Chapter 7
Fuzzy Multi-Objective Association Rule Mining Using Evolutionary Computation

Ganghishetti Pradeep
IDRBT, India

Vadlamani Ravi
Institute for Development and Research in Banking Technology, India

ABSTRACT

In this chapter, we model association rule mining as a Fuzzy multi-objective global optimization problem by considering several measures of strength such as support, confidence, coverage, comprehensibility, leverage, interestingness, lift and conviction by utilizing various fuzzy aggregator operators. In this, pdel, each measure has its own level of significance. Three fuzzy multi-objective association rule mining techniques viz., Fuzzy Multi-objective Binary Particle Swarm Optimization based association rule miner (FMO-BPSO), a hybridized Fuzzy Multi-objective Binary Firefly Optimization and Threshold Accepting based association rule miner (FMO-BFFOTA), hybridized Fuzzy Multi-objective Binary Particle Swarm Optimization and Threshold Accepting based association rule miner (FMO-BPSOTA) have been proposed. These three algorithms have been tested on various datasets such as book, food, bank, grocery, click stream and bakery datasets along with three fuzzy aggregate operators. From these experiments, we can conclude that Fuzzy-And outperforms all the other operators.

INTRODUCTION

Association rule mining, an important task in data mining, is used to find the frequent patterns, potential associations, correlations, of objects among huge data. The rules help find unknown relationships that can be the basis of decision making. It is extensively applied in many diverse areas: market basket analysis, graph mining, sequential pattern mining, medical diagnosis, times series pattern mining, text mining, fraud detection in web, churn modeling in credit card business, click stream mining, discovering protein sequences. The ideas of association rules was pioneered by Agrawal et al. (1993). Here, an association

DOI: 10.4018/978-1-5225-0997-4.ch007

rule is an inference as follows: A→B, where A and B represent antecedent and consequent parts which in turn are item sets (I) or a group of features/items. An item set is a subset of all possible items $\{i_1, i_2,..., i_m\}$ in a transactional database They proposed a priori algorithm (Agrawal & Srikant, 1994) for mining association rules from large customer datasets. The algorithm can be divided into two stages: Frequent item set generation stage and Rule generation stage. In the first stage, frequent item-sets are generated level by level from 1 item-sets by using a measure called support count (SUP). In the second stage, various rules are generated using another user-defined parameter called confidence.

The algorithm suffers from a drawback of repeated database scan during every item-set generation. This drawback has been overcome by FP-Growth algorithm, proposed by Han et al. (2000), which is built on the lines of divide and conquer strategy and requires just two database scans. During its first database scan, it generates F-List i.e., it computes a list of frequent items sorted by frequency in descending order. The second database scan is utilized for transformation F-List into a hierarchical structure called FP-tree. Later on, the algorithm begins to extract rules from FP-tree for each item whose support is larger than the minimum support by recursively building its conditional FP-tree.

Other algorithms include the AIS (Agrawal et al, 1993), the set-oriented algorithm (Houtsma & Swami, 1995), DHP algorithm (Part et al, 1995), the DIC (Sergey et al, 1997), the sampling algorithm (Toivonen, 1997), the cluster-based association rule (CBAR) approach (Tsay et al, 2005) which concentrate on improving the efficiency and accuracy. The well-known drawbacks of most of these algorithms include the need to specify minimum support and confidence, finding the right combination of user-defined parameters. The latter is mostly based on either domain knowledge or trial & error approach. Therefore, here we propose the use of MO-BPSO, MO-BFFO-TA and MO-BPSO-TA developed by Pradeep and Ravi (2014) for association rule mining considering several objectives of interest.

The remaining chapter is organized as follows. In section2, we present a review of literature related to association rule mining and the utilization of Metaheuristics to solve rule mining problems. In Section 3, we introduce different optimization algorithms used in this research followed by Section 4 where we briefly present various steps of our devised strategies. In section 5, we present results and discussion related to various datasets followed by Conclusions Section.

BACKGROUND

Evolutionary computational techniques have been widely applied for mining association rule. Saggar et al. (2004) suggested a technique wherein Genetic Algorithm was used to optimize the association rules obtained by the a priori algorithm. Waiswa et al (2008)proposed Pareto-based GA-centered multi objective evolutionary algorithm mining technique to extract the association rules considering several measures like J-measure, interestingness and comprehensibility as changed objectives of association mining problems. Ghosh et al. (2004) proposed multi-objective association rule mining framework using pareto based GA by considering comprehensibility, interestingness, predictive accuracy and found it suitable for large databases. Anandhavalli et al. (2009) used GA centered approach for mining association rules containing negative attributes with more than one attribute in consequent part. Gupta (2012) used weighted particle swarm optimization (WPSO) for association mining to determine appropriate threshold values for minimum confidence and minimum support. Asadi et al (2012)adopted particle swarm optimization (PSO) to find the minimum confidence and minimum support threshold values for the a priori algorithm. Minaei-Bidgoli et al (2013) proposes multi-objective GA for mining numerical association rules considering