International Journal of Information Processing, 7(3), 45-57, 2013
ISSN : 0973-8215
IK International Publishing House Pvt. Ltd., New Delhi, India

# An Ant-Miner System without Rule Pruning for Discovering Classification Rules

Saroj, Renu Bala, Jyoti[a]

[a]Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology 125001 India,
Contact: ratnoo.saroj@gmail.com, renu0805@gmail.com, kwatra.jyoti@gmail.com

Ant Colony algorithms have been employed successfully to discover classifications rules with high predictive accuracy and comprehensibility. In this context, Ant Colony-based Data Mining Algorithm, named as Ant-Miner, is a premier and leading contribution. Ant-Miner discovers rules with predictive accuracy competitive with the other techniques of classification rule discovery like CN2 and decision tree rule induction. It discovers lesser number of rules with smaller number of terms per rule. Hence, the performance of Ant-Miner is better on the index of comprehensibility. Ant-Miner constructs rules by adding terms (attribute-value conditions) probabilistically one by one. Subsequently, it removes the irrelevant terms through a rule pruning procedure. In fact, rule pruning step involves $k^3$ database scans for a rule with $k$ terms. A database scan is exorbitantly expensive operation for most of the data mining applications. A large number of database scans involved in rule pruning step makes application of Ant-Miner infeasible to bigger datasets with larger number of attributes. This paper proposes an Ant-Miner system which eliminates the pruning step completely and improves run time efficiency of the earlier algorithm. The performance of the proposed Ant-Miner without rule pruning is compared to the original Ant-Miner and one of its important extensions as AntMiner+ with respect to predictive accuracy and comprehensibility across several datasets. A comparison is also made between running time of Ant-Miner and the proposed Ant-Miner without rule pruning. The results obtained show that the suggested Ant-Miner system is competitive with Ant-Miner and AntMiner+ in terms of predictive accuracy. It performs better on comprehensibility and achieves manifold gains in running time.

**Keywords:** Ant Colony Optimization, Ant-Miner, Data Mining, Discovery of Classification Rules, Swarm Intelligence.

## 1. INTRODUCTION

Swarm Intelligence is an area of computational intelligence that draws its inspiration from collective behavior exhibited by social insects. A swarm is defined as a set of mobile agents which communicate often indirectly by acting on their local environment. Individuals like ants, bees, termites and wasps are homogeneous organisms working in swarms, follow few simple rules and do not possess much cognitive ability. Yet, as a group they carry out a distributed problem solving without any centralized control and are able to accomplish complex tasks like building nests, collecting honey and finding shortest paths to a food sources from their nests. Swarm intelligence has been recently applied to several optimization problems in data mining.

An Ant Colony Optimization(ACO) algorithm is a popular technique that comes under the umbrella of swarm intelligence. The ACO algorithms are based on foraging behavior of real ants and employed to solve discrete optimization problems with large search spaces. The Ants are able to indirectly communicate through pheromone (a chemical substance which every ant deposits while operating in its environment) trails to find the shortest path between a food source and their nest. Initially, ants start their search through random move-

Table 4
Running Time of Ant-Miner and Ant-Miner Without Rule Pruning

| Datasets | Time(sec) | #Dbs | Avgerage Terms | Time(Sec) | #Dbs Proposed work | Max Terms | Speed Gain |
|---|---|---|---|---|---|---|---|
| Dermatology | 69.27 | 426529 | 17.5 | 3.61 | 22179 | 8 | 19.19 |
| Hepatitis | 1.44 | 23436 | 8.5 | 0.109 | 1749 | 3 | 13.21 |
| Vote | 1.84 | 14916 | 7.5 | 0.26 | 2088 | 4 | 7.08 |
| Mushroom | 132.66 | 66531 | 10.16 | 12.06 | 5981 | 5 | 11.00 |
| Nursery | 36.957 | 17304 | 4.2 | 7.62 | 3534 | 2 | 4.85 |

