Rule Discovery for Binary Classification Problem using ACO based Antminer

Sanjeev Gupta
Assistant Professor, CS&IT Dept.
Moradabad Institute of Technology
Moradabad, India

Sanjeev Bhardwaj
Assistant Professor, CS&IT Dept.
IFTM University
Moradabad, India

ABSTRACT

Data mining can be performed by number of ways. Classification is one of them. Classification is a data mining technique that assigns items to a predefined categories or classes or labels. The aim of classification is to predict the target class for the inputted data. On the other hand biology inspired algorithms such as Genetic Algorithms (GA) and Swarm based approaches like Particle Swarm Optimization (PSO) and Ant Colonies Optimization (ACO) were used in solving many data mining problems and currently the most prominent choice in the area of swarm intelligence. In this paper binary classification is considered as an area of problem and a modified AntMiner is used to solve the problem. The basic algorithm of AntMiner has been modified with a different classification accuracy function.

Keywords

Ant Colony optimization (ACO), Particle Swarm Optimization (PSO), Classification models

1. INTRODUCTION

Data is collected and available in every dimension of life. The field of data mining deals with conversion of raw data into useful information. One important technique of data mining is classification [14]. The objective of classification is to build one or more models based on the training data, which can correctly predict the class of test objects. Classification has several important applications in lives [1],[15],[22],[26]. Examples include customer behavior prediction, portfolio risk management, identifying suspects, medical applications, sports and fraud detection.

On the other hand biology inspired algorithms such as Genetic Algorithms (GA) and Swarm based approaches like Ant Colonies based algorithms have been successfully used by many researchers. Particle Swarm Optimization (PSO) [2] and Ant Colonies Optimization (ACO) are currently the most popular algorithms in the area of swarm intelligence. While Ant Colonies Optimization (ACO) algorithms were introduced around 1990 [8], [9], [10], [11].These algorithms were inspired by the behavior of ant colonies.

2. OVERVIEW OF ACO & CLASSIFICATION RULES

Ant Colony Optimization (ACO) is a field of swarm intelligence. ACO was designed to keep in mind the behavior of biological ant colonies. ACO was proposed by M. Dorigo et al. [8],[11] as meta-heuristic method for solving optimization problems.

The behavior of real ants was the source of motivation for the design of novel algorithms for the solution of optimization and distributed control problems. Individual ant looks simple but ant colonies are able to solve enough complex problems like finding short path from the nest to the food source. To achieve these complex tasks the whole swarm communicates indirectly with each other. The indirect mechanism is a special chemical known as pheromone.

ACO is encouraged by this foraging behavior of ant colonies, and well suited for discrete optimization problems such as quadratic assignment [27], job scheduling [28], subset problems [25], network routing [4], vehicle routing [7], graph coloring problem, bioinformatics [8] and data mining [6] which is the concern of this paper.

The structure of the classification rules is IF <conditions> THEN<class>. The <conditions> part is also called antecedent and is a logical combination of independent or predictor attributes. The form of the antecedent part is like term1 AND term2 AND... . Each term is a triplet like <attribute, operator, value>.For example <Outlook=Sunny>, <Wind=Strong> etc. The operator is a relational operator in the set like(=,<,>) etc. The <class> part is called consequent of the rule and contains the class label value predicted for cases whose predictor attributes satisfy the <conditions>or antecedent part of the rule.

Several methods have been proposed for the rule induction process such as ID3 [27], C4.5 [16], CN2 [5], CART [3] and others. All of these algorithms fall into two broad categories namely, sequential covering algorithms and simultaneous covering algorithms Simultaneous covering algorithms like ID3 and C4.5 generate the entire rule set at once, while the sequential covering algorithms like AQ15 and CN2 learn the rule set in an incremental fashion.

3. ACO BASED ANTMINERS AND RULE DISCOVERY

Parpinelli et al [24] incorporated ACO to rule induction algorithms by developing a new sequential covering algorithm called "AntMiner".

