



## A Novel Method for Extracting Classification Rules based on Ant-Miner

Babak Fakhar

Department of Computer Science, Mahshahr branch, Islamic Azad University, Mahshahr, Iran

Fakharrbabak@yahoo.com

### Article history:

Received August 2013

Accepted September 2013

Available online September 2013

### Abstract

In this paper, we propose a new method for data mining based on Ant Colony Optimization (ACO). The ACO is a metaheuristic inspired by the behavior of real ants, where they search for optimal solutions by considering both local heuristic and previous knowledge, observed by pheromone changes. Good results of the ant colony algorithms for solving combinatorial optimization problems suggest an appropriate effectiveness of the approach also in the task of discovering classification rules in data mining. Mining classification rules is an important research area in data mining. Ant-Miner is an Ant Colony Optimization algorithm for classification task. This paper proposes an improved version of Ant-Miner named Ant-Miner4, which is based on Ant-Miner3. By changing the heuristic function used in the Ant-Miner3, and implementing it based on correction function of Laplace, we tried to redesign Ant-Miner to gain rules with high predictive accuracy. We compared Ant-Miner4 with the previous version (Ant-Miner3) using four data sets. The results indicated that the accuracy of the rules discovered by the new version was higher than the ones gained by the previous version.

**Key words:** Ant Colony Optimization Algorithm, Classification Rules, Data Mining, Laplace Correction

## 1. Introduction

Classification is one of the most frequently occurring tasks of human decision making. A classification problem encompasses the assignment of an object to a predefined class according to its characteristics. Many decision problems in a variety of domains, such as engineering, medical sciences, human sciences, and management science can be considered as classification problems. Popular examples are speech recognition, character recognition, medical diagnosis, bankruptcy prediction, and credit scoring. The main objective of Ant-Miner is extracting knowledge from data sets. Data mining is a science based on machine learning, statistic rules and data sets.

It shall be emphasized that in data mining, in contrast to statistical methods, the goal is, only and only, extracting knowledge; a knowledge which not only should be correct, but it shall convey universality for the operator [1, 2]. Data mining process consists of several tasks such as classification, regression, independent modeling, clustering, etc. [1].

The classification is the process of finding a model that describes and distinguishes data classes or concepts, and its goal is to use the model to predict the class of objects whose class label is unknown. In this research, we tried to change the Ant-Miner algorithm to gain rules with high power of forecasting.

The process of finding rules is an important part of data mining. During this process, a symbolic set of rules is presented for explaining each class or family of data [12]. Compared to other methods, this method of presenting data is more comprehensive for an operator [5].

For the first time, Mr. Parepinelli presented a system named Ant-Miner to explore the classification rules of data sets [13]. This system is designed based on the ant colony optimization. In the next version of Ant-Miner (Ant-Miner2), a new method was presented to estimate the heuristic function. In this version, the amount of the heuristic function was dependent on the rate of replication of a term in a class. In Ant-Miner3, a new strategy was presented for updating the pheromone tracks [11]. In this research, we tried to apply Laplace correction function to design the heuristic function of the Ant-Miner. Our new method, without increasing the running time of the algorithm, can explore rules, which have high accuracy in forecasting. We named our new algorithm as Ant-Miner4. To test the efficiency of Ant-Miner4, we compared it with Ant-Miner3 using four data sets. Based on the results, the accuracy of the rules obtained by Ant-Miner4, on average, is higher than the ones obtained by Ant-Miner3.

## 2. Ant Colony System (ACS)

Ant Colony System is a branch of artificial intelligence, which is known as Swarm intelligent. Ants are independent insect that have collaborative activities [6]. Ants can search for the shortest path from

their nest to the food source, without any central controlling, centralized surveillance or imaging information. Ants contact with each other indirectly using a substance named as pheromone, and share the information of the path with each other. Each ant comes across different paths while solving the problem. When an ant passes a path, he increases the pheromone amount on that path based on the quality of that path. Therefore, when an ant confronts with several paths, the probability of selecting the path with a high amount of pheromone increases. The ants always choose the shortest path to the food source. The main point in the behavior of the ants or subjects is collaboration and adaptation [4]. By using this system, we can present a meta heuristic and evolutionary method for solving the problems of NP optimization [9].

