Ant Colony Algorithms for Data Learning

Mohamed Hamlich, Computer Science Lab, UH2, FSTM, Mohammadia, Morocco
Mohammed Ramdani, Computer Science Lab, UH2, FSTM, Mohammadia, Morocco

ABSTRACT

Fuzzy Ant-Miner algorithm processes data with nominal class and has the disadvantage of not treating the data with continuous class. In this paper, after presenting the Fuzzy Ant Miner algorithm, the authors propose a new learning method to partition heterogeneous data with continuous class. This method in a first step finds the optimal path between the data using algorithms of ants. Distance adopted in their optimization method takes into account all types of data. The second step vise to divide the data into homogeneous groups by browsing the optimal path found. A new test probability is estimated based on the distance and the amount of pheromone deposited by ants in the transitions between the data. A third step is to find the prototype of each cluster to identify the cluster membership of any new data injected.

Keywords: Ant Colony Algorithms, Classification, Continuous Class, Data Clustering, Fuzzy Ant-Miner

1. INTRODUCTION

Ant miner Parpinelli, Lopes, & Freitas, 2002 is a classification algorithm that learns rules from training sets by using the simulation of real ants. Each artificial ant is a mere agent in the system and finds a solution to the problem and communicates with other ants indirectly through the amount of pheromone deposited.

Ant miner is an efficient algorithm with discrete attributes. But with the continuous attributes, the ant colony algorithm uses intervals with strict boundaries. This can lead to false predictions of the target attribute, especially if the attribute’s value is close to the borders of discretization.

To address this problem, Fuzzy Ant-Miner (Hamlich & Ramdani, 2012) treats continuous attributes using the concepts of fuzzy logic (Zadeh, 1965): Continuous attributes are discretized into fuzzy partitions that will be used in a Fuzzy Ant-Miner algorithm. The originality of this approach is that it generates fuzzy rules by using the concept of fuzzy entropy and fuzzy fitness of a rule (Hamlich & Ramdani, 2011). The basic idea of this algorithm is that the probability with which ant chooses a value of a fuzzy continuous attribute among all those possible depends on two things: the amount of pheromone deposited by the ant, and the fuzzy entropy of the value. After obtaining the complete rule, the method proceeds to pruning based on the quality of a fuzzy rule. This provides simplified and easy to interpret rules.

Fuzzy Ant-Miner algorithm (Hamlich & Ramdani, 2012) processes data with nominal

DOI: 10.4018/jaec.2013070101
class and has the disadvantage of not treating the data with continuous class. To process the data with continuous class, we propose a clustering method based on ant colony algorithm.

In this paper we first describe the base Fuzzy Ant-Miner algorithm in section 2. In the third section, we explain the improvements made by our clustering method. In the section 4, we present an illustration of the method.

2. FUZZY ANT-MINER ALGORITHM

2.1. Fuzzy Discretization

In order to avoid the crisp discretization used by cAnt-Miner (Otero, Freitas, & Johnson, 2008), our method extends the crisp partition by discretizing the values of an attribute into fuzzy partitions. First a threshold $x$ is determined in the same way as cAnt-Miner (the one that leads to the minimum entropy (Quinlan, 1993)). We assume that the threshold is a fuzzy number around $x$.

The membership functions of the continuous attribute $A_i$ to a fuzzy value $A_i < x$ is calculated by:

$$
\mu_{A_i < x}(v) = \begin{cases} 
1 & \text{if } b \leq v \leq c \\
\frac{v - b}{b - a} & \text{if } a < v < b \\
0 & \text{if } v \geq b
\end{cases}
$$

(1)

The membership functions of the continuous attribute $A_i$ to a fuzzy value $A_i < x$ is calculated by:

$$
\mu_{A_i < x}(v) = \begin{cases} 
1 & \text{if } v \geq d \\
\frac{v - d}{d - c} & \text{if } c \leq v \leq d \\
0 & \text{if } v \leq c
\end{cases}
$$

(2)

where parameters $(a, b, c, d)$ determine the boundaries of the fuzzy area and so the degree of fuzziness. The user sets these parameters for the best accuracy. Note that other ways of generating the fuzzy partitions Liu, Hussain, Tan, & Dash, 2002, Marsala, 1998 can be used in our method.

2.2. Fuzzy Ant-Miner Algorithm

In order to deal with fuzzy concepts we have extended the algorithm Ant-miner to treat these concepts as presented in Algorithm 1:

Figure 1. Fuzzy discretization of a continuous attribute
Supporting Business Cases for PHM: Return on Investment and Availability Impacts
www.igi-global.com/chapter/supporting-business-cases-phm/69683?camid=4v1a