Expert Systems with Applications 41 (2014) 7524–7535

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Handling numeric attributes with ant colony based classifier for medical decision making

Matej Pičulin *, Marko Robnik-Šikonja

Faculty of Computer and Information Science, Tržaška 25, 1000 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Available online 17 June 2014

Keywords:
Ant Colony Optimization
Ant-Miner
Numeric attributes
Rule learning
Classification
Medical data mining

ABSTRACT

In data mining many datasets are described with both discrete and numeric attributes. Most Ant Colony Optimization based classifiers can only deal with discrete attributes and need a pre-processing discretization step in case of numeric attributes. We propose an adaptation of AntMiner+ for rule mining which intrinsically handles numeric attributes. We describe the new approach and compare it to the existing algorithms. The proposed method achieves comparable results with existing methods on UCI datasets, but has advantages on datasets with strong interactions between numeric attributes. We analyse the effect of parameters on the classification accuracy and propose sensible defaults. We describe application of the new method on a real world medical domain which achieves comparable results with the existing method.

1. Introduction

The task of classification rule mining is to describe data by IF-THEN rules. The advantage of such rules is that they are comprehensible even to a non expert. This makes classification rule mining useful in cases where we want to explain the reason for each decision. The downside of classification rules is that they usually achieve lower classification accuracy compared to other black-box approaches.

ACO (Ant Colony Optimization) was introduced by Dorigo, Maniezzo, and Colorni (1996) is a general and successful optimization approach. ACO is an swarm intelligence approach that mimics the behavior of foraging ants which leave pheromone trails on the ground to guide other ants to the food source.

In this way ants in nature indirectly communicate. ACO algorithms simulate pheromone trails to guide the search through the graph representation of the problem and use other heuristics as a background knowledge to guide the search at the beginning, where there are no pheromones.

ACO based algorithms can be applied and adapted to many different problems. A recent overview of the use of ACO in engineering was made by Chandra Mohan and Baskaran (2012). In this article we will focus on applying it on syncope dataset for medical decision making. In this kind of problems both high classification accuracy and comprehensibility of the constructed model are important. We chose ACO based approach since it achieves both good accuracy and good comprehensibility. We show that the method can detect some dependencies among attributes and can therefore detect some interesting rules.

Several rule mining algorithms based on ACO exist for mining classification rules. You can see some of the work by Parpinelli, Lopes, and Freitas (2002), Martens et al. (2007) and Otero, Freitas, and Johnson (2008). ACO based rule mining algorithms build a discrete search space, represented by a graph, in which ants try to find the best rule set by discrete optimization. Since ACO is used for discrete optimization this can be problematic when datasets are described with numeric or mixed attribute types. The most frequent solution is to use discretization as a pre-processing step, which can lead to a potential loss of information and inferior models. This is especially pronounced on datasets with interactions between attributes, where the loss of information due to uninformed discretization is prohibitive. The problem with discretization was recognised and addressed in cAnt-Miner by Otero et al. (2008) which uses dynamic discretization. We propose a method that handles numeric attributes without any prior discretization step and thereby keeps all the information. This was also recently done for other swarm intelligence methods, namely particle swarm optimization, approach by Beiranvand, Mobasher-Kashani, and Bakar (2014). To achieve this we use a different graph representation for numeric attributes and simultaneously update multiple paths in the search graph.

The proposed method, called nAnt-Miner, is based on ACO and can deal with nominal, ordinal and numeric attributes without

* Corresponding author. Tel.: +386 1 4768 459.
E-mail address: matej.piculin@fri.uni-lj.si (M. Pičulin).

http://dx.doi.org/10.1016/j.eswa.2014.06.017
0957-4174/

pre-processing or using on-the-fly discretization. For each type of attributes nAnt-Miner uses a different representation in the search graph.

The remainder of the paper is split into four sections. Section 2 briefly describes similar approaches. In Section 3 we describe the new graph presentation for numeric attributes and its parameters. Section 4 presents empirical evaluation and Sections 5 draws conclusions.

2. Related work

This section gives a quick overview of ACO based methods for classification rule mining.

The first implementation of ACO for mining classification rules was introduced by Parpinelli et al. (2002) and named Ant-Miner. The method was later extended by Wang and Feng (2004) and Liu, Abbass, and McKay (2004) mainly by different pheromone updating rules favoring exploration. These methods can only deal with nominal attributes. The main idea of these approaches is to construct a discrete search space from given data. Ants are then allowed to run trough the graph from the start to the end node and the path they make describes a classification rule. The found rules are evaluated and based on their quality, the paths by which they were constructed are strengthened by artificial pheromones. This process is repeated until all or most of the ants converge to a single path and then the corresponding rule is added to the rule set. The examples covered by this rule are then removed from the training data and the process is repeated until no more data remains.

