Expert Systems with Applications 36 (2009) 11844–11852

Contents lists available at ScienceDirect

Expert Systems with Applications

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

An ACS-based framework for fuzzy data mining ☆

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

[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

Keywords:
Ant colony system
Data mining
Fuzzy set
Membership function
Association rule

ABSTRACT

Data mining is often used to find out interesting and meaningful patterns from huge databases. It may generate different kinds of knowledge such as classification rules, clusters, association rules, and among others. A lot of researches have been proposed about data mining and most of them focused on mining from binary-valued data. Fuzzy data mining was thus proposed to discover fuzzy knowledge from linguistic or quantitative data. Recently, ant colony systems (ACS) have been successfully applied to optimization problems. However, few works have been done on applying ACS to fuzzy data mining. This thesis thus attempts to propose an ACS-based framework for fuzzy data mining. In the framework, the membership functions are first encoded into binary-bits and then fed into the ACS to search for the optimal set of membership functions. The problem is then transformed into a multi-stage graph, with each route representing a possible set of membership functions. When the termination condition is reached, the best membership function set (with the highest fitness value) can then be used to mine fuzzy association rules from a database. At last, experiments are made to make a comparison with other approaches and show the performance of the proposed framework.

1. Introduction

Data mining is most commonly used in attempts to induce association rules from transaction data. An association rule is an expression $X \rightarrow Y$, where X is a set of items and Y is a single item (Agrawal & Srikant, 1994). It means in the set of transactions, if all the items in X exist in a transaction, then Y is also in the transaction with a high probability. For example, assume whenever customers in a supermarket buy bread and butter, they will also buy milk. From the transactions kept in the supermarkets, an association rule such as "Bread and Butter → Milk" will be mined out. Most previous studies focused on binary-valued transaction data. Transaction data in real-world applications, however, usually consist of quantitative values. Designing a sophisticated data mining algorithm able to deal with various types of data presents a challenge to workers in this research field.

Recently, the fuzzy set theory has been used more and more frequently in intelligent systems because of its simplicity and similarity to human reasoning (Kandel, 1992). The theory has been applied in fields such as manufacturing, engineering, diagnosis, and economics. Several fuzzy learning algorithms for inducing rules from given sets of data have been designed and used to good effect with specific domains. As to fuzzy data mining, Hong, Kuo, and Wang (2004) proposed a mining approach that integrated fuzzy set concepts with the apriori mining algorithm to find fuzzy interesting itemsets and association rules in quantitative transaction data. In that approach, the memberships functions used for fuzzy data mining have to be defined in advance. In (Hong, Chen, Wu, & Lee, 2006), a GA-based fuzzy data mining method for extracting both association rules and membership functions from quantitative transactions was thus proposed. The proposed GA-based method was divided into two phases: mining membership functions and mining fuzzy association rules. In the phase of mining membership functions, GA is used to derive the membership functions suitable for mining problems. In the phase of mining fuzzy association rules, the best membership functions derived by genetic algorithms are used to fuzzify the quantitative transactions. Then a fuzzy mining approach proposed by Hong, Kuo, and Chi (1999) can be used to find fuzzy association rules.

Recently, ant colony systems (ACS) have been successfully applied to optimization problems. They are inspired from the behavior of social insects and are s heuristic approach. Ants deposit their chemical trails called "pheromone" on the ground for communicating with others. According to the pheromone, ants can find the

☆ This is an extended version of the paper "Extracting membership functions in fuzzy data mining by ant colony systems", presented in the 2008 International Conference on Machine Learning and Cybernetics.

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

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.04.016

shortest path between the source and the destination. The characteristics of an ant colony include positive feedback and distributed computation. It also uses a constructive greedy heuristic (Kuo, Chiu, & Lin, 2004) to search for solutions.

The research about data mining based on the ant colony system is still rare. Previous works on ACS-based rule discovery were proposed by Parpinelli, Lopes, and Freitas (2001) and Cordon and Herrera (2002), in which they proposed the mining of classification rules for fuzzy control systems. Very few other researches explore the association rules. Therefore, in this work, we propose an ACS-based framework to extract membership functions from quantitative data for fuzzy data mining. Numerical experiments on the proposed algorithm are also performed to show its effectiveness.

The remaining parts of the paper are organized as follows. Section 2 reviews ACS and fuzzy data mining. An ACS-based mining framework is then presented in Section 3. The details about how to use ACS on fuzzy data mining are explained in Section 4. The proposed algorithm based on the above framework is described in Section 5. An example demonstrating the proposed algorithm is given in Section 6. Numerical simulations are shown in Section 7. Conclusion and future work are given in Section 8.

2. Background

This section reviews some basic concepts related to this paper. They are fuzzy data mining and ant colony systems.

