International Journal of Electronics Communication and Computer Engineering
Volume 3, Issue 2, ISSN 2249 –071X

# Comparative Study of All Sub Steps of Ant-Miner Algorithm

**Swati B. Patil**
Asstt. Professor
Vishawakarma Institute of Information Technology, Information Technology Department, Kondhwa (B.K.), Pune

**Snehal Rathi**
Asstt. Professor
Vishawakarma Institute of Information Technology, Information Technology Department, Kondhwa (B.K.), Pune

**Saroj Kulkarni**
Asstt. Professor
Vishawakarma Institute of Information Technology, Information Technology Department, Kondhwa (B.K.), Pune

**Abstract - This paper proposes an algorithm for data mining called Ant-Miner which shows the importance of sub steps by results. The goal of Ant-Miner is to extract classification rules from data. Rules are more accurate and omprehensible to the user when algorithm executes with all substeps.In this paper I have shown the results by removing the sub steps.**

**Keywords - Ant Colony Optimization Algorithm (Ant-Miner), data mining, Rule pruning, Pheromone updatation.**

## I. INTRODUCTION

The goal of data mining is to extract knowledge from data. At the same time we have to check the accuracy of extracted data. To improve the accuracy of Ant-miner first I have checked in this paper the importance of sub steps of Ant-Miner. By removing the sub steps I have checked the accuracy parameter so that it can be extended to improve the further accuracy in which substeps means rule pruning or pheromone updatation.

The rest of this paper is organized as follows. Section II presents a experimental approach. Section III brief description of Ant-miner Section IV discusses experimental results Section V conclusion VI Acknowledgement Section VII References & Bibliography.

## II. EXPERIMENTAL APPROACH

Ant-miner extracts classification rule from data and checked the accuracy but it's not 100% .So by processing in which steps I can increases the percentage of accuracy.

### *A. Collecting the experimental data*

For the purpose of comparing the performance of all substeps of Ant-Miner, data sets are obtained from the UCI (University of California at Irvine) Machine Learning Repository <http:// www.ics.uci.edu/mlearn/ MLRepository.html>)[3].

In total, six data sets are used. The main properties of the data sets are summarized in Table 2.1. The rest column of this table gives the data set name, the second column the total number of examples in the data set, the third column the number of predicting categorical attributes, the fourth column the number of predicting numerical attributes, and finally the fifth column gives the number of class values of the data[6].

| Data Set | Total # Examples | Nominal Attributes | Continuous Attributes | Classes |
|---|---|---|---|---|
| Ljubljana breast cancer | 286 | 9 | - | 2 |
| Wisconsin breast cancer | 699 | - | 9 | 2 |
| Tic-tac-toe | 958 | 9 | - | 2 |
| Dermatology | 366 | 33 | 1 | 6 |
| Hepatitis | 155 | 13 | 6 | 2 |
| Cleveland heart disease | 303 | 8 | 5 | 5 |

Table 1 Data sets used in the Experiments:

## III. APPLICATION OF ANT-MINER ALGORITHM TO CHECK ACCURACY

### *A. Description of Ant-miner:*

Ant-miner Algorithm mainly consists of 3 sub steps.

1) Rule construction
2) Rule Pruning
3) Pheromone updatation

This is the description of Ant—miner Algorithm Begin Initialize with empty list;

Evaluate population;

While (condition satisfies) repeat Rule Construction;Rule pruning; Pheromone updatation. until (termination criteria); end.

Fig. 1 A pseudo-code for a Simple Ant-miner

Conditions applied to algorithm:

No_of_ants = 3000;

Min_cases_per_rule = 10;

Max_uncovered_cases = 10;

No_rule_converg = 10;

### *B. Rule Construction*

Ant-miner constructs rule by calculating entropy and information-theoretic heuristic function. By substituting these values it calculates the probability of selecting the terms and forms the rules [1].

### *C. Rule Pruning*

Rule pruning means to remove the unwanted terms. I have calculated the accuracy parameter of Ant-miner with pruning and without pruning.

### *D. Pheromone Updation*

In order to exchange information about which path should be followed, ants communicate with each other by means of

chemical substance known as pheromone. Again checked the accuracy parameter of Ant-miner with and without pheromone updatation.

## IV EXPERIMENTAL RESULTS

### *A. Results of Rule pruning*

Rule pruning is a necessary process to avoid over fitting to noisy training data. For decision lists (Rules constructed by Ant-Miner) there are also two ways for rule pruning.

The first is builds the entire rule sets then eliminates attributes from individual rule. The second is simplifying rule immediately after it has been generated.

Ant-Miner, the algorithm implemented in this project adopts second method. In order to analyze the influence of the rule pruning execute Ant-Miner algorithm, Ant-Miner was also run without rule pruning.

To make a fair comparison between Ant-miner with and without pruning, the experiments without pruning used same parameter only sub step of pruning is removed. It has seen that Ant-miner without pruning obtained a lower predictive accuracy than Ant-miner with pruning in two data sets, namely, Wisconsin breast cancer and Tic-tac-toe.

