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

Saroj, Renu Bala, Jyoti

Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology 125001 India,
Contact: ratn00sar@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

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

Figure 4. Speed Gain

Figure 5. Percent Database Scan Saved

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 disjuncts in the dataset. It would be interesting to discover such rules in the rule + exceptions framework as given in [30,31].

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

REFERENCES
2. Dorigo M, Caro G D and Gambardella L M.


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.