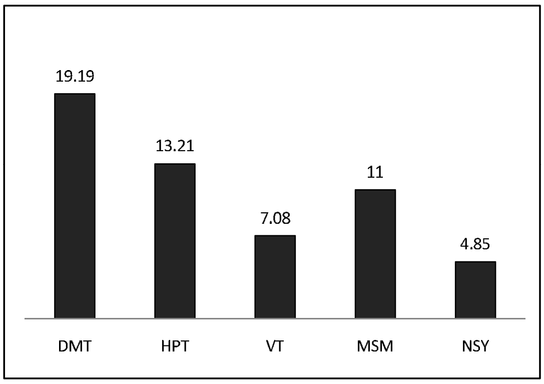


Figure 4. Speed Gain

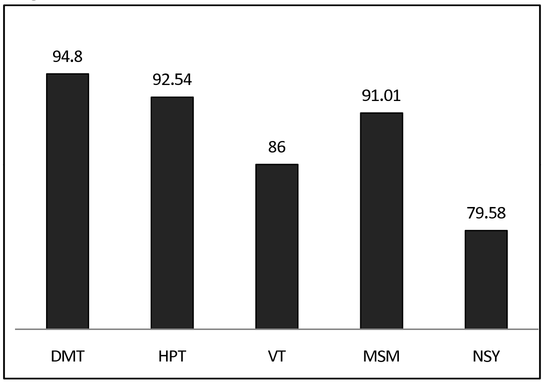


Figure 5. Percent Database Scan Saved

curacy, simplicity and running time to that of original Ant-Miner and AntMiner+. The suggested ACO based rule mining algorithm has achieved competitive predictive accuracy rates. It has performed better on the index of simplicity and run time as a large number of database scans are prohibited by the new approach.

We have noticed that all the Ant-Miner implementations use sequential covering algorithm. As a result successive rules are learnt from the remaining objects in the dataset after removal of the objects which get covered by the rules discovered in earlier iterations of the algorithms. While working on the Mushroom datasets, we noticed some conflicting rules, like '*if* (odor = none) *then* Mushroom is edible' and '*if* (odor = none) *then* Mushroom is poisonous', are being getting discovered because sequential covering algorithm is not able to cope up with problem of attribute and rule interactions. The second rule '*if* (odor = none) *then* Mushroom is poisonous' gets discovered due to interaction of the attribute value pair 'odor = none' with some other exceptional conditions arising out of several smaller disjunts in the dataset. It would be interesting to discover such rules in the rule + exceptions framework as given in [30,31].

$$dbs\ saved = \frac{\#dbs(AM) - \#dbs(PAM)}{dbs(AM)} \times 100 \quad (10)$$

## REFERENCES

1. Dorigo M, Maniezzo V and Colorni A. The Ant System: Optimization by a Colony of Cooperating Agents, *IEEE Transactions On Systems, Man And Cybernetics-Part B*, 26(1):29−41, 1996.
2. Dorigo M, Caro G D and Gambardella L M.

Ant Algorithms for Discrete Optimization, *Artificial Life*, 5:137−172, 1999.

3. Dorigo M and Sttzle T. Ant Colony Optimization, *MIT Press*, 2004.
4. Dorigo M and Stutzle T. Any Colony Optimization: Overview and Recent Advances, *In Handbook of Meta-heuristics, Gendreau M, Potvin J. (eds.)*, Springer, pages 227−263, 2010.
5. Dorigo M, Birattari M, Blum C, Gambardella L M, Mondada F and Sttzle T (Eds.). Lecture Notes in Computer Science, *Ant Colony Optimization and Swarm Intelligence*, 3172, Springer, Brussels, Belgium, 2004.
6. Tsai C-F, Tsai C-W, Wu H-C and Yang T. ACODF: A Novel Data Clustering Approach for Data Mining in Large Databases, *Journal of Systems and Software*, 73(1):133−145, 2004.
7. Parpinelli R S, Lopes H and Freitas A A. Ant Ant Colony based System for Data Mining: Applications to Medical Data, *In Spector L E, Goodman E et al. (Eds.), Proceeding of Genetic and Evolutionary Computation Conference (GECCO-2001)*, San Francisco, USA, Morgan Kaufmann, pages 791−798, 2001.
8. Parpinelli R S, Lopes H S and Freitas A A. Data Mining with an Ant Colony Optimization Algorithm, *IEEE Transactions on Evolutionary Computation*, 6:321−332, 2002.
9. Liu B, Abbas Hussein A and McKay Bob. Density-based Heuristic for Rule Discovery with Ant-Miner, *In Proceedings of 6th Australasia-Japan Joint Workshop on Intelligent and Evolutionary Systems*, pages 180−184, 2002.
10. Liu B, Abbas Hussein A, and McKay B. Classification Rule Discovery with Ant Colony Optimization, *IEEEWIC International Conference on Intelligent Agent Technology (IAT 2003)*, 3(1):83−88, 2003.
11. Chan A and Freitas A A. A New Classification-rule Pruning Procedure for an Ant Colony Algorithm, *In Proceedings of 7th International Conference on Artificial Evolution*, Springer-Verlag Berlin,Heidelberg, pages 25−36, 2006.
12. Martens D, Backer M De, Haesen R, Baesens B and Holvoet T. Ants Constructing Rule based Classifiers, *In Abraham A, Grosan C and Ramos V (Eds.), Studies in Computational Intelligence: Swarm Intelligence in Data Mining*, 34, 21−43, Springer-Verlag, Berlin, 2006.
13. Martens D, De Backer M, Haesen R, Vanthienen J, Snoeck M and Baesens B. Classification with Ant Colony Optimization, *IEEE Transactions on Evolutionary Computation*, 11(5):651−665, 2007.
14. Yu J, Chen Yun and Wu J. Modeling and Implementation of Classification Rule Discovery by Ant Colony Optimization for Spacial Land use Suitability Assessment, *Computer Environment and Urban Systems*, 35:308−319,2011.
15. Otero F E B, Freitas A A and Johnson C G. A New Sequential Covering Strategy for Inducing Classification Rules with Ant Colony Algorithms, *IEEE Transactions on Evolutionary Computations*, 17(1):64−76, 2013.
16. Witten I and Frank E. Data Mining: Practical Machine Learning Tools and Techniques, *Morgan Kaufmann, 2nd ed.*, 2005.
17. Freitas A A. Data Mining and Knowledge Discovery with Evolutionary Algorithms, *Springer-Verlag*, 2002.
18. Jiang Yi, Wang L and Chen Li. Discovering Interesting Classification Rules with Particle Swarm Algorithm, *In AdIshikawa, Y. et al. (Eds.),LNCS*, Springer-Verlag, Berlin Heidelberg, 4977, 65−73, 2008.
19. Matrtens D, Beaesens B and Fawcett T. Editorial Survey: Swarm Intelligence in Data Mining,*Machine Learning*, 82(1):1−42, 2011.
20. Otero F, Freitas A A and Johnson C. cAnt-Miner: An Ant Colony Classification Algorithm to Cope with Continuous Attributes, *In Dorigo M, Birattari M, Blum C, Clerc M, Stutzle T and Winfield A. (Eds.), Proceedings of the 6th International Conference on Swarm Intelligence (ANTS-2008), LCNS*, 5217, 48−59,Springer-Verlag, Berlin.(2008).
21. Otero F, Freitas A A and Johnson C. Handling Continuous Attributes in Ant Colony Classification Algorithms, *In Proceedings of the IEEE Symposium on Computational Intelligence in Data Mining (CIDM-2009)*, Nashville, USA, 225−231, 2009.
22. Galea M and Shen Q. Simultaneous Ant Colony Optimization Algorithms for Learning Linguistic Fuzzy Rules, In Abraham A, Grosan C and Ramos V. (Eds.), *Swarm Intelligence in Data Mining*,Springer Berlin, Heidelberg, 34, 75−99, 2006.
23. Ganji M F and Abadeh M S. A Fuzzy Classification System based on Ant Colony Optimization for Diabetes Disease Diagnosis, *Expert Systems with Applications*, 38:14650−14659, 2011.