Several algorithms have been developed on the basis of the work proposed by Parpinelli et al for rule induction and shows the competitiveness of AntMiners in different versions with other comprehensible classifiers [17-21,23].

3.1 Overview Of Basic AntMiner

AntMiner discovers an ordered list of classification rules.An ant attempts to discover a rule by selecting terms for antecedent part. Term is decided probabilistically according to a heuristic function and pheromone amount for the term.

After construction of a rule, the ant updates the pheromone for the term selected by increasing the pheromone for selected term and

evaporates pheromone for the rest term. The best rule is selected among the rules constructed by the ants and added to the discovered rule set. The algorithm is repeated until the discovered rules cover a sufficient portion of the given dataset or predefined stopping criteria is not met. The algorithm is shown below in box.

```
Training Set(TS)=All Training Cases
WHILE(no of uncovered cases in the TS>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 than the
          previous no_rules_converge-1 ants)
          Select the best rule among all
          constructed rules;
          Remove the cases correctly covered by
          the selected rule from the TS;
END WHILE
```

3.2 Rule Construction

A rule is constructed incrementally by adding a terms to the current partial rule that an ant holds. As mentioned before, $term_{ij}$ is in form of $A_i = V_{ij}$, where A_i is *i-th* attribute and V_{ij} is the *j-th* value of the domain of A_i . The probability that termij is selected by the ant to be added to the current partial rule is given by the Eq.1.

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

Where:

oη_{ij} is the value of a problem-dependent heuristic function for $term_{ij}$.

oτ_{ij} is the current amount of the pheromone on $term_{ij}$ for the current ant through its current trial.

oa is the total number of attributes.

ob_i is the number of values in domain of the *i-th* attribute.

Higher the value of $\eta_{ij,}$ better the $term_{ij}$ higher the probability of it to be selected.

3.3 Heuristic Function

This heuristic function is based on information theory. The local heuristic function applied in Ant-Miner is an entropy measure of individual terms which is similar to the entropy used in C4.5.

$$H(W|A_i=V_{ij})=-\sum_{w=1}^{k}P\big(w\big|A_i = V_{ij}\big).log_2 P\big(w\big|A_i = V_{ij}\big) \tag{2}$$

Where:

oW is the class attribute and k is the number of class values,

oA_i is the *i*th attribute and V_{ij} is the *j*th attribute value of the *i*th attribute,

o$(P(w \mid A_i = V_{ij}))$ is the probability of observing class c conditional on observing $A_i = V_{ij}$.

Higher value of entropy means that value V_{ij} in attribute is more uniformly distributed among the classes, and so the lower the predictive power of $term_{ij}$. The terms to be selected should have a high predictive power to be added to the current partial rule. So higher value of entropy, smaller the probability of $term_{ij}$ to be added by an ant to its partial rule.

The ant prefers to choose a term with higher heuristic value.

A normalized heuristic function:

$$\eta_{ij} = \frac{log_2 k - H(C|A_i = V_{ij})}{\sum_{i=1}^{a} x_i . \sum_{j=1}^{b}(log_2 k - H(C|A_i = V_{ij}))} \tag{3}$$

Parameter a,b,k and x_i has the same meaning as previous.

3.4 Rule Quality

The quality of a rule, denoted by Q, is defined by the following formula called sensitivity/specificity:

$$Q = \frac{TP}{TP + FN} \cdot \frac{TN}{TN + FP} \tag{4}$$

where:

oTP (true positive) is the number of examples covered by the rule that belong to the class predicted by the rule.

oFP (false positive) is the number of examples covered by the rule that belong to a class different from the class predicted by the rule.

oFN (false negative) is the number of examples that are not covered by the rule, but that belong to the class predicted by the rule.

oTN (true negative) is the number of examples that are not covered by the rule and that do not belong to the class predicted by the rule.

