# Tackling Ant Colony Optimization Meta-Heuristic as Search Method in Feature Subset Selection Based on Correlation or Consistency Measures

Antonio J. Tallón-Ballesteros and José C. Riquelme

Department of Languages and Computer Systems,
University of Seville, Spain
atallon@us.es

**Abstract.** This paper introduces the use of an ant colony optimization (ACO) algorithm, called Ant System, as a search method in two well-known feature subset selection methods based on correlation or consistency measures such as CFS (Correlation-based Feature Selection) and CNS (Consistency-based Feature Selection). ACO guides the search using a heuristic evaluator. Empirical results on twelve real-world classification problems are reported. Statistical tests have revealed that InfoGain is a very suitable heuristic for CFS or CNS feature subset selection methods with ACO acting as search method. The use of InfoGain is shown to be the significantly better heuristic over a range of classifiers. The results achieved by means of ACO-based feature subset selection with the suitable heuristic evaluator are better for most of the problems comparing with those obtained with CFS or CNS combined with Best First search.

**Keywords:** Feature selection, classification, ant colony optimization, heuristic evaluator, filter, feature subset selection.

## 1 Introduction

Ant colony optimization (ACO) meta-heuristic [6] was proposed by Dorigo et al. and is inspired in the behaviour of real ant colonies. Depending on the amount of pheromone deposited by the ant in their walk there would be some points that are more likely to be visited by the next ants [4]. The (artificial) ants in ACO define a randomized construction heuristic which makes probabilistic decisions depending on the strength of artificial pheromone trails and available heuristic information. As such, ACO can be interpreted as an extension of traditional construction heuristics, which are readily available for many combinatorial optimization problems. Yet, an important difference with construction heuristics is the adaptation of the pheromone trails during algorithm execution to take into account the cumulated search experience. As construction algorithms work on partial solutions trying to extend these in the best possible way to complete problem solutions.

In essence, feature selection (FS) is a NP-hard combinatorial optimization problem. Often, FS is tackled with classical search methods for avoiding the prohibitive exhaustive search. Instead of looking for the optimal solution, obtaining a good solution in a reasonable time might be preferable for certain problems. On one hand, the typical NP-hard Travelling Salesman Problem has been treated successfully with ACO. On the other hand, ACO was applied for feature selection in [9] in the context of rough set reducts with very promising results. Here, we apply ACO as a stochastic procedure for quickly finding high quality solutions, in the scope of ordinary or crisp sets for FS in classification tasks. In this context, feature subset selection (FSS) problem is formulated using a graph with the purpose of getting a subset of attributes that is relevant for the problem at hand. FSS needs a search method, that usually is any kind of artificial intelligence heuristic technique. The current proposal shifts the search from an heuristic non-stochastic perspective to a stochastic angle.

This paper goals to address the suitability of using the ant colony optimization meta-heuristic as a search method built-in in the CFS and CNS feature subset selectors for classification problems.

The rest of this article is organized as follows: Sect. 2 describes some concepts about ACO meta-heuristic and feature selection; Sect. 3 presents our proposal; Sect. 4 details the experimentation; then Sect. 5 shows and analyzes statistically the results obtained; finally, Sect. 6 states the concluding remarks.

## 2 Background

### 2.1 Ant Colony Optimization Meta-Heuristic

Artificial ants used in ACO are stochastic solution construction procedures that probabilistically build a solution by iteratively adding solution components to partial solutions by taking into account a) heuristic information about the problem instance being solved, if available, and b) (artificial) pheromone trails which change dynamically at runtime to reflect the agents' acquired search experience.

The problem representation is a graph where the nodes represent the different points that the ants can visit and the edges are the link between points. Links are unidirectional and there are no cycles, so it is not possible to go back to a point previously visited. At the beginning of the algorithm every ant is located in a point and will construct a solution taking several decisions until the stop condition is met. At the end each ant has found a candidate solution to the problem at hand.

