rule $X \Rightarrow Y$ holds in the transaction set D with confidence c if c % of transactions in T that contain X also contain Y.

3.2 Support

If any item i_k, present in Transaction T, then T is said to support a subset of items $X \subseteq I$, if T support each item i_k in X. An itemset $X \subseteq I$ have a support s in D, denoted by s(X), if s% of transactions in D support X.

$$Support\ (X \Rightarrow Y) = Support\ (X \cup Y) / |D| \quad \text{------ (1)}$$

3.3 Confidence

The confidence of rule $X \Rightarrow Y$ is the fraction of transactions in D containing X that also contain Y and indicates the strength of rule.

$$(X \Rightarrow Y) = Support\ (X \cup Y) / Support\ (X) \quad \text{------ (2)}$$

IV. MULTIOBJECTIVE ASSOCIATION RULE MINING

The interesting, relevant and useful association rule can not be identified by using only support and confidence. There are other parameters are also available to determine appropriate association rules. For that reason , ARM can be looking as a multiobjective optimization problem, where the objective to find out the association rules who fulfill other decisive factor concurrently. Even though support and confidence are two major objectives to identify association rules but there are some others factors are also important to find out interesting, relevant and useful association rules.

Other than Support and confidence, some other objectives such as Comprehensibility, Coverage, Cosine, lift, Laplace, Jaccard, J-measure, prevalence, surprise, recall, conviction, surprise and so on are also available in the literature[9,10]. Author also proposed two novel objectives for high and low correlation for 2 variables and 3 variables [11].

4.1 Comprehensibility

Comprehensibility of an association rule is quantified by the following expression:

$$Comprehensibility = \frac{\log\ (1+|Y|)}{\log\ (1+|X \cup Y|)} \quad \text{------ (3)}$$

where |itemset| denote the attributes exist in the itemset.

In other words, if the number of conditions in the antecedent part is less, the rule is more comprehensible.

4.2 Interestingness

Interestingness determine how much the rule is surprising for the user. The main objective of association rule mining is to find unknown information, it extract rules that have relatively less occurrence in the database. Interestingness is quantified as follows:

$$Interistngness = \left(\frac{Support\ (X \cup Y)}{Support\ (X)}\right) \times \left(\frac{Support\ (X \cup Y)}{Support\ (Y)}\right) \times \left(1 - \frac{Support\ (X \cup Y)}{Support\ (Z)}\right) \quad \text{------ (4)}$$

where Support(Z) is the number of records in the database [13,17].

However, most researchers have adopted Piatetsky-Shapiro's [14] argument that a rule cannot be interesting, if its antecedent and consequent are statistically independent.

4.3 Lift

Lift compute the ratio between the rule's confidence and the support of the itemset in the rule consequent. Lift is equivalent to the ratio of the observed support to that expected if X and Y were statistically independent.

$$Lift = \frac{Confidence\ (x \rightarrow y)}{(Support\ (y))} \quad \text{------ (5)}$$

4.4 Interest Factor

For binary variables, lift is equivalent to another objective called interest factor which is defined as follows :

$$I(A,B) = \frac{S(A,B)}{S(A)\ X\ S(B)} = \frac{N\ f11}{f(1+)\ f(+1)} \quad \text{------ (6)}$$

V. MULTIOBJECTIVE EVOLUTIONARY ALGORITHM

Some research has been done to find out multiobjective association rule mining, by using evolutionary algorithms and soft computing techniques. In literature MOO tool used and compare encoding techniques , objective functions , evolutionary operators and methods to acquire solution. MOGA([12,13], PARETO BASED COEVLOUTIONARY[14], NSGA II[15- 17], MODE[18] SPEA VARIENT [19-21] are some examples of the MOEA based association rule mining. [22] explain that most of the methods have used a Michigan encoding and thus all the non-dominated solutions are treated as final solutions without needing a particular solution from the set.

VI. CONCLUSION

In literature different techniques have been describe. MODENAR AND NSGA – II QAR measure up to rule interestingness metrics . Except these two methods, all other techniques compare their performance with regard to single objective evolutionary and other non-evolutionary techniques.

It is established that association rules must be based on multionjective parameters and the result of this, MOEA based ARM Algorithms fetch attention in last few years. Literature shows that all the experiments and results are based on UCI repository datasets. The implementation of these techniques on real database is required.

Although there is a need of MOEA based efficient technique for real life application and some applications in areas like biological data, text and web mining, financial data, temporal database and others.