3.5 Pheromone Initialization & Updation

The initial amount of pheromone deposited at each $term_{ij}$ is inversely proportional to the number of values of all attributes.

$$\tau_{ij}(t = 0) = \frac{1}{\sum_{i=0}^{a} b_i} \tag{5}$$

As an ant finishes constructing the rule, the amount of pheromone at each $term_{ij}$ in the construction rule is updated.

This pheromone updating process has two steps:

o Increasing the amount of pheromone associated with each term occurred in the constructed rule.

o Decreasing the amount of pheromone associated with each term not occurred in the constructed rule. This is known as pheromone evaporation.

Increasing of the pheromone can be done using the formula

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \tau_{ij}(t) . Q \qquad \forall i, j \in R_t \qquad (6)$$

where:

o $\tau_{ij}(t)$ is the pheromone level of the *term$_{ij}$* at iteration t.

o Q is the quality of the rule constructed

For the terms not used in constructed rule, pheromone level is decreased by dividing the pheromone value of the *term$_{ij}$* by the sum of all pheromone values.

These procedures, i.e. from B to E are repeated until every artificial ant has generated a rule, or the current rule has been generated by previous ants. The best rule among the rules generated by all ants is added to the induced rule set. The training dataset is appropriately updated by removing all the examples covered by the best rule.

4. PROPOSED WORK

In many real world problems, classification is used as one of the important decision making technique. Classification task can be used when a tuple or sample needs to be classified into a predefined set of classes based on some set of attributes. There are many real world problems that can be categorized as classification problems like weather forecasting, credit risk evaluation, medical diagnosis problem, bankruptcy prediction etc.

Generally, there are two types of classification problems: binary class problems and multiple class problems. In binary class problems a set of attributes is categorized to one class between two classes like a decision of Yes or No. While in a multiple classification problem a tuple is categorized to one class having number of classes as a solution like Class A, Class B and Class C, i.e. more than two classes as a solution or predicted class.

In basic AntMiner the quality function which has been used to measure the quality of a rule was according to Eq.4.

In this paper a different quality function is used which is simpler in mathematical operation i.e. less no of multiplication and division operation, suitable for binary classification and produces good results. The function is described in Eq.7.

$$Q = \frac{TP + TN}{TP + TN + FP + FN} \qquad (7)$$

All the parameters passed in quality function have the same meaning as in basic AntMiner.

5. EXPERIMENTAL RESULTS

The experiments are performed by using GUI AntMiner developed in Java[13]. The performance is measured using two public dataset having only two classes. The datasets are obtained from UCI Learning Repository [12]. The considered datasets have only two classes. The detail description of datasets is given in Table .The values of input parameters are taken randomly. Optimization of these variables values is still a future scope of the problem.
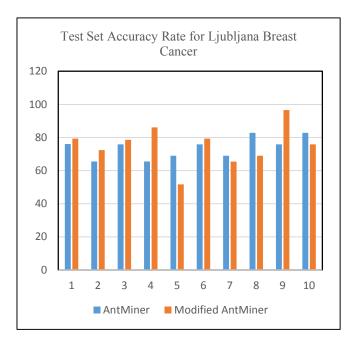
Table 1: Dataset Description

Data Sets	Instances	Attributes	Classes
Ljubljana Breast Cancer	286	10	2
Tic-Tac-Toe	958	10	2

Table 2: Input Parameters

Parameter Name	Value
Folds	10
Number of Ants	10
Min. Cases per Rule	5
Max. uncovered Cases	5
Rules for Convergence	5
Number of Iterations	10

Results after applying basic AntMiner and Modified AntMiner algorithm is shown in Table 3 and Table 4.Test set accuracy is shown in Table 5 and Table 6.

