Hybrid Intelligent Method for Association Rules Mining Using Multiple Strategies

Y. Djenouri, University of Algiers, Algiers, Algeria
H. Drias, University of Algiers, Algiers, Algeria
Z. Habbas, University of Lorraine, Metz, France

ABSTRACT

Association rules mining has attracted a lot of attention in the data mining community. It aims to extract the interesting rules from any given transactional database. This paper deals with association rules mining algorithms for very large databases and especially for those existing on the web. The numerous polynomial exact algorithms already proposed in literature processed the data sets of a medium-size in an efficient way. However, they are not capable of handling the huge amount of data in the web context where the response time must be very short. Moreover, the bio-inspired methods have proved to be paramount for the association rules mining problem. In this work, a new association rules mining algorithm based on an improved version of Bees Swarm Optimization and Tabu Search algorithms is proposed. BSO is chosen for its remarkable diversification process while tabu search for its efficient intensification strategy. To make the idea simpler, BSO will browse the search space in such a way to cover most of its regions and the local exploration of each bee is computed by tabu search. Also, the neighborhood search and three strategies for calculating search area are developed. The suggested strategies called (modulo, next, syntactic) are implemented and demonstrated using various data sets of different sizes. Experimental results reveal that the authors’ approach in terms of the fitness criterion and the CPU time improves the ones that already exist.

Keywords: Association Rules Mining, Bees Swarm Optimization, Large Scale Data, Optimization Method, Tabu Search

INTRODUCTION

Association Rules Mining (ARM) is one of the most important and well studied techniques of Data Mining tasks. It aims at extracting frequent patterns, associations or causal structures among sets of items from a given transactional database. Formally, the association rule problem is as follows: let T be a set of transactions \( \{t_1, t_2, ..., t_n\} \) representing a transactional database, and I be a set of m different items or attributes \( \{i_1, i_2, ..., i_m\} \), an association rule is an implication of the

DOI: 10.4018/ijamc.2014010103
form \( X \rightarrow Y \) where \( X \subseteq I \), \( Y \subseteq I \), and \( X \cap Y = \emptyset \). The itemset \( X \) is called antecedent while the itemset \( Y \) is called consequent and the rule means \( X \) implies \( Y \).

Association rule mining is concerning by discovering a set of rules covering a large percentage of data and tends to produce an important number of rules. However, since the databases are increasingly large, the user no longer looks for all the rules but only a subset of useful rules.

Two basic parameters are commonly used for measuring usefulness of association rules, namely the support of a rule and a confidence of a rule.

The support of an itemset \( I' \subseteq I \) is the number of transactions containing \( I' \). The support of a rule \( X \rightarrow Y \) is the support of \( X \cup Y \) and the confidence of a rule is the ration between \( \text{support}(X \cup Y) \) and \( \text{support}(X) \). Confidence is a measure of strength of the association rules. An association rule \( X \rightarrow Y \) with a confidence of 80% means that 80% of the transactions that contain \( X \) also contain \( Y \) together. So, association rules mining consists in extracting from a given database, all interesting rules, that is rules with support \( \geq \text{MinSup} \) and confidence \( \geq \text{MinConf} \). Romero, C., and Zafra, A., (2012) where MinSup and MinConf are two thresholds predefined by users. The main goal is to do so efficiently.

BSO for Bees Swarm Optimization and TS for Tabu Search have been widely used for solving complex problems. BSO was successfully applied to Web Information Retrieval Agrawal, and Ramakrishan, S. (2004).

The determination of search areas for exploration by the bees and defined in BSO results in a good diversification search. Moreover, TS was adapted to manage any neighborhood exploration search. Motivated by the diversification aspect of BSO and the intensification strategy of TS, we propose a new algorithm called HBSO-TS (HBSO-TS for Hybrid Bees Swarm Optimization and Tabu Search Algorithm) for association rule mining. Furthermore three strategies of search area are developed with three called respectively (Modulo, Next, Syntactic).

The rest of this paper is organized as follows. The next section relates about association rules mining algorithms. In section 4, we introduce the Bees Swarm Optimization meta heuristic and in section 5, we present the HBSO-TS algorithm with three different strategies of search area. Section 6 summarizes our experimental results by illustrating the performance of the proposal algorithm and compared it to the state of the art of ARM methods using different data sets (small, average, large) with respect to solution quality and execution time. Section 7 concludes this paper by some remarks and some perspectives for a future work.

**RELATED WORKS**

Many algorithms for generating association rules have been proposed in literature. Some well known exacts algorithms are AIS Agrawal and Imielinski (1993), Apriori, Agrawal, R. and Ramakrishan (2004), Eclat Zaki (2000) and FP-Growth Han, J., Pei, J.(2004). AIS is very space consuming and requires too many passes over the whole database. Apriori is the best known algorithm for association rules mining. It is based on breadth first search strategy to count the supports of itemsets and uses a candidate generation function to exploit the downward closure property of support. FP-growth uses a FP-tree structure to compress the database and a divide-and-conquer approach, to decompose the mining tasks and the database as well. In Zaki, M. J. (1999) the authors present an interesting survey about different exact and polynomial algorithms. However, because of the fast web development and growth of databases, they have become very quickly inefficient. Indeed, even if these polynomial algorithms can still calculate the association rule in a very short
Nature-Inspired Metaheuristics for Automatic Multilevel Image Thresholding
[www.igi-global.com/article/nature-inspired-metaheuristics-for-automatic-multilevel-image-thresholding/119648?camid=4v1a](www.igi-global.com/article/nature-inspired-metaheuristics-for-automatic-multilevel-image-thresholding/119648?camid=4v1a)

An Improved Ant-IS Algorithm for Intrusion Detection
[www.igi-global.com/article/an-improved-ant-is-algorithm-for-intrusion-detection/113829?camid=4v1a](www.igi-global.com/article/an-improved-ant-is-algorithm-for-intrusion-detection/113829?camid=4v1a)