The prerequisite of solving a problem based on the Ant Colony System is defining the structure of that problem in a graph format. After presenting this definition, we can use the ACS for finding the shortest path, which is the solution for the problem. The general function of the ants for selecting the nodes in paths depends on a probability function, which is shown in formula (1). This formula is discussed in section 3.

### 3. Studying the operation of Ant-Miner1

Ant-miner is a learning method with supervision for extracting classification rules of a data set. Supervision has three steps: 1. selecting a training set, 2. presenting a method for data mining and analysis (learning) and 3. synthesis and combining forecasting rules. Selection of a training set is a critical step in data mining. The selected training set should be simple as small as possible and shall be general enough [5, 12]. The output of the Ant-Miner is a set of rules shown in figure (1). The general view of the Ant-miner is shown in figure (2).

IF ( term1 AND term2 AND ...) THEN <class>
---

Figure (1): The output of Ant-Miner

Each term is as  $term_{ij} = A_i = V_{ij}$

$A_i$ :  $i$ th variable.

$V_{ij}$ :  $i$ th amount from  $A_i$  domain.

In Ant-Miner, first, an ant starts his move from an empty rule (if then); next, at each step, one term is selected and it is added to the current rule. In other words, an ant adds a new term to the previous step by selecting a path. In this way, ants make the structure of the problem by moving in the graph, and add new terms to the current rule by passing through different nodes. After making a rule, its

forecasting capability is checked; next, it is stored temporarily in a list. Now, it is the next ant's turn. This ant, based on the amount of the pheromone stored in the paths by previous ants, and based on the function (1), starts making a new rule. The operator (Repeat loop) can determine the number of ants at this point. At the end, the best rule (the best quality) is select and it is added to the list of the explored rules. The explored rule is capable of covering the samples of a specific class. In a similar way, the same process is repeated to find a new rule to cover the samples of another class. This loop (While) is repeated until all or most of the available samples in the training set are covered with explored rules. Selection of the path or term for adding into the rule which is under construction by ants is done based on function (1) and also based on the heuristic function, which is dependent on the problem and the amount of the pheromone present in the path.

$$P_{ij} = \frac{\eta_{ij} \cdot \tau_{ij}}{\sum_{i=1}^a \sum_{j=1}^{b_i} (\eta_{ij} \cdot \tau_{ij}(t))} \tag{1}$$

$\eta_{ij}$ : the amount of the heuristic function, which is dependent on the problem for termij

$\tau_{ij}$ : the amount of pheromone at i to j path or termij

```

Training set = all training cases;
WHILE (No. of uncovered cases in the Training set > max_uncovered_cases)
  i=0;
  REPEAT
    i=i+1;
    Anti-incrementally constructs a classification rule;
    Prune the just constructed rule;
    Update the pheromone of the trail followed by Anti;
  UNTIL (i ≥ No_of_Ants) or (Anti constructed the same rule as the
  previous No_Rules_Converg-1 Ants)
  Select the best rule among all constructed rules;
  Remove the cases correctly covered by the selected rule from the training set;
END WHILE
    
```

Figure (2): General view of Ant-Miner [12]

Ant-miner estimates the amount of  $\eta_{ij}$  function for any applied termij in each rule, to help the ants in selecting the terms. This function is designed based on the structure of the problem's theory [3, 11].

In the original version of the Ant-Miner, the value of the heuristic function is obtained using (entropy) formula (2).

$$\eta_{ij} = \frac{\text{Log}(K) - \text{Info}T_{ij}}{\sum_i^a \sum_j^{b_i} \log_2^{(k)} - \text{Info}T_{ij}} \tag{2}$$

$$InfoT_{ij} = -\sum_{w=1}^k \left[ \frac{freqT_{ij}}{|T_{ij}|} \right] \times \log_2 \left[ \frac{freqT_{ij}}{|T_{ij}|} \right] \quad (3)$$

K: number of classes.

$|T_{ij}|$ : total number of samples in  $T_{ij}$  section (each section consists of several samples for which  $A_i=V_{ij}$ ).

$InfoT_{ij}$ : total number of terms' replication in  $T_{ij}$  section of W class.

a: total number of the specific attribute (variables).

bi: total number of the values of a specific domain of the i specific attributes.