2.1. Fuzzy data mining

The fuzzy set theory was proposed by Zadeh (1965). His primary idea is to use natural languages to represent the concepts, in which words may have ambiguous meanings. It may be especially useful for quantifying and reasoning. Fuzzy sets can be thought as one of the extensions of traditional crisp sets, in which each element must either be in or not in a set. The role of fuzzy sets in data mining helps transform quantitative values into linguistic terms, thus reducing possible itemsets in the mining process. Hong et al. then proposed a fuzzy mining algorithm to mine linguistic association rules (Hong et al., 1999; Hong, Kuo, & Chi, 2001). They first transformed each quantitative value into several fuzzy sets labeled with linguistic terms by membership functions. The algorithm then calculated the scalar cardinality of each linguistic term on all the transaction data. The mining process based on fuzzy counts is then performed to find fuzzy association rules. Hong et al. then modified the previous algorithm and proposed a new fuzzy data mining algorithm for extracting both association rules and membership functions (Hong et al., 2006).

Besides, Kaya et al. proposed a GA-based clustering method to derive a predefined number of membership functions for getting a maximum profit within an interval of user-specified minimum-supported values (Kaya & Alhajj, 2003). Later, the idea of a multi-objective GA to find a number of Pareto-optimal rule sets and an automatic method for mining association rules came into existence (Kaya et al., 2004; Kaya & Alhajj, 2005). There is still a fairly large body of literature about fuzzy data mining (Cordón, Herrera, & Villar, 2001; Lee, Hong, & Lin, 2004; Roubos & Setnes, 2001; Setnes & Roubos, 2000; Subramanyam & Goswami, 2005; Wang, Hong, & Tseng, 1998; Wang, Hong, & Tseng, 2000; Yue, Tsang, Yeung, & Shi, 2000). In this paper, we attempt to propose a new model based on the ant colony system for fuzzy data mining.

2.2. The ant colony system

The ant system was first introduced in 1991 (Colorni, Dorigo, & Maniezzo, 1991; Dorigo, Maniezzo, & Colorni, 1996), and then extended to the ant colony system (Dorigo & Gambardella, 1997). The idea of the ant system is from the observation on the real colonies of ants searching for food. Ants are capable of cooperating to solve complex problems such as searching for foods, carrying food and so on. They can find the shortest path between their nests and food without using vision. They will deposit pheromone on the paths for their companions. When the next ants go through the paths, they select the path with high density of pheromone. Ants thus determine the next node on the route according to the pheromone density. Once all ants have terminated their tours, the amount of pheromone on the tours will have been modified.

The ant algorithms have thus been designed to simulate the above ant behavior for solving optimization problems. Especially, the process of modifying the amounts of pheromone on the tours is called the updating rule, which is designed to give more pheromone to the best path. Currently, the ant algorithms have been applied to solve several difficult NP-hard problems, such as Traveling Salesman Problem (TSP) (Dorigo & Gambardella, 1997), Quadratic Assignment Problem (QAP) (Maniezzo, Colorni, & Dorigo, 1994; Stutzle & Dorigo, 1999), Vehicle Routing Problems (VRP) (Kuo et al., 2004; Montemanni, Gambardella, Rizzoli, & Donati, 2005; Wade & Salhi, 2004), Job Schedule Problem (JSP) (Colorni, Dorigo, Maniezzo, & Trubian, 1994) etc. (Maniezzo & Arbonaro, 2000; Merkle, Middendorf, & Schmeck, 2002).

The ant colony system (ACS) (Dorigo & Gambardella, 1997) proposed by Dorigo et al. is based on the ant system (Dorigo et al., 1996) and is applied to extract membership functions in this paper. The algorithm for finding solutions to an optimization problem is shown in Fig. 1 (Dorigo & Gambardella, 1997).

The three rules used in the ACS algorithm are described below.

1. State transition rule: it defines how an ant probabilistically changes its current state to a next state (node) to form a solution.
2. Global updating rule: it defines how the pheromone of the best tour passed by the ants will be updated after all ants have completed their tours.
3. Local updating rule: it defines how the pheromone of a path is updated when an ant constructs the path.

3. The ACS-based fuzzy mining framework

In this section, a fuzzy mining framework based on the ACS algorithm is proposed to discover both useful association rules and suitable membership functions from quantitative values. The proposed framework is shown in Fig. 2 where each item has its own membership function set. These membership function sets are then fed into the ant colony system to search for the final appropriate sets. When the termination condition is reached, the

```
Initialize
  Loop
  /* at this level each loop is called an iteration */
    Each ant is positioned on a starting node
    Loop
     /* at this level each loop is called a step */
    Each ant applies a state transition rule to
    incrementally build a solution.
    A local pheromone updating rule is applied
    Until
   All ants have built a complete solution
   A global pheromone updating rule is applied
  Until
End_condition
```

Fig. 1. The ACS algorithm.