



HAL
open science

ON THE USE OF ANT COLONY IN ASSOCIATIVE CLASSIFICATION

Kifaya Qaddoum, Fadi Thabtah, Cyrille Bertelle

► **To cite this version:**

Kifaya Qaddoum, Fadi Thabtah, Cyrille Bertelle. ON THE USE OF ANT COLONY IN ASSOCIATIVE CLASSIFICATION. ESM'2007, Oct 2007, St Julian, Malta. pp.315-316. hal-00431235

HAL Id: hal-00431235

<https://hal.science/hal-00431235>

Submitted on 11 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ON THE USE OF ANT COLONY IN ASSOCIATIVE CLASSIFICATION

Kifaya Qaddoum & Fadi Thabtah
Philadelphia University
Amman, Jordan
E-mail: sweetkq@hotmail.com,
ffayez@philadelphia.edu.jo

Cyrille Bertelle
LITIS – University of Le Havre
25 rue Philippe Lebon, BP 540,
76058 Le Havre Cedex, France
E-mail: cyrille.bertelle@litislab.eu

KEYWORDS

Ant Colony, associative classification, data mining, classification rules

ABSTRACT

So far, ant algorithms have showed magnificent research interest within the search/optimisation community. In Data Mining search space, Associative Classification [3] (AC) approaches proved to be strongly able to extract more accurate classifiers than traditional classification techniques. In this paper the two are brought together. An investigation of the ant colony algorithms to present a variant of them for associative classification mining. In the proposed algorithm, each class is labeled and assigned to represent the consequent part of the rules within a processor. The ants groups are attached to the main processor to search for the antecedent part of the rules. The ants, depending on the pheromone, and heuristic information, select the values of the attributes according to the importance of an attribute to the class. The proposed algorithm is expected to discover classification rules with magnificent enhancement on accuracy and less redundancy than comparing methods.

INTRODUCTION

Building effective classification systems is one of the main tasks of data mining and machine learning. Past researches have produced many techniques (e.g. Decision Tree [12], Naive-Bayes [10], Support Vector Machines [5], Neural Networks [13], and Statistical approaches [26], [8]). In recent years a new classification approach that integrates classification with association rule mining called associative classification arises, which is considered a promising technique that produces highly accurate classifiers, AC in many experiments proved to give more accurate rules, and less error rate relating to the produced rules, comparing with other leading techniques in Data Mining Classification.

AN EXACT DEFINITION OF ASSOCIATIVE CLASSIFICATION

Classification rule mining and association rule mining are two important data mining techniques [4]. The target of mining is not pre-determined for association rule mining,

while for classification rule mining there is one and only one pre-determined target, i.e., the class.

Associative classification [3] proposed the integration of association rule mining and classification. Association rule is unsupervised learning that describes the co-occurrence among data items in a large amount of collected data [1]. Whereas, associative classification is a supervised task that predicts the class label of test cases.

This section introduces the associative classification technique. A training dataset R is characterized by a schema (A_1, \dots, A_k, c) , where (A_1, \dots, A_k) are k distinct attributes and c is a class attribute. Each tuple in R is a data object, which is associated to a unique identifier called tid . The attributes may have either a categorical or a continuous domain. For categorical attributes, all values in the domain are mapped to consecutive positive integers. For continuous attributes, the value range is discretized into intervals, which are mapped into consecutive positive integers. In this way, all attributes are treated uniformly.

Each data object in R can be described as a collection of pairs (attribute, integer value), plus a class label (a value belonging to the domain of the class attribute c). Each pair (attribute, integer value) is called item in the remaining part of the paper.

In associative classification, classification rules are used to model the most relevant properties characterizing classes of data, and to predict the class label for unknown (unlabeled) data. A classifier is a function from 2^I to C that allows the assignment of a class label to a data object. A classifier, able to predict the class label for data objects with high accuracy, is generated from a collection of transactions (i.e., data objects with a label, also called training dataset). In associative classification, the classifier is generated by selecting the most appropriate set of association rules. A rule $r : X \rightarrow c$ classifies or matches a data object d when $X \subseteq d$. In this case, rule r assigns class label c to data object d .

Several measures have been proposed to quantify the "interestingness" or the quality of an association (and a classification) rule. Frequently used quality indices are support and confidence [1]. In our setting, these measures correspond to the support and confidence of the l-itemset

encoding the classification rule. For an arbitrary classification rule $\mathcal{X}c$, the support $\text{sup}(\mathcal{X}c)$ is the fraction of transactions in D which contain \mathcal{X} and are labeled by class label c . For the rule antecedent \mathcal{X} , the support $\text{sup}(\mathcal{X})$ is the fraction of transactions in D which contain \mathcal{X} .