In the above formula, by increasing the value of  $InfoT_{ij}$ , the probability of selecting  $Term_{ij}$  by an ant, that is making a rule, decreases. Because, the value of  $InfoT_{ij}$ , varies between zero and  $Log_2(k)$ . If the value of  $V_{ij}$  does not happen for  $A_i$  in the training set, then the value of  $InfoT_{ij}$ , becomes maximum or equals to  $Log_2(k)$ ; this makes the  $term_{ij}$  to have the least power of forecasting. When all samples belong to a same class, then the value of the  $InfoT_{ij}$ , equals zero indicates that the probability of choosing  $term_{ij}$  increases. The heuristic function designed for Ant-Miner1 is similar to the heuristic function of CN2 decision tree [7].

### 3.1. Rule pruning

After the ant completes the construction of its rule, the rule pruning procedure is performed for each ant. Rule pruning is a common process in data mining. The main goal of rule pruning is to improve the rule quality and increasing the simplicity by removing irrelevant terms which might have been added during the rule construction process. [8, 12]. Here, the main idea is starting the operation with a rule, which includes several terms, and next at each replication, one term is omitted from the rule and the quality of the rule under construction is determined by formula (4). This process is continued until the quality of the rule increases by omitting a term. It shall be mentioned that this process may change the forecasting class of a rule.

$$Q = \left( \frac{TruePos}{TruePos + FalseNeg} \right) \times \left( \frac{TrueNeg}{FalsePos + TrueNeg} \right) \quad (4)$$

where TruePos is the number of cases covered by the rule and having the class predicted by the rule, FalsePos is the number of cases covered by the rule and having a class different from the class predicted by the rule, FalseNeg is the number of cases that are not covered by the rule while having the class predicted by the rule, and TrueNeg is the number of cases that are not covered by the rule and having a class different from the class predicted by the rule.

## 5. Evaluating the Ant-Miner3

In Ant-Miner3, there is a new method for updating the pheromone in all paths. In this algorithm, the value of a pheromone increases based on formula (9), considering the terms used in the construction of the rule. In a similar way, the amount of the pheromone is normalized by using the applied terms [12].

$$\tau_{ij}(t) = (1 - p) \cdot \tau_{ij}(t-1) + (1 - \frac{1}{1+Q}) \tau_{ij}(t-1) \cdot \forall i, j \in R \quad (9)$$

$Q$ : determines the quality of the rule, which is constructed.

$P$ : by using this parameter, the evaporation rate of the pheromone can be determined in a path.

In this formula, the value of  $Q$  ranges between 0 and 1. The maximum value of  $Q$  indicates the maximum rate of increase of the pheromone for the applied term; while the minimum value of  $Q$  indicates the decrease of the pheromone amount for the terms, which are not used in the construction of the rule.

The main property of the  $P$  parameter is that we can control the effect of time on the construction of the current rule [12]. The value of this parameter also ranges between 0 and 1. In addition, in Ant-Miner3 there is a new method for choosing the terms in order to be added to the previous rule. In this method, choice of a term is done based on algorithm (1).

Algorithm (1):

If  $q_1 \leq \varphi$

  Loop

    If  $q_2 \leq \sum_{j \in ji} P_{ij}$  then

      choose term<sub>ij</sub>

    end-loop

else

  Choose term<sub>ij</sub> with max  $P_{ij}$

$q_1, q_2$ : random numbers

$\varphi$ : a parameter between 0 and 1

$ji$ : the number of  $i$ th value for  $A_i$

$p_{ij}$ : is the probability, and it is determined by formula (5).

$$\tau(t=0) = \frac{1}{\sum_{i=1}^a b_i} \quad (5)$$

$a$ : the number of the attributes (variables).

$b_i$ : the possible values for  $a$ .

When an ant completes construction of a rule, the amount of pheromones is updated based on formula (6) considering the terms used in the construction of the rule.

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \tau_{ij}(t) \cdot Q \forall i, j \in R \quad (6)$$

To simulate the process of pheromone evaporation, the amount of pheromone associated with each term, which does not occur in the constructed rule must be decreased. For this purpose, the formula (7) is used.

$$\tau_{ij} = \frac{\tau_{ij}}{\sum \tau_{ij}} \forall i, j \quad (7)$$

$\tau_{ij}$ : the amount of pheromone in  $i$  to  $j$  path or for each value of term  $ij$

#### 4- Evaluating the Ant-Miner2

In Ant-Miner2, which is also known as density-based Ant-Miner [11], the heuristic function is estimated by formula (8).