The main steps of the algorithm are the following:

1. *Initialisation.* The algorithm starts and all the pheromone variables are initialized to a value $\tau_0$ which is a key parameter.
2. *Construct Ant Solutions.* This action starts the algorithm loop and is related with sending ants around the construction graph. An ant in the node $i$ chooses the $j$ one according to a probabilistic decision rule, which is a function of the pheromone $\tau$ and the heuristic $\eta$. A set of $n_a$ ants constructs solutions to the concrete problem being tackled.

3. *Update Pheromones*. The purpose of this part is to change the values of the pheromones, by both depositing and evaporating.

Among the several variants of ACO, hereinafter we focus on Ant System (AS) that was introduced in 1991 by Dorigo and published a few years later by Dorigo et al. [5].

### 2.2 Feature Selection

Feature selection methods try to pick a subset of features that are relevant to the target concept [2]. According to Langley [10], the different approaches for feature selection can be divided into two broad categories (i.e., filter and wrapper) based on their dependence on the inductive algorithm that will finally use the selected subset. Filter methods are independent of the inductive algorithm, whereas wrapper methods use the inductive algorithm as the evaluation function.

FS involves two stages: a) to obtain a list of attributes according to an attribute evaluator and b) to perform a search on the initial list. All candidate lists would be assessed using a measure evaluation and the best one will be returned.

Two of the most widespread feature subset selectors are Correlation-based Feature Selection (CFS) [8] and Consistency-based feature selection (CNS) [3] that work in combination with a search method such as Greedy Search, Best First (BF) or Exhaustive Search. Generally speaking, BF is a powerful search method [7] which is the reason to be used very frequently by the machine learning community nowadays. We have chosen CFS and CNS as representative FSS methods, because they are based on different kind of measures, have few parameters and have provided a good performance inside the supervised machine learning area. Often, BF search is the preferable option by the researchers for both FSS algorithms. CFS is probably the most used FSS in data mining. CNS is also powerful, however the amount of published works is more reduced.

## 3 Proposal

Firstly, the graph meaning may be reformulated in order to deal with a feature selection problem by means of ACO meta-heuristic. Here, the nodes represent features and edges the link between nodes and the possibility to add another feature to the current solution. The search for a candidate solution is a walk through the graph. Once an ant visits an edge it contains a weight indicating the strength of this solution component.

In the current work, ACO, implemented following the AS model, is considered as search strategy in the context of CFS and CNS methods after the attribute evaluation phase. ACO guides the search by means of a heuristic evaluator. As heuristic evaluators, on one hand we have considered for CFS and CNS approaches the own attribute evaluator, obtaining the pure versions for CFS-AS and CNS-AS. On the other hand, we have tried Information Gain (InfoGain as abbreviation) [1] as evaluator resulting an hybrid approach. Moreover, CFS,

CNS and InfoGain compute different kinds of measure to evaluate the relevance, such as correlation, consistency and information, respectively.

The probabilistic transition rule is defined in the same way as in [9] and is the most widely used in AS [5]:

$$p_{ij} = \frac{\tau_{ij}^{\alpha} \cdot [\eta_{ij}]^{\beta}}{\sum_{il \in \mathcal{N}(x)} \tau_{il}^{\alpha} \cdot [\eta_{il}]^{\beta}}, \quad \forall ij \in \mathcal{N}(x). \tag{1}$$

where $p_{ij}$ represents the probability that current ant at feature $i$ would travel to feature $j$, $\tau_{ij}$ is the amount of pheromone on the $ij$ edge, $\eta_{ij}$ is the heuristic desirability of the $ij$ transition and $\mathcal{N}(x)$ the set of current feasible components. Lastly, $\alpha$ and $\beta$ are parameters that may take real positive values –according to the recommendations on parameter setting in [5]– and are associated with heuristic information and pheromone trails, respectively.