Ant-Miner+ by Martens et al. (2007) can handle ordinal attributes by using different graph representation. cAnt-Miner by Otero et al. (2008) can handle numeric attributes using implicit discretization. HACO (Hybrid ACO) by Xiao and Li (2011) is a newer hybrid method that can also deal with numeric attributes. Seçkiner, Eroğlu, Emrullah, and Dereli (2013) also developed a method addressing the same problem using solution archives. These approaches show that there is a need to handle numeric attributes. These approaches have comparable predictive performance to other rule mining methods like CN2 by Clark and Niblett (1989) and Ripper by Cohen (1995) and have several spawned variants. We give their short description below.

2.1. Ant-Miner

Ant-Miner and its variants use a separate and conquer approach as described by Fürnkranz (1999) which means that the algorithm generates one rule, removes (separates) the covered examples from the dataset and then learns the remaining rules (conquers) from the remaining instances. An alternative is to use Pittsburgh approach described by Otero, Freitas, and Johnson (2012), which tries to find the best set of rules instead of the set of best rules. Pittsburgh approach evaluates the whole set of rules together and not every rule independently.

Ant-Miner by Parpinelli et al. (2002) can only handle discrete attributes and does not differentiate between nominal and ordinal attributes. The search graph is composed of groups of nodes for nominal attributes, with one node for each nominal value as shown in Fig. 1. Each column in the graph represents one attribute. Attribute Sex has values M and F, attribute Age values 5, 10, 15 and 20 and attribute Country values Fra, Ger and Spa. The bolded line shows one possible path an ant can take to construct the rule IF Age = 15 AND Sex = *M* AND Country = *Ger* THEN class, where class is the majority class of the instances the rule covers. The order of groups is unimportant and the Start node is connected to all nodes except the End node. Nodes representing the same attribute are not connected but are connected to all other nodes. There is also a restriction that if one node of a group is selected then links to this group can not be selected again by an ant preventing construction of conflicting rule terms. The graph complexity and consequentially search space is larger than in our approach.

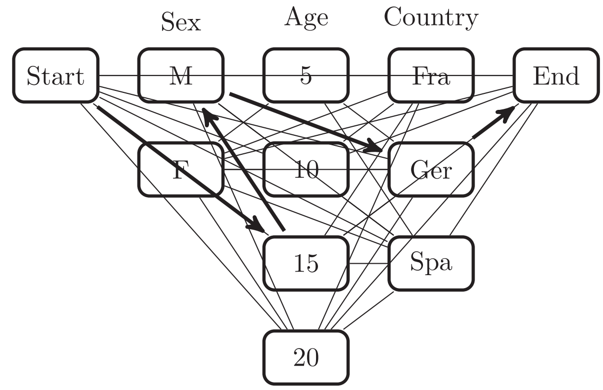

Fig. 1. An example of graph representation in Ant-Miner.

The quality of the rule is calculated as sensitivity times specificity of the rule. The pheromone values are saved in nodes instead of edges, as in later approaches. The pheromone update rules for evaporation are the same as described in Section 3.3, but pheromones values are not restricted.

2.2. cAnt-Miner

cAnt-Miner by Otero et al. (2008) is the first ACO based data mining algorithm that handles numeric attributes. In cAnt-Miner discretization is done on-the-fly when an ant chooses a numeric attribute. The graph representation is similar to the original Ant-Miner with a single node for each numeric attribute.

The method calculates a heuristics for each node to guide the search at the beginning when there are no pheromones. For discrete attributes this is similar to approach in Section 3.4.

cAnt-Miner uses a single node a_i to represent a numeric attribute. When this node is visited, the ant has to select a cut point v such that it will split values of a_i into two parts $a_i < v$ and $a_i \geqslant v$. The merit of a split value v is determined by:

$$\mathrm{split}_v(a_i) \equiv \frac{|S_{a_i<v}|}{|S|}\,\mathrm{entropy}\,(a_i < v) + \frac{|S_{a_i\geqslant v}|}{|S|}\,\mathrm{entropy}\,(a_i \geqslant v) \qquad (1)$$

where $|S|$ is the number of training instances, $|S_{a_i<v}|$ and $|S_{a_i\geqslant v}|$ are the number of instances in partition $a_i < v$ and $a_i \geqslant v$, respectively. The approach tries to find v minimizing $\mathrm{split}_v(a_i)$. The splits depend on the order of attributes selected on the way through the graph as the instances used to find the split v depend on previous choices. An extension of cAnt-Miner solves this problem by using pheromones on edges. A detailed description of these problems is found by Otero, Freitas, and Johnson (2009) in their work.

2.3. Ant-Miner+

Ant-Miner+ by Martens et al. (2007) is an extension of Ant-Miner using $\mathcal{MAX}$-$\mathcal{MIN}$ Ant System algorithm (Stützle & Hoos, 2000) The method can handle ordinal attributes but it cannot handle numeric attributes.

The method introduces two additional groups in the graph representation placed immediately after the Start node. They serve for selection of parameters influencing the relative importance of pheromones and heuristic values. Nominal and ordinal attributes are handled the same as in our nAnt-Miner.