On the other hand, Ant-Miner without pruning obtained a predictive accuracy somewhat larger than Ant-Miner with Pruning in one data set, namely, Dermatology. In the other data sets, Ljubljana breast cancer, Hepatitis and Cleveland heart disease, the predictive accuracy with and without pruning was almost equal.

Results shows (Table 3.1 & Table 3.2) that Ant-Miner with pruning and without pruning shows that rule pruning seems to be beneficial more often than it is harmful, concerning predictive accuracy. But also we can say this rule pruning procedure is fine in the case where numbers of attributes are less than number of instances. If the attributes will increase e.g. news where number of attributes are more than instances such a pruning procedure will not work efficiently. So we will use Hybrid Rule Pruner for Ant-Miner which is a future scope of this paper [2].

The result of Ant-Miner without pheromone and with pheromone is reported in Table 4.1 & Table 4.2. Results show that Ant-Miner without pheromone achieved a lower predictive accuracy than Ant-miner with pheromone in also rules discovered by pheromone updation process is simple and comprehensible to the user.

### *B. Pheromone Updation Process*

Initially each term consist of same amount of pheromone. Whenever ant constructs the rule that rule is pruned. The amount of pheromone must be updated. This updation is done by 2 ways.

1) Amount of pheromone is increased for those terms which are present after rule pruing.
2) Amount of pheromone is decreased for those terms which are not present in rule after rule pruing.
3) Increasing the Pheromone of Used Terms The quality of rule determines the amount of pheromone.
   Q = sensitivity • specificity
   Where *sensitivity = TP/ (TP + FN)*
   *Specificity = TN/(FP+TN)*
   - *TP* (true positives) is the number of cases covered by the rule that have the class predicted by the rule.
   - *FP* (false positives) is the number of cases covered by the rule that have a class different from the class predicted by the rule.
   - *FN* (false negatives) is the number of cases that are not covered by the rule but that have the class predicted by the rule.
   - *TN* (true negatives) is the number of cases that are not covered by the rule and that do not have the class predicted by the rule.
4) Decreasing the Pheromone of Unused Terms
   The amount of pheromone of unused term is decreasing to simulate pheromone evaporation. Pheromone evaporation for unused terms is achieved by normalizing the value of each pheromone $\tau_{ij}$.[5].

To make a fair comparison between Ant-Miner with pheromone and without pheromone, the experiments without pheromone also used the parameter setting in Algorithm described in Figure 4.1 only sub step of pheromone updation is removed.

Four data sets, namely, Ljubljana breast cancer, Wisconsin breast cancer, Tic-tac-toe and Hepatitis. For other two datasets, namely, Dermatology and Cleveland heart disease where Predictive accuracy was almost same for both versions of Ant-Miner. With respect to the simplicity of the discovered rule lists, there is not much difference Ant-Miner without pheromone and Ant-Miner with pheromone. Overall, the result shows that pheromone updating procedure is important to improve the predictive accuracy of the rules discovered by Ant-Miner.

Therefore, we can say that in Ant-Miner algorithm all the sub steps i.e. rule pruning and pheromone updation are important. Therefore we can say that Ant-Miner with pruning and with pheromone updation seems particularly advantageous when it is important to obtain accurate and comprehensible rule.

| Data set | Ant-Miner Without Pruning Predictive accuracy (%) | Ant-Miner With Pruning Predictive accuracy (%) |
|---|---|---|
| Ljubljana breast cancer | 77.44% | 75.17% |
| Wisconsin breast cancer | 91.42% | 92.70% |
| Tic-tac-toe | 76.92% | 97.80% |
| Dermatology | 85.82% | 79.80% |
| Hepatitis | 83.83% | 83.83% |
| Cleveland heart disease | 72.27% | 72.31% |

Table 2 Comparison of the predictive accuracy of rules generated using Ant-Miner without pruning and with pruning.

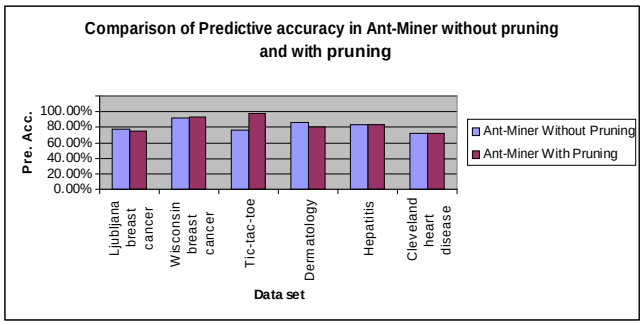


Figure 2 A graph showing the Comparison of the predictive accuracy of rules generated using Ant-Miner without pruning and with pruning.

| Data set | Ant-Miner Without Pruning Predictive accuracy (%) | Ant-Miner With Pruning Predictive accuracy (%) |
|---|---|---|
| Ljubljana breast cancer | 77.44% | 75.17% |
| Wisconsin breast cancer | 91.42% | 92.70% |
| Tic-tac-toe | 76.92% | 97.80% |
| Dermatology | 85.82% | 79.80% |
| Hepatitis | 83.83% | 83.83% |
| Cleveland heart disease | 72.27% | 72.31% |

Table 3 Comparison of Predictive accuracy in Ant-Miner without pheromone and with pheromone.