All ants update pheromone level with an increase of small quantities, depending directly on the heuristic desirability of the $ij$ transition given by the measure (merit) of the subset attribute evaluator used as heuristic evaluator and inversely proportional to the subset size.

## 4 Experimentation

Table 1 depicts the feature subset selection methods applied in the experimental process. We have grouped them according to the attribute evaluator and search method.

Table 2 summarizes the main parameters along with their symbols and numerical or conceptual values for all the feature subset selection methods used for the experiments. On one hand, in relation to ACO-based feature subset selection, the $n_a$ and $gen$ parameters have been set to fix values to our choice. For $\tau_0$ parameter, in [5] there is a suggestion to assign a small positive constant and hence we have defined a value of 0.5. The trade-off between $\alpha$ and $\beta$ parameters may influence in the behaviour of the algorithm thus for their determination a preliminary experimental design by means of a five-fold cross validation on the training set has been carried out with a couple of values for each one parameter (1 and 2). On the other hand, for BF-based search method in the context of CFS and CNS, we have followed the recommendations of the authors ([8] and [3] for the number of expanded nodes; the search direction has been fix according to our previous experiences).

Table 3 represents the data sets employed throughout the experimentation. They come mostly from binary and multi-class classification real-world problems (*Cl.* column specifies the number of classes) taken from the public UCI repository. The number of instances ranges from more than one hundred to approximately fourteen thousands, thus problems with a medium size, and the dimensionality varies between nineteen and one hundred and twenty nine. Also, we have included the number of selected features for every feature subset selector obtained in the training set. Last row shows the dimensionality reduction (higher is better) in mean over the original data sets for each filter.

**Table 1.** List of feature subset selectors for the experimentation

| Attribute evaluator | Search method | Heuristic evaluator | *Abb. Name* |
|---|---|---|---|
| $CFS$ | $AntSearch$ | $CFS$ | $CFS-AS\ h1$ |
| | | $InfoGain$ | $CFS-AS\ h2$ |
| | $BestFirst$ | $-$ | $CFS-BF$ |
| $CNS$ | $AntSearch$ | $CNS$ | $CNS-AS\ h1$ |
| | | $InfoGain$ | $CNS-AS\ h2$ |
| | $BestFirst$ | $-$ | $CNS-BF$ |

**Table 2.** Parameter values for ACO-based feature subset selection approaches (CFS-AS and CNS-AS) and BestFirst-based ones (CFS-BF and CNS-BF)

| *Search method* | Parameter | Symbol | Value |
|---|---|---|---|
| *Ant search* | $Number\ of\ ants$ | $n_a$ | 10 |
| | $Number\ of\ generations$ | gen | 10 |
| | $Pheromone\ trail\ influence$ | $\alpha$ | 1 |
| | $Heuristic\ informacion\ value$ | $\beta$ | 2 |
| | $Pheromone\ initial\ value$ | $\tau_0$ | 0.5 |
| *Best First* | *Consecutive expanded nodes without improving* | | 5 |
| | *Search direction* | | $Forward$ |

