A Model Based on Ant Colony System and Rough Set Theory to Feature Selection

R. Bello
Computer Science Dept, Universidad Central de Las Villas, Cuba, 53-0142-281416
rbellop@uclv.edu.cu

A. Nowe
CoMo Lab, Computer Science Dept, Vrije Universiteit Brussel, Belgium
asnowe@info.vub.ac.beI

Y. Caballero
Informatics Dept, Universidad de Camagüey, Cuba, 53-0142-281416
gpmtn@shine.cmw.sld.cu

Y. Gómez
Computer Science Dept, Universidad Central de Las Villas, Cuba,
ygomezd@uclv.edu.cu

P. Vrancx
CoMo Lab, Computer Science Dept, Vrije Universiteit Brussel, Belgium
pvrancx@vub.ac.be

ABSTRACT

In this paper we propose a hybrid approach to feature selection based on Ant Colony System algorithm and Rough Set Theory. Rough Set Theory offers the heuristic function to measure the quality of a single subset. We have studied the influence of the setting of the parameters for this problem, in particular for finding reducts. Experimental results show this hybrid approach is a promising method for features selection.

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning – *Induction.*

General Terms

Algorithms, Experimentation.

Keywords

Feature selection, reducts, ACO meta-heuristic, rough set theory.

1. INTRODUCTION

Feature selection (FS) is useful in different computational tasks, for instance, in machine learning processes. An appropriate representation space for learning by selecting relevant attributes to the problem domain is a crucial issue for learning systems. FS is useful to reduce the dimensionality of training sets (TS); it results not only in improving the speed of data manipulation, but even in improving the classification rate by reducing the influence of noise.

FS can be viewed as a particular case of a more general subset selection problem in which the goal is to find a subset maximizing some performance criterion, such as accuracy of classification. Feature selection methods search through the subsets of features and try to find the best subset among the competing 2^N-1 candidate subsets according to some evaluation measure, where N denotes de number of features. All FS methods contain two important components: an Evaluation function used to judge the performance of a candidate feature subset and a Search algorithm to search through the feature space.

The approach presented in this paper is based on the Rough Sets Theory (RST) and the Ant Colony Optimization meta-heuristic (ACO). The former is used to build the evaluation function and the latter to implement the search method.

An important issue in the RST is feature reduction based on the reduct concept. A reduct is a minimal set of features that preserves the partitioning of universe and hence the ability to perform classifications. Usually smaller reducts are preferable. However, the problem of finding a globally minimal reduct for a given information system is NP-hard. For that reason, heuristic methods for calculating near optimal reducts have been developed.

R. Jensen and Q. Shen [2] have proposed a method which combines ACO and RST to find reducts with promising results. Our research work addresses the same issue, but we propose an a different ant based approach and we have developed a study on thesetting of the parameters of this algorithm apllied to the problem of finding reducts because the setting of parameters is crucial for the performance of the ant's algorithm [1]. The ACO is used to generate subsets of features using a filter approach based on forward selection. The process starts with an empty set of attributes and successively adds the one with the best performance. RST offers the heuristic function to measure the quality of one feature subset.

2. ABOUT ANT COLONY OPTIMIZATION AND ROUGH SET THEORY.

ACO [1] is a master strategy (metaheuristic) used to guide other heuristics in order to obtain better solutions than those that are generated by local optimality methods. Ant System (AS) is the first ACO algorithm. Ant Colony System (ACS) is an advanced successor algorithm of AS. ACS uses an extra pseudorandom proportional rule (expression 2) in which parameters Alpha denotes the relative weight of trail intensity, Beta point outs the relative weight of heuristic value, and a new parameter q_0 (tunable parameter of transition rule) is introduced.

Let SD=(U, A$\cup$\{d\}) a TS, where U is a set of examples, A is a non-empty set of features, and d$\notin$A is the decision feature. Let B$\subseteq$A and X$\subseteq$U. B defines an equivalence relation. The set X can be approximated using only the information contained in B by constructing the B-lower and B-upper approximations of X, denoted by B_*X and B^*X respectively, where B_*X=\{ x : $[x]_B \subseteq X$

} and $B^*X=\{ x : [x]_B \cap X \neq \phi \}$, and $[x]_B$ denotes the class of x according to B-indiscernible relation. The objects in B_*X are sure members of X, while the objects in B^*X are possible members of X. If $B^*X - B_*X$ is not empty, then X is a rough set.