Table 3: Predictive Accuracy Comparison

	Ljubljana Breast Cancer	Tic-Tac-Toe
AntMiner	73.81% +/- 2.01%	73.57% +/- 2.71%
Modified AntMiner	75.61% +/- 2.89%	74.39% +/- 2.8%

Table 4: Number of Rules Discovered

	Ljubljana Breast Cancer	Tic-Tac-Toe
AntMiner	7.1 +/- 0.1	8.7 +/- 0.6
Modified AntMiner	7.5 +/- 0.22	9.2 +/- 0.44

Table 5: Test Set Accuracy Rate for Ljubljana Breast Cancer

Run No	AntMiner	Modified AntMiner
1	76.0	79.31
2	65.51	72.41
3	75.86	78.57
4	65.51	86.20
5	68.96	51.72
6	75.86	79.31
7	68.96	65.38
8	82.75	68.96
9	75.86	96.55
10	82.75	75.86

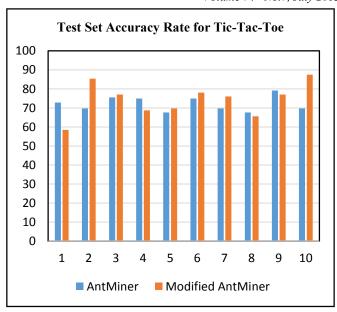**Fig 1: Test Set Accuracy Rate for Ljubljana Breast Cancer**

Fig 2: Test Set Accuracy Rate for Tic-Tac-Toe

6. CONCLUSION & FUTURE SCOPE

Data is a vital and valuable asset. There is always a demand for efficient and effective data mining algorithms and it is a subject of ongoing research, One important technique of data mining and of the concern of this paper is classification. Researchers are interested in finding efficient and accurate classification models that achieve higher accuracy rate and better performance in reasonable time. Various researchers has shown that Swarm Intelligence based AntMiner produces competitive results with the conventional algorithms.

On the other hand for binary classification same Antminer can be used with little modification. A little modification in quality function of basic AntMiner algorithm makes it effective for binary classification. The experimental results shown in previous indicates that it produces competitive results with basic AntMiner.

From the point of future scope lots of optimization is already done and lots of opportunity is still exists. The optimize behavior of real ants encourages more towards in solving of complex problems. Quality function, Pheromone updating criteria, Heuristic function may be optimize more to make it more effective in solving problems which has continuous attribute set. Optimization of input parameters may also be an area of future research.

Table 6: Test Set Accuracy Rate for Tic-Tac-Toe

Run No	AntMiner	Modified AntMiner
1	72.916	58.51
2	69.791	85.41
3	75.531	77.08
4	75.0	68.75
5	67.70	69.79
6	75.0	78.125
7	69.79	76.04
8	67.70	65.62
9	79.16	77.08
10	69.79	87.5

7. REFERENCES

[1] A.C. Tessmer, "What to learn from near misses: an inductive learning approach to credit risk assessment," Decision Sciences, Vol. 28, No. 1, pp. 105-120, 1997.

[2] Abraham, A., Grosan, C., Ramos V.: "Swarm Intelligence in Data Mining". Studies in Computational Intelligence, vol. 34, (2006).

[3] Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. Classification and Regression Trees. Belmont, CA: Wadsworth, 1984.

[4] Bremner D, Demaine E, Erickson J, Iacono J, Langerman S, Morin P, Toussaint G (2005). "Output-sensitive algorithms for computing nearest-neighbor decision boundaries". Discrete and Computational Geometry 33(4):593– 604.doi:10.1007/s00454-004-1152-0.

[5] Clark P, Boswell R, "Rule Induction with CN2: Some Recent Improvements". in Proceedings of the Fifth European Conference on Machine Learning, pages 151-163. Berlin, Springer-Verlag, 1991.

[6] D. Martens, M. de Backer, R. Haesen, J. Vanthienen, M. Snoeck, and B. Baesens, "Classification with ant colony optimization," IEEE Transactions on Evolutionary Computation, Vol. 11, No. 5. Oct. 2007

[7] Domingos, Pedro & Michael Pazzani (1997) "On the optimality of the simple Bayesian classifier under zero-one loss". Machine Learning, 29:103–137.