24. Holden N and Freitas A A. A Hybrid Particle Swarm/Ant Colony Algorithm for the Classification of Hierarchical Biological Data, *In Proceedings of IEEE of Swarm Intelligence Symposium (SIS 2005)*, pages 100−107, 2005.
25. Salama K, Abdelbar A and Freitas A A. Multiple Pheromone Types and Other Extensions to the Ant-Miner Classification Rule Discovery Algorithms, *In Dorigo M. et al.(Eds.), Swarm Intelligence, LNCS*, 5(3-4), 149−182, 2011.
26. Sttzle T and Hoos H. Improving the Ant System: A Detailed Report on the Max-Min Ant System, *Technical Report, Department of Computer Science, Technical University of Darmstadt*, 1996.
27. Sttzle T and Hoos H. MAX-MIN Ant System, *Future Generation Computer Systems*, 16(9):889−914, 2000.
28. Asuncion A and Newman D. UCI Machine Learning Repository *University of California, Irvin, School of Information and Computer Sciences*, (online) Available at: http://www.ics.uci.edu/ mlearn/MLRepository.html.
29. Ant Colony Optimization Public Software (online) Available at: http://iridia.ulb.ac.be/ mdorigo/ACO/aco-code/public-software.html.
30. Yogita, Saroj, Kumar D and Pal Vipin. Rules + Exceptions: automated discovery of comprehensible decision rules, *In Proceeding of IEEE International Advance Computing Conference, IACC 2009*, Patiala, India, pages 1479−1484, 2009.
31. Saroj and Bharadwaj K K. Discovery of Exceptions: a step towards perfection, *In Proceedings of IEEE Workshop on Data Mining and Artificial Intelligence held in Conjunction with Third International Conference on Network and System Security*, Gold Coast, Australia, pages 540−545, 2009.

**Dr. Saroj** is a Professor in the Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Hisar, Haryana. She earned her postgraduate degree in Computing Science from University of London. She obtained her Ph.D from Jawaharlal Nehru University, New Delhi. She has published in refereed International Journals and Conference Proceedings.

**Ms Renu Bala** is pursuing her doctoral research in the Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Hisar, Haryana. She is a Junior Research Fellow in the Department.

**Ms Jyoti** is also pursuing her doctoral research in the Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Hisar, Haryana. She earns her Fellowship from TEQIP scheme.