$$\eta_{ij} = \frac{\text{majority\_class}T_{ij}}{|T_{ij}|} \quad (8)$$

$\text{majority\_class}T_{ij}$ : for estimating this value, first, all samples for which  $A_i = V_{ij}$  are separated ( $T_{ij}$ ), next the samples are screened to detect their appropriate class. Estimating the heuristic function based on formula (8) ameliorate the time complexity of the Ant-Miner. However, it does not affect the rules' power of forecasting. Estimated in formula (1). The goal of the random numbers  $q1$  and  $q2$  is making a balance between use of the knowledge extracted by the previous ants and new knowledge extraction. Therefore, if  $q1 \leq \phi$ , it indicates the new knowledge extraction; otherwise, it indicates previous knowledge use. This phenomenon provides a situation to find out the paths, which have not been passed through by previous ants, therefore, if there is new knowledge, it can be found by the following ants.

## 6. Designing a new algorithm (Ant-Miner4) our contribution

To make a new rule, an ant starts the operation with an empty rule. As a rule is made to cover the samples of a specific class, it is important to choose an appropriate  $A_i = V_{ij}$  to cover the samples of that class with enough precision. In other words, the question is for covering the samples, which belong to class C, what term ( $A_i = V_{ij}$ ) has the high informative value? In Ant-Miner3, All terms have the equal value of pheromone. Therefore, at the first step, the only factor affecting the choice of a term is the amount of the heuristic function of that term. Considering that it is more probable that the next ants choose the paths, which have been chosen by previous ants. Therefore, the heuristic function is important in guiding the ants.

We used Laplace correction (formula 10) for designing heuristic function in Ant-Miner4.

$$\eta_{ij} = \frac{|term_{ij}, K| + 1}{|term_{ij}| + No\_of\_classes} \quad (10)$$

$|term_{ij}|$ : indicates the number of training sets that consists of  $term_{ij}$  in the current k class

$No\_of\_classes$ : defines the numbers related to the variable class

Using Laplace correction function, increases the accuracy of the rules, decreases over fitting and helps managing subgroups in the Ant-Miner. Over fitting occurs when a training set is not covered by any rule, or it is covered by more than one rule into different classes. For example, consider a set of data, which has a sample with some absolute values, which belong to a class. There are two classes. In this case, the heuristic function of Ant-Miner3 (formula 8) considers the value of  $1/1=100\%$  as the informative value of the term. The value of 100% destroys the generalization property of the rule made by the current term. The formula (8) does not consider the number of classes of the data sets when it estimates the informative value of a term. While, in the case of Laplace function, by considering the number of classes, the value of  $(1+1)/(1+2)=67\%$  is assigned for the current term. This value is more close to reality and has more appropriate generalization property. Due to appropriate output of Laplace correction function, we redesigned the Ant-Miner using this heuristic function.

## 7. Comparing Ant-Miner3 with Ant-Miner4

For evaluating the output of the Ant-Miner4, four data sets (table 1) were used.

TABLE (1):  
THE DATA SETS USED FOR THE EXPERIMENT



Data set	#cases	#categ. attrib.	#contin. attrib.	#classes
Breast cancer	683	-	9	2
Tic-tac-toe	953	9	-	2
Dermatology	366	33	1	6
Hepatitis	155	13	6	2

These data sets were obtained from the database of UCI [14]. The information in table (1), from left to right, is as follows: data set, number of cases in the data set, number of variables, number of contiguous values, and number of classes.

After running Ant-Miner3 and Ant-Miner4 on the above-mentioned data sets, the outputs of systems in forecasting are presented in table (2). It shall be mentioned that in the reference [13], for proving the efficiency of the Ant-Miner1, it was compared to CN2. Table (2) shows that the Ant-Miner4 had a high power of forecasting in two cases, a similar power of forecasting in one case, and a less power of forecasting in another case.

TABLE (2):  
COMPARING THE EFFICIENCY OF THE NEW VERSION OF THE ANT-MINER WITH THE PREVIOUS VERSION.

