Information Sciences 182 (2012) 3–14

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A multi-level ant-colony mining algorithm for membership functions

Tzung-Pei Hong [a,b,*], Ya-Fang Tung [c], Shyue-Liang Wang [d], Yu-Lung Wu [c], Min-Thai Wu [b]

[a] Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan
[b] Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung 804, Taiwan
[c] Institute of Information Management, I-Shou University, Kaohsiung 840, Taiwan
[d] Department of Information Management, National University of Kaohsiung, Kaohsiung 811, Taiwan

ARTICLE INFO

Article history:
Available online 11 January 2011

Keywords:
Data mining
Ant system
Ant colony system
Fuzzy set
Membership function
Multi-stage graph

ABSTRACT

Fuzzy data mining is used to extract fuzzy knowledge from linguistic or quantitative data. It is an extension of traditional data mining and the derived knowledge is relatively meaningful to human beings. In the past, we proposed a mining algorithm to find suitable membership functions for fuzzy association rules based on ant colony systems. In that approach, precision was limited by the use of binary bits to encode the membership functions. This paper elaborates on the original approach to increase the accuracy of results by adding multi-level processing. A multi-level ant colony framework is thus designed and an algorithm based on the structure is proposed to achieve the purpose. The proposed approach first transforms the fuzzy mining problem into a multi-stage graph, with each route representing a possible set of membership functions. The new approach then extends the previous one, using multi-level processing to solve the problem in which the maximum quantities of item values in the transactions may be large. The membership functions derived in a given level will be refined in the subsequent level. The final membership functions in the last level are then outputted to the rule-mining phase to find fuzzy association rules. Experiments are also performed to show the performance of the proposed approach. The experimental results show that the proposed multi-level ant colony systems mining approach can obtain improved results.

1. Introduction

Knowledge Discovery and Data Mining (KDD) refers to the application of nontrivial procedures for identifying effective, coherent, potentially useful, and previously unknown patterns in large databases [11]. The fast growth of KDD has spurred the development of many related techniques and applications based on different approaches including classification rules, clustering, association rules, and others. Since 1993, the practice of inducing association rules from transaction data has been commonly used in KDD [1]. An association rule is an expression $X \rightarrow Y$, where X is a set of items and Y is usually a single item. In the set of transactions, if all the items in X exist in a transaction, then there is a high probability of Y also being in the transaction. For example, assuming that transactions including the purchase of bread are usually accompanied by the purchase of milk, then the association rule "bread → milk" will be induced.

Most previous studies focused on binary valued transaction data. Transaction data in real-world applications, however, usually consist of quantitative values. Hong et al. [13] thus proposed a mining approach that integrated fuzzy-set concepts with the *apriori* mining algorithm to find interesting fuzzy itemsets and association rules in quantitative transaction data. In

* Corresponding author at: Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan.
E-mail addresses: tphong@nuk.edu.tw, tphong@ksmail.seed.net.tw (T.-P. Hong), m9522030@stmail.isu.edu.tw (Y.-F. Tung), slwang@nuk.edu.tw (S.-L. Wang), wuyulung@isu.edu.tw (Y.-L. Wu), d953040015@student.nsysu.edu.tw (M.-T. Wu).

0020-0255/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2010.12.019

that approach, each item has its own set of membership functions which need to be defined in advance, prompting the development of approaches for automatically finding appropriate membership functions in fuzzy data mining [12,14,18].

The ant colony system (ACS) algorithm is an Ant Colony Optimization (ACO) technique which has recently shown promising for finding nearly optimal problem solutions [5,9]. In the past, we proposed a mining algorithm to find suitable membership functions for fuzzy association rules based on ACS [14]. In that approach, precision was limited by the use of binary bits to encode the membership functions. A large maximum quantity of an item value would result in the code growing too long. Thus, the single-level ACS could not easily obtain a good solution. This paper extends our previous approach through the addition of multi-level processing to increase accuracy. In particular, we propose a multi-layered ant colony algorithm to address large maximum quantities of an item value in the transactions. In the multi-level ACS, the maximum coding length for a feasible solution is set and the approach derives a rough set of membership functions at the first level. The set of membership functions is then refined at the subsequent levels by reducing the search space, resulting in an improved set of membership functions. Thus, different levels have different search ranges. At each level, the same approach is executed on the solution derived from the previous level but with a smaller search space for adjusting membership functions. Regardless of the size of the maximum item value, the coding length can thus be kept fixed without reducing the quality of the derived membership functions. Experimental results show that the proposed approach can generate better membership functions than the previous one, though with greater computational time due to the multi-level processing.