REFERENCES

[1] R. Agrawal ,T. Imielinski and A. Swami, "Mining Association Rules between sets of Items in large Databases". *In Proc. Int. Conf. Management of Data (SIGMOD'93)*,1993, p 207-216.

[2] R. Agarwal and R. Srikant, "Fast Algorithm for Mining Association Rules". *In Proc. of Int. conf. Very Large Data Bases (VLD'94)*, 1994, p 487-499.

[3] M. Houtsma and A. Swami. "Set-Oriented Mining of Association Rules". Research Report RJ 9567, IBM Almaden Research Centre ,San Jose ,California, Oct.'93.

[4] H. K. Soni, S. Sharma, P.K. Mishra. "Association Rule Mining : A data profiling and prospective approach". *International Journal of Current Engineering and Scientific Research*. Vol. 3, pp. 57-60, 2016.

[5] Sotiris Kotsiantis, Dimitris Kanellopoulos, "Association Rules Mining: A Recent Overview", *GESTS International Transactions on Computer Science and Engineering*, Vol.32 (1), pp. 71-82, 2006.

[6] H. K. Soni, S. Sharma, M. Jain. "Frequent Pattern Generation Algorithms for Association Rule Mining : Strength and Challenges". *In proceedings of IEEE International Conference on Electrical, Electronics and Optimization Techniques (ICEEOT)*, 2016, p. 3744-3747.

[7] M. Kumar, H. K. Soni. "A Comparative Study of Tree- based and Apriori-based Approaches for incremental Data Mining" . *International Journal of Engineering Research in Africa*, vol. ,23,pp 120-130, 2016.

[8] Neelam Mishra, Hemant Kumar Soni, Sanjiv Sharma. "A Comprehensive Survey of Data Mining Techniques on Time Series Data for Rainfall Prediction". Submitted to *Journal of ICT Research and Applications.*

[9] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1650–1654, Dec. 2002.

[10] P.-N. Tan, V. Kumar, and J. Srivastava, "Selecting the right interestingness measure for association patterns," in *Proc. 8th ACM SIGKDD Int. Conf. KDD*, 2002, p. 32–41.

[11] Hemant Kumar Soni, S. Sharma, A.K. Upadhyay, "Two novel pioneer objectives of association rule mining for high and low correlation of 2-variables and 3-variables", *International Journal of Engineering and Technology",* vol. 9, No.2, 2017, pp. 695-703.

[12] K. Kim, R. B. McKay, and B.-R. Moon, "Multiobjective evolutionary algorithms for dynamic social network clustering," in *Proc. 12th Ann. Conf. GECCO*, 2010, p. 1179–1186.

[13] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.

[14] G Piatetsky-Shapiro, *Discovery, analysis, and presentation of strong rules*. In: G. Piatetsky-Shapiro, W. Frawley (eds.) Knowledge Discovery in Databases, AAAI/MIT Press ,1991, p 229–248.

[15] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the Strength Pareto Evolutionary Algorithm," in *Proc. EUROGEN, 2001*, p. 95–100.

[16] J. Horn and N. Nafpliotis, "Multiobjective Optimization using the Niched Pareto Genetic Algorithm," Univ. Illinois at Urbana- Champaign, Champaign, IL, USA, Tech. Rep. IR-93005, 1993.

[17] A. Mukhopadhyay, S. Bandyopadhyay, and U. Maulik, "Clustering using multiobjective genetic algorithm and its application to image segmentation," in *Proc. IEEE Int. Conf. SMC*, vol. 3, p. 2678–2683, 2006.

[18] T. O¨ zyer, Y. Liu, R. Alhajj, and K. Barker, "Multi-objective genetic algorithm based clustering approach and its application to gene expression data," in *Proc. 3rd Int. Conf. ADVIS*, 2004, p. 451–461.

[19] G. N. Demir, A. S. Uyar, and S. G. O¨ gu¨du¨cu¨, "Multiobjective evolutionary clustering of web user sessions: A case study in web page recommendation," *Soft Comput.*, vol. 14, no. 6, pp. 579–597, 2010.

[20] J. Handl and J. Knowles, "Exploiting the trade-off—The benefits of multiple objectives in data clustering," in *Proc. 3rd Int. Conf. EMO*, 2005, p. 547–560.

[21] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, NY, USA: Plenum, 1981.

[22] Anirban Mukhopadhyay Ujjwal Maulik, Sanghamitra Bandyopadhyay and Carlos A. Coello Coello, "Survey of Multiobjective Evolutionary Algorithms for Data Mining: Part II", *IEEE Transactions on Evolutionary Computation*, vol.. 18, NO. 1, Feburary 2014 ,pp 24 – 35.