ANT COLONY ALGORITHM FOR CLASSIFICATION RULES DISCOVERING

Ant colony techniques have been successfully applied to many problems including the Travelling Salesman Problem [8], Quadratic Assignment Problem [8], Job-Shop Scheduling Problems [9], generic constraint satisfaction problems [10], University Course Timetabling Problems [11], cutting and packing [2] graph colouring problems [9] and load balancing in simulation distribution [15]. The standard technique is a constructive one: a colony of ants moves in a solutions space and evolve in collaborative way to find the best solution. Each ant contributes to this collaborative construction by making decisions stochastically, using existing problem constraints and heuristics combined with experience (which is analogous to a substance called pheromone). The colony then reinforces decisions in the construction process according to their successes by adding pheromone, which also decays to mitigate against poorer decisions. Our purpose in this paper is to investigate the ant algorithm technique as a means of constructing effective class labels associated with an attribute value in the training data set rather than just the single most obvious class.

For our purpose, we define a weighted graph $G = (V,E)$ such that V is the set of vertices corresponding to the class labels and E is the set of edge (u,v) represent the transactions between two nodes u and v . The edge weight value corresponds to $w(u,v)$ is $\text{sup}(u)$, the fraction of transactions which contain u . At each step, a ant will choose to move, crossing an edge (u,v) , according to a probabilistic value $p(u,v)$:

$$p(u,v) = (w(u,v))^\alpha (c(u,v))^\beta / (\sum_{w \in \gamma_u} (w(u,w))^\alpha (c(u,w))^\beta) \quad (1)$$

Where γ_u is all the state reachable from u , α and β are arbitrary parameters and $c(u,v)$ is the pheromone rate deposited by the ant colony on the edge (u,v) . When a ant crosses an edge, it deposit a pheromone value on it. After a complete iteration, the whole pheromone rate on each edge (u,v) is computed by

$$c^{t+1}(u,v) = \sum_{k \in A} c^{t+1}(u,v,k) + \rho c^t(u,v) \quad (2)$$

where $c^{t+1}(u,v,k)$ is the pheromone value deposited by the ant k from the set A , at the iteration $t+1$, ρ is an evaporation factor which allow to decrease the previous pheromone rate. This negative feed-back allow to forget some initial and non optimal solution to make emerge new ones in dynamic way.

This distributed computation contribute to find appropriate transactions path for emergent association classification.

REFERENCES

- [1] Agrawal, R. and Srikant, R. 1994. "Fast algorithms for mining association rules." VLDB-94, 1994.
- [2] E. Burke and G. Kendall, Applying Ant Algorithms and the No Fit Polygon to the Nesting Problem, Proceedings of 12th Australian Joint Conference on Artificial Intelligence, Sydney, Australia, December 6-10, 1999, Lecture Notes in Artificial Intelligence vol. 1747, Foo, N. (Ed), pp. 453-464, 1999.
- [3] Baralis, E. And Garza, P. 2002. A lazy approach to pruning classification rules. In Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM'02). IEEE Computer Society Press, Los Alamitos, CA, 35-42.
- [4] .B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rule mining. In Proc. of 4th Intl. Conf. on Knowledge Discovery and Data Mining (KDD), Aug. 1998.
- [5] Cristianini, N. And Shawe-Taylor, J. 2000. An Introduction to Support Vector Machines. Cambridge University Press, Cambridge, U.K.
- [6] D. Costa and A. Hertz, Ants can colour graphs, Journal of the Operations Research Society 48, pp. 295-305, 1997.
- [7] Duda, R. And Hart, P. 1973. Pattern Classification and Scene Analysis. John Wiley and Sons, New York, NY.
- [8] M. Dorigo, V. Maniezzo and A. Colorni, The Ant System: Optimization by a colony of cooperating agents, IEEE Transactions on Systems, Man and Cybernetics Part B, vol. 26, no. 1, 1996, pp. 1-13.
- [9] M. Dorigo, V. Maniezzo and A. Colorni. Distributed optimization by ant colonies Proceedings of the first European Conference on Artificial Life Paris, ed. By FJ Varela and P Bourguine, MIT/Press/Bradford Books, Cambridge, Massachusetts: pp. 134-142, 1992.
- [10] S. Pimont and C. Solnon, A Generic Ant Algorithm for Solving Constraint Satisfaction Problems, Abstract Proceedings of ANTS' 2000 From Ant Colonies to Artificial Ants: Second International Workshop on Ant Algorithms, pp. 100-108, 2000.
- [11] K. Socha, J. Knowles and M. Sampels, A MAX-MIN Ant System for the University Course Timetabling Problem, Proceedings of the Third International Workshop on Ant Algorithms (ANTS'02), Springer LNCS vol. 2463, pp. 1-13, 2002.
- [12] Quinlan. J. C4.5: Programs for machine learning. Morgan Kaufmann, 1993.
- [13] R. Lippmann. An introduction to computing with neural nets. IEEE ASSP Magazine, 4(22), 1987.
- [14] Thabtah, F, Cowling, P and Peng, Y, 2004, MMAC: A new multi-class, multi-label associative classification approach. In Proceedings of the 4th IEEE International Conference on Data Mining (ICDM'04), Brighton, UK, pp. 217-224.
- [15] Bertelle C, Dutot A, Guinand F and Olivier D, 2007, "Organization detection for dynamic load balancing in individual-based simulation", Multiagent and Grid System, pp 141-163, Volume 3, Number 1, 2007.