The remaining parts of this paper are organized as follows. Some related work is briefly reviewed in Section 2. A multi-level ACS-based mining framework is presented in Section 3. The proposed multiple-level ACS-mining approach based on the framework is described in detail in Section 4. An example to illustrate the proposed approach is given in Section 5. Experimental results are presented in Section 6, and conclusions and suggestions for future work are given in Section 7.

2. Related works

Fuzzy set theory has been increasingly used in intelligent systems due to its simplicity and similarity to human reasoning [17]. The role of fuzzy sets in data mining may help to transform quantitative values into linguistic terms, thus generating linguistic or fuzzy knowledge. Fuzzy set theory has been applied in many fields such as manufacturing, engineering, medical diagnostics, economics, and others. It can also be easily combined with other techniques [3]. Examples include a fuzzy rule-based back-propagation method for training binary multilayer perceptrons proposed by Delgado et al. [8] and a genetic algorithm to find useful fuzzy concepts for pattern classification proposed by Hu [15].

As for fuzzy data mining, Hong et al. integrated the fuzzy-set concepts and the *apriori* mining algorithm to obtain fuzzy association rules [13]. Defining an appropriate set of membership functions is important in that the set may have a critical impact on the final results of fuzzy data mining. Thus, some GA-based fuzzy data-mining methods for extracting both association rules and membership functions from quantitative transactions have been proposed [12].

Swarm intelligence studies the collective behavior of unsophisticated agents that interact locally through their environment [2], and is inspired by the behavior of social insects and animals, such as ants, fish and birds. The ant system was first introduced by Colorni et al. in [5,9]. Ants are capable of cooperating to solve complex problems such as searching for foods and carrying food. Without using vision, they can find the shortest path between their nests and a food source. As they search, they deposit pheromones on their paths for their companions to find. When another ant crosses these paths, it will select the path with high pheromone density, which thus determines the next node on an ant's route. Once all ants have ended their tours, the amount of pheromone on the tours will have been modified. Ant algorithms have been designed to simulate this type of behavior for solving optimization problems.

The ant colony system (ACS) algorithm [10] is based on the ant system [5,9] and, in this paper, is used to extract membership functions. ACS has been successfully used to obtain nearly optimal solutions for difficult NP-hard problems, such as the Traveling Salesman Problem (TSP), Job Schedule Problem (JSP), Vehicle Routing Problem (VRP) [6,10,23] and other applications [4,22]. The ACS uses three rules: the state transition rule (pseudo random proportional rule), pheromone global updating rule and local updating rule. The ant system is normally adapted according to the different characteristics of the problems to be solved, resulting in many variants of the ant system algorithm, such as the Ant Colony System, the Ant System Elitist, the Rank-Based Version of the Ant System and the Max–Min Ant System. These algorithms differ in several respects, including how they update the pheromone, the state transition, the heuristic functions used, etc. Stützle et al. organized these variants into an integrated system named the Ant Colony Optimization (ACO) [21], and these variants may thus be thought of as the application tools generated by the Ant Colony Optimization. The Ant System, the Ant Colony System and the Max–Min Ant System differ in some aspects of state transaction, local update and global update methods (see Table 1). For example, the ant system and the Max–Min Ant System use the same methods for state transition and local update, but use different methods for global update.

ACO techniques have also shown promise for discovering useful and interesting information from databases, such as classification rules [16,19]. However, research on data mining based on the ant colony system is still rare. Parpinelli et al. proposed an ACS-based approach to find rules from medical data [20]. Cordon and Herrera also proposed using the ACS approach to mine classification rules for fuzzy control systems [7]. Very little research, however, has investigated the mining of association rules.