RST [4] offers several measures about TS. These measures are based on B-lower and B-upper approximations. Among them is the quality of the approximation of classification (expression 1). It expresses the percentage of objects which can be correctly classified into classes $Y_1, ..., Y_n$ employing the set B.

$$\gamma_B(Y) = \frac{\sum_{i=1}^{n} |B_* Y_i|}{|U|} \qquad (1)$$

3. A HYBRID MODEL TO FEATURE SELECTION

FS is an example of a difficult discrete problem. This problem can be modeled using the ant approach in the following way. Let $A=\{a_1, a_2, ..., a_{na}\}$ be a set of features. This set can be viewed as a network in which nodes represent features, and all nodes are connected by bidirectional links. Pheromone values τ_i are associated to nodes a_i, therefore τ_i represents the absolute contribution of the feature a_i to a reduct. In the first step, each ant k is assigned to one node, and b_k is set to $\{a_i\}$, where b_k is the subset ant k builds up. Ants perform a forward selection in which each ant k expands its subset b_k step-by-step by adding new features; for performing it, each ant k looks for all features in the set $A-b_k$ and selects the next feature to include in b_k according to the rule (2) given below. This rule is the pseudorandom proportional rule used ACS. It consists of two components, one being the heuristic component which is in our case the quality of the approximation of the classification measure of RST (given by expression (1)), the second one being the amount currently present in the attribute. The following pseudorandom proportional rule uses this value and the pheromones:

$$i = \begin{cases} \mathrm{argmax}\left(\tau_i \cdot \left(\gamma_B^k \cup \{a_i\}(Y)\right)^\beta\right) & if \quad q \leq q0 \\ J & else \end{cases} \qquad (2)$$

Where J is given by the random proportional rule of ACO. Our ACS-RST-FS algorithm looks for reducts these are subsets B for which $\gamma_B(Y) = \gamma_A(Y)$.

Considering the computation effort for finding either lower or upper approximations involved in the heuristic is $O(na*n^2)$, where na and n are number of attributes and examples respectively, we can estimate the complexity of these algorithms as $O(NC*m*na^2*n^2)$, where NC is the number of cycles, m is the number of ants.

4. EXPERIMENTAL RESULTS.

The aim of this experimental study is to evaluate the algorithm, and determining rules of thumb for the setting its parameters.

We studied three methods to calculate initial pheromone, i.e.: random values, 1/na, and $\sigma_{A,D}(a_i)$, where $\sigma_{A,D}(a_i)$ is the relevance of feature a_i according to RST. The first observation is that the

method to calculate the initial amount of pheromone only has a minor effect in the algorithm performance. For this reason, further experiments were done using random values.

The following rules of thumb are proposed to determine the number of ants (m) as a function of the quantity of features (m=f(na)): R1: If na ≤ 19 then m=na; R2: If $20 \leq na \leq 49$ then [If $0.666*na \leq 24$ then m=24 else m=Round[0.66*na]]; and R3: If na≥ 50 then [If $0.5*na \leq 33$ then m=33, else m=Round[0.5*na]].

We have developed this study using databases from UCI Repository. Our research work shows the relation between Beta(β) and q_0 parameters has the strongest impact in the performance of the algorithm. The combination ($\beta=1$, $q_0=0.3$) generates about 2 times more reducts than combination ($\beta=5$, $q_0=0.9$), regardless of the quantity of the ants and the number of cycles. But, in this case the combination ($\beta=5$, $q_0=0.9$) obtained shorter reducts than other combinations.

We compare the ACS-RST-FS algorithm with other three algorithms. First based on a genetic algorithm [6], second is an Estimation of Distribution Algorithm [3], and the last one is based in pattern recognition techniques [5]. ACS-RST-FS algorithm obtained the shortest reducts and it found a more quantity of these.

5. CONCLUSION.

We have presented a study about the possibilities of applying a hybrid model (ACO + rough sets) to features selection. Our approach is based on Ant Colony System. In this algorithm, the combination of parameters (β, q_0) controls a trade off between the number of reducts found and the size of the reducts. We have determined the method to calculate initial value for pheromone has a minor effect in the algorithm performance.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Dorigo, M. and Stutzle, T. Ant Colony Optimization. MIT Press. 2004.

[2] Jensen, R. and Shen. Q. Finding rough set reducts with ant colony optimization. Proceedings of UK Workshop on Computational Intelligence, pp. 15-22. 2003.

[3] Mühlenbein H. The equation for the response to selection and its use for prediction. Evolutionary Computation 5(3), pp. 303-346, 1998.

[4] Pawlak, Z. Rough sets. International Journal of Information & Computer Sciences 11, 341-356, 1982.

[5] Santiesteban, Yovanis; Pons, Aurora. LEX: LEX: a new algorithm to calculate typical testors. Mathematics Sciences Journal, Vol. 21, No. 1, 2003.

[6] Wroblewski, J. Genetic algorithms in decomposition and classification problems. In Polkowski, L. and Skowron, A. (Eds.). Rough sets in Knowledge Discovery 1: Applications, Case Studies and Software Systems. Physica-Verlag, pp. 472-492. 1998