[8] Dorigo, M., Colorni, A., Maniezzo, V.: "The Ant System: Optimization by a colony of cooperating agents". IEEE Transactions on Systems, Man, and Cybernetics-Part B., vol. 26, pp. 29–41 (1996).

[9] Dorigo, M., Di Caro, G.: The ant colony optimization meta-heuristic, in New Ideas in Optimization. McGraw-Hill., p.11, (1999).

[10] Dorigo, M., Maniezzo, V., Colorni, A., "Positive Feedback as a Search Strategy, Technical report". Dipartimento di Elettronica, pp. 91-016 (1991).

[11] Dorigo, M.: "Optimization Learning and Natural Algorithms, Ph.D. thesis (in Italian)". Dipartimento di Elettronica, Politecnico di Milano. (1992).

[12] http://www.ics.uci.edu/mlearn/MLRepository.html

[13] http://www.joinville.udesc.br

[14] J. Han, and M. Kamber, Data Mining: Concepts and Techniques, 2nd ed., Morgan, Kaufmann Publishers, 2006.

[15] J. Pesce, "Stanching hospitals, "Financial hemorrhage with information technology," Health Management Technology, Vol. 24, No. 8, pp. 6-12, 2003.

[16] J.R .Quinlan, C4.5: Programs for Machine Learning. Los Altos, California: Morgan Kauffman, 1994.

[17] K. Salama and A. Abdelbar, "Extensions to the Ant-Miner Classification Rule Discovery Algorithm," in Proceedings of the 7th International Conference on Swarm Intelligence (ANTS 2010), Lecture Notes in Computer Science 6234, 2010, pp. 167–178.

[18] K. Salama, A. Abdelbar, and A. Freitas, "Multiple pheromone types and other extensions to the ant-miner classification rule discovery algorithm," Swarm Intelligence, vol. 5, no. 3-4, pp. 149–182, 2011.

[19] Liu, H.A. Abbass, and B. McKay, "Classification rule discovery with ant colony optimization," in Proceedings of IEEE/WIC International Conference on Intelligent Agent Technology, 2004, pp. 83–88.

[20] Liu, H.A. Abbass, and B. McKay, "Classification rule discovery with ant colony optimization," IEEE Computational Intelligence Bulletin, Vol. 3, No. 1, Feb. 2004.

[21] Liu, H.A. Abbass, and B. McKay, "Density-based heuristic for rule discovery with ant-miner," in Proceedings of 6th Australia-Japan Joint Workshop on Intelligent Evolutionary Systems. Canberra, Australia, 2002.

[22] M.J. Berry, and G. Linoff. Data Mining Techniques for Marketing, Sales, and Customer upport. New York: John Wiley, 1997.

[23] N. Holden, and A.A. Freitas, "A hybrid PSO/ACO algorithm for classification,"in Proceeding of GECCO-2007 Workshop on Particle Swarms: The Second Decade, ACM Press, New York, 2007.

[24] Parpinelli, R.S., Lopes, H.S., Freitas, A.: ―Data mining with an ant colony optimization algorithm‖. IEEE Transactions on Evolutionary Computation, vol. 6, pp.321–332 (2002).

[25] R.E. Schapire, and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," Machine Learning, Vol. 37(3), pp. 297–336, 1999.

[26] W. Ceusters, "Medical natural language understanding as a supporting technology for data mining in healthcare," Chapter 3 in: Cios K.J., eds. Medical Data Mining and Knowledge Discovery, Heidelberg: Springer-Verlag, pp. 32-60, 2000.

[27] W. Cohen, "Fast effective rule induction," in Machine Learning: Proceedings of the Twelfth International Conference (ML95), pp. 852-857, 1995.

[28] Z. Wu and C. Li, "Feature selection for classification using transductive support vector machines", in Feature Extraction, Foundations and Applications, Springer, Verlag, Berlin, 2006.