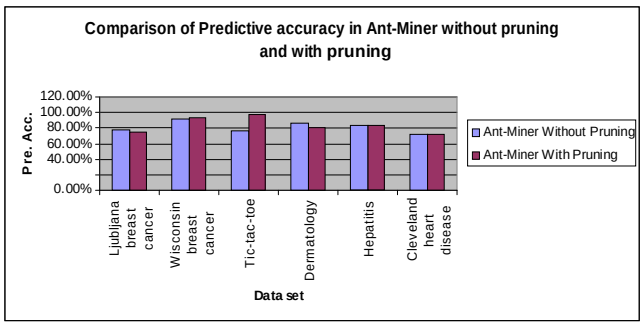


Figure 3 A graph showing the Comparison of the predictive accuracy of rules generated using Ant-Miner without pheromone and with pheromone.

## V. Conclusion

In this paper, we have compared the results of Ant-miner with and without rule pruning as well as with and without pheromone updation in six public domain data sets. The results show that for predictive accuracy of Ant-Miner obtained better results with rule pruning and with pheromone updation. But still if we want to increase the accuracy, from results we can say that by enhancing rule pruning procedure we will improve more compare to enhancing the pheromone updation procedure.

## VI. Acknowledgments

Balaprakash Doctoral researcher in the CACHE Institute project, for his continuous encouragement, patience, and promptness in answering all my questions and offering his

### *A. Directions for Future Work*

First, it would did not involve datasets with a large number of attributes like BBC news so to extend Ant-Miner to cope with Hybrid rule pruner and this helps to form more comprehensible rules and also increase the percentage of accuracy [2].

Second, we can modify heuristic function so that it will increase predictive accuracy.

## References

[1] Rafael S. Parpinelli, Heitor S. Lopes, Member, IEEE, and Alex A. Freitas Data Mining With an Ant Colony Optimization Algorithm. IEEE Transactions on evolutionary computing, vol.6, no. 4, August 2002.

[2] A New Classification-Rule Pruning Procedure for an Ant Colony Algorithm Allen Chan and Alex Freitas Computing Laboratory, University of Kent, Canterbury, CT2 7NF, UK

[3] UCI (University of California at Irvine) Machine Learning Repository( < http://www.ics.uci.edu/ mlearn/MLRepository.html >).

[4] M. Bohanec and I. Bratko, "Trading accuracy for simplicity in decision trees," Machine Learning, vol. 15, pp. 223-250, 1994.

[5] E. Bonabeau, M. Dorigo and G. Theraulaz, Swarm Intelligence: From Natural to Artificial Systems. New York, NY: Oxford University Press, 1999.

[6] L. A. Brewlow and D. W. Aha, "Simplifying decision trees: a survey," The Knowledge Engineering Review, vol. 12, no. 1, pp. 1-40, 1997.

[7] J. Catlett, "Overpruning large decision trees," In: Proceedings International Joint Conference on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 1991.

## Books

[1] Pang-Ning Tan, Michael Steinbach and Vipin Kumar.\Introduction to Data Min-ing (First Edition)", Perason Education, Inc.in 2007.

[2] K. P. Soman, Shyam Diwakar and V. Ajay.\Insight into Data Mining Theory and Practice (First Edition)", Prentice-Hall India, Inc. in 2006.

[3] Ian H. Written and Eibe Frank,\Data Mining Practical Machine Learning Tools and Techniques(Second Edition)", Elsevier Publication, inc. in 2005.

[4] Jawai Han and M. Kamber.\Data Mining Concepts and Techniques (Second Edi-tion)",Morgan Kaufmann Publisher, Inc. in 2000.

[5] URL : www.java.sun.com, Information obtained, Core Java Language.

[6] URL : www.javaworld.com, Information obtained, Core Java Language

## AUTHOR'S PROFILE

**Snehal R.Rathi**
Asstt. Professor
Vishawakarma Institute of Information Technology,
Information Technology Department,Kondhwa (B.K.),
Pune-411048

**Swati Bula Patil**
Asstt. Professor
Vishawakarma Institute of Information Technology,
Information Technology Department,Kondhwa (B.K.),
Pune-411048
Member of LMISTE, MCSI
i. Providing Data security in WLAN by detecting unauthorized access points and attacks International Journal of engineering science and technology(IJEST) ISSN : 0975-5462 May 2011
ii. Detection of Illegal access points using filters and network attacks using signatures for providing security in WLAN International J. of Engg. Research & Indu. Appls. (IJERIA). ISSN 0974-1518 August 2011

**Saroj M. Kulkarni**
Asstt. Professor
Vishawakarma Institute of Information Technology,
Information Technology Department,Kondhwa (B.K.),
Pune-411048