Data sets	Forecasting accuracy of Ant-Miner3	Forecasting accuracy of Ant-Miner4
Breast cancer	94.32±0.94	95.52±0.88
Tic-tac-toe	76.58±2.53	78.38±0.52
Dermatology	95.18±0.7	91.45±1.216
Hepatitis	91.1±2.4	91.22±1.98

## 8. Conclusion

The induction decision tree is one of the best methods in finding the rules of classification in data sets. Mr. Parepinelli proved that by using Ant-Miner1, we could obtain simpler rules with higher accuracy and precision compared to CN2 decision tree. In this research, Ant-Miner1, Ant-Miner2 and Ant-Miner3 were explored and the process of obtaining a new version was studied. In the new version, the accuracy of the obtained rules was increase by using the heuristic function of Laplace correction. This property increases the application of the Ant-Miner for extracting knowledge from real world data. The future research can be focused on designing a new version of Ant-Miner, which can find an unsorted list of rules. In addition, in Ant-Miner4, no attempt was made to configure the parameters; therefore, more research can be done on this part.

## References

- [1] S. M. Weiss and C. A. Kulikowski, *Computer Systems that Learn*, San Francisco, CA: Morgan Kaufmann, 1991.
- [2] A. A. Freitas and S. H. Lavington, *Mining Very Large Databases with Parallel Processing*, London, UK: Kluwer, 1998.
- [3] T. M. Cover and J. A. Thomas, "Elements of Information Theory, New York, NY: John Wiley & Sons, 1991.

- [4] M. Dorigo, G. Di Caro and L. M. Gambardella, "Ant algorithms for discrete optimization," *Artificial Life*, vol. 5, no. 2, pp. 137-172, 2000.
- [5] U. M. Fayyad, G. Piatetsky-Shapiro and P. Smyth, "From data mining to knowledge discovery: an overview," In: *Advances in Knowledge Discovery & Data Mining*, U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy (Eds.) Cambridge, MA: AAAI/MIT, pp. 1-34, 1996.
- [6] Dorigo, M., & Maniezzo, V. "The ant system: optimization by a colony of cooperating agent"s. *IEEE Transactions on Systems, Man, and Cybernetics*, 26(1), 1-13 1996.
- [7] Ziqiang Wang, Boqin Feng, *Classification Rule Mining with an Improved Ant Colony Algorithm*, Lecture Notes in Computer Science, Volume 3339, Jan 2004.
- [8] L. A. Brewlow and D. W. Aha, "Simplifying decision trees: a survey," *The Knowledge Engineering Review*, vol. 12, no. 1, pp. 1-40, 1997.
- [9] M. Dorigo, A. Colorni and V. Maniezzo, "The Ant System: optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, vol.26, no.1, pp.29-41, 1996.
- [10] M. P. Oakes, "Ant Colony Optimization for Stylometry: The Federalist Papers." *International Conference on Recent Advances in Soft Computing*, November 2004.
- [11] Bo Liu , Hussein A. Abbass , and Bob McKay " Classification Rule Discovery with Ant Colony Optimization " *IEEE Computational Intelligence. Bulletin* February 2004
- [12] Lotti Admane, Karima Benatchba, Mouloud KOUDIL, Habiba "Using ant colonies to solve data-mining problems" *DFUAS IEEE International Conference on Systems, Man and Cybernetics Said GHAROUT and Nacer HAMANI Institut National de formation en Informatique*, BP 68M, 16270, Oued Smar, Alg6rie. 2003.
- [13] Parepinelli, R. S., Lopes, . 'An Ant Colony Algorithm for Classification Rule Discovery'. In H. A. a. R. S. a. C. Newton (Ed.), *Data Mining: Heuristic Approach: IEEE Idea Group Publishing*.2002
- [14] Hettich, S. and Bay, S.D.(1999). *The UCI KDD Archive*, Retrieved September, 2002, from the World Wide Web:<http://kdd.ics.uci.edu>. 2002.
- [15] R. Shakerian, S. H. Kamali, M. Hedayati, M. Alipour , "Comparative Study of Ant Colony Optimization and Particle Swarm Optimization for Grid Scheduling", *Journal of mathematics and computer Science(JMCS)* 2011 Issue: 3 Pages: 469-474 (2011)
- [16] Rouhollah Maghsoudi, Arash Ghorbannia Delavar, Somayye Hoseyny, Rahmatollah Asgari, Yaghub Heidari , "Representing the New Model for Improving K-Means Clustering Algorithm based on Genetic Algorithm", *Journal of mathematics and computer Science(JMCS)* 2011 Volume: 2 (2011) Issue: 2 ,Pages: 329 - 336