International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Review of Multi-objective Evolutionary Algorithm

Sheetal Ashok Naredi

JSPMs RSCOE Tathwade , Pune, India

Abstract: ***Datamining is an analytic process designed to explore data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data instructions***

Keywords: Mining, Association Rule, Apriori, Association rule mining**,** Closed Patterns

1. Introduction

Data mining is the extraction of implicit, valid, and potentially useful knowledge from large volumes of raw data. The extracted knowledge must be not only be accurate but also readable, comprehensible and ease of understanding. Association rule (AR) mining is one of the important research problems in data mining field where the goal is to derive multi-feature (attribute) correlations from databases. Data mining functions include clustering, classification, prediction, and link analysis (associations). One of the most important data mining applications is that of mining association rules. The objective of this paper is to provide a thorough survey of previous research on association rules mining.

2. Methods of Data Mining

Major The traditional association rules mining algorithm "Apriori", mainly, supports to solve two problems, one is how to reduce the number of candidate frequent item sets and the times of browsing database, the other is how to easily generate candidate frequent item sets and fleetly compute the support of candidate frequent item sets.

In order to solve the first problem, some algorithms were presented, such as Max-Miner [2], Pincer-Search [3], DMFI [4] and DMFIA [5]. And then, in order to solve the second problem, some algorithms, based on binary were presented, such as B-Apriori [6]. The type of algorithm computes support by binary logic operation to improve efficiency. However, the algorithm is merely suitable for mining relative short frequent item sets, if frequent item sets are too long, the efficiency of "B_Apriori" will be badly influenced. Another algorithm based on up-down [3] search strategy, denoted by B_DMFIA, is only suitable for mining long frequent item sets, too.

Association rule mining [6] plays a key role in boosting the research, development and application of data mining techniques. This leads to a many significant technologies and methodologies for identifying association rules. These techniques mainly focus on algorithms for scale-up and data reduction from the efficiency point of view. However, these mining algorithms are mostly based on the assumption that the users can specify the minimum support appropriate to their databases, and thus referred to as the Apriori-like algorithms [7] have pointed out that setting the minimum support is quite subtle, which can hinder the widespread applications of these algorithms.

Our own experiences of mining transaction databases also tell us that the setting is by no means an easy task. Recognizing the above limitation, a great many techniques have been developed to attack this issue. Han et al. [9] designed a strategy to mine top-k frequent closed patterns for effectiveness and efficiency.

Piatetsky Shapiro and Steingold [10], proposed a method to identify only the top 10% or 20% of the prospects with the highest score for marketing.

Roddick and Rice [11] presented the independent thresholds and context dependent thresholds to measure time-varying interestingness of events for temporal data.

Hipp and Guntzer [12] explored a new mining approach that postpones constraints from mining to evaluation.

Wang, He, Cheung, and Chin [13] designed a confidence-driven mining strategy without minimum support to identify new patterns.

Cheung and Fu [14] developed a technique to identify frequent itemsets without the support threshold.

Zhang et al. [15] advocated a fuzzy-logic-based method to acquire user threshold of minimum support for mining association rules. However, most of these approaches attempt to avoid specifying the minimum support.
It has been pointed out that the usual framework to assess association rules, based on support and confidence as measures of importance and accuracy, has several drawbacks. In particular, the presence of items with very high support can lead to obtain many misleading rules, even of the order of 95% of the discovered rules in some of our experiments.

This paper introduces a different framework, based on Shortliffe and Buchanan’s certainty factors and the new concept of very strong rules. The paper also, discusses some intuitive properties of the new framework. Both the theoretical properties and the experiments, we have performed show that it is possible to avoid the discovery of misleading rules, improving the manageability and quality of the results [16]

Licensed Under Creative Commons Attribution CC BY

References

[1] Han J, Kamber M (2001) Data mining: concepts and techniques. Morgan Kaufmann, Academic, San Fransisco, New York
[2] J. Han and M. Kamber,Data Mining: Concepts and Techniques, 2nd ed.Burlington, MA, USA: Morgan Kaufmann, 200
[3] C. Zhang and S. Zhang, "Association rule mining: Models and algorithms," in Lecture Notes Computer Science (LNAI 2307). Berlin/Heidelberg, Germany: Springer-Verlag, 2002
[4] R. Srikant and R. Agrawal, "Mining quantitative association rules in large relational tables," in Proc. ACM SIGMOD, 1996, pp. 1–12
[5] K. Sun and B. Fengshan, "Mining weighted association rules without preassigned weights,"IEEE Trans. Knowl. Data Eng., vol. 20, no. 4, pp. 489–495, Apr. 2008
[6] C. Silverstein, S. Brin, and R. Motwani, "Beyond market baskets: Generalizing association rules to dependence rules," Data Mining Knowl. Discovery, vol. 2, no. 1, pp. 39–68, 1998
[7] Agrawal, R., Imielinski, T., & Swami, T. (1993). Mining association rules between sets of items in large databases. InProceedings of ACM SIGMOD international conference on management of data (SIGMOD'93) (pp. 207–216)
[8] Zhang, C., & Zhang, S. (2002).Association rules mining: Models and algorithms. Lecture notes in computer science(Vol. 2307, p. 243). Springer-Verlag
[9] Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. InProceedings of the ACM SIGMOD international conference on management of data (SIGMOD)(pp. 1–12)
[10] Piatetsky-Shapiro, G., & Steingold, S. (2000). Measuring lift quality in database marketing.SIGKDD Explorations, 22, 76–80
[11] Roddick, J. F., & Rice, S. (2001). What's interesting about cricket? – On thresholds and anticipation in discovered rules.SIGKDD Explorations, 31, 1–5
[12] Hipp, J., & Guntzer, U. (2002). Is pushing constraints deeply into the mining algorithms really what we want?SIGKDD Explorations, 41, 50–55
[13] Wang, K., He, Y., Cheung, D., & Chin, F. (2001). Mining confident rules without support requirement. InProceedings of the 10th ACM international conference on information and knowledge management (CIKM 2001), (pp 89–96)
[14] Cheung, Y., & Fu, A. (2004). Mining frequent itemsets without support threshold: With and without item constraints. IEEE Transactions on Knowledge and Data Engineering(available online)
[15] Zhang, S., Lu, J., & Zhang, C. (2004). A fuzzy-logic-based method to acquire user threshold of minimum-support for mining association rules. Information Sciences(available online)
[16] Spain.E. Shortliffe and B. Buchanan, "A model of inexact reasoning in medicine," *Math. Biosci.*, vol. 23, nos. 3–4, pp. 351–379, 1975

Licensed Under Creative Commons Attribution CC BY