**Table 3.** Summary of the data sets used and selected features for each feature subset selector

| *Data set* | *Size* | *Train* | *Test* | *Feat.* | *Cl.* | *Selected features* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | *CFS* | | | *CNS* | | |
| | | | | | | *AS* | | *BF* | *AS* | | *BF* |
| | | | | | | $h1$ | $h2$ | | $h1$ | $h2$ | |
| *batch(gas)* | 13910 | 10432 | 3478 | 129 | 6 | 7 | 13 | 20 | 7 | 10 | 6 |
| *cardiotoc.* | 2126 | 1595 | 531 | 22 | 10 | 12 | 12 | 12 | 15 | 15 | 13 |
| *hepatitis* | 155 | 117 | 38 | 19 | 2 | 10 | 9 | 10 | 15 | 11 | 11 |
| *ionosphere* | 351 | 263 | 88 | 33 | 2 | 9 | 10 | 11 | 9 | 9 | 9 |
| *libras* | 360 | 270 | 90 | 90 | 15 | 8 | 34 | 23 | 47 | 55 | 20 |
| *lymph.* | 148 | 111 | 37 | 38 | 4 | 11 | 8 | 12 | 19 | 13 | 10 |
| *promoter* | 106 | 80 | 26 | 58 | 2 | 10 | 10 | 10 | 15 | 8 | 8 |
| *satimage* | 6435 | 4435 | 2000 | 36 | 6 | 20 | 21 | 23 | 13 | 13 | 12 |
| *sonar* | 208 | 104 | 104 | 60 | 2 | 4 | 7 | 8 | 14 | 9 | 9 |
| *soybean* | 683 | 511 | 172 | 82 | 19 | 22 | 22 | 25 | 39 | 58 | 16 |
| *SPECTF* | 267 | 80 | 187 | 44 | 2 | 10 | 8 | 12 | 12 | 10 | 8 |
| *waveform* | 5000 | 3750 | 1250 | 40 | 3 | 12 | 14 | 14 | 14 | 13 | 12 |
| *Averages* | 2479.08 | 1812.33 | 666.75 | 54.25 | 6.08 | 11.25 | 14.00 | 15.00 | 18.25 | 18.67 | 11.17 |
| *Dim. reduction* | | | | | | 79.26 | 74.19 | 72.35 | 66.36 | 65.59 | 79.42 |

The experimental design follows a stratified hold-out cross validation with three and one quarters for the training and test sets, respectively. Sometimes, these proportions do not match since the original data are prearranged. For the statistical analysis between two feature subset selection methods we have carried a Wilcoxon signed-ranks test with the test accuracy results obtained.

## 5 Results

Tables 4 and 5 report the accuracy test results for the FSS based on ACO with CFS and CNS -that is, CFS-AS and CNS-AS, respectively- with two different heuristic evaluators, their difference ($Diff.$) and its ranking ($R.$). Ten executions with different seeds were run and the most frequent solution was considered for the assessment. We have carried out experiments with three kind of deterministic classifiers such: a) C4.5, based on decision trees, b) SVM, founded in support vectors, and c) PART, a rule-based approach. The reason for the choice of these classifiers is motivated by the fact that their overall performance is good in the feature selection scope [11]. The best results, excluding ties, for each pair (classifier, FSS method) have been highlighted with boldface. According to Wilcoxon signed-ranks test, since there are 12 data sets, the $T$ value at $\alpha = 0.05$ should be less or equal than 14 (the critical value) to reject the null hypothesis. On one hand, in relation to CFS-AS, for classifiers C4.5 and SVM the h2 heuristic evaluator is significantly better than h1. On the other hand, for CNS-AS the performance of h2 with PART is the significant best option.

**Table 4.** CFS-AS: Accuracy test results and statistical tests

| *Data set* | *C*4.5 | | | | SVM | | | | PART | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $CFS-AS$ | | | | $CFS-AS$ | | | | $CFS-AS$ | | | |
| | h1 | h2 | $Diff.$ | R. | h1 | h2 | $Diff.$ | R. | h1 | h2 | $Diff.$ | R. |
| *batch(gas)* | 93.99 | **96.87** | 2.88 | 8 | 72.28 | **78.21** | 5.92 | 9 | 93.73 | **96.35** | 2.62 | 7 |
| *cardiotoc.* | 61.58 | 61.58 | 0.00 | 2 | **67.98** | 67.80 | −0.19 | 2 | 60.45 | **62.52** | 2.07 | 5 |
| *hepatitis* | 84.21 | **89.47** | 5.26 | 10 | 86.84 | **89.47** | 2.63 | 6 | 84.21 | **86.84** | 2.63 | 8 |
| *ionosphere* | **90.91** | 87.50 | −3.41 | 9 | 82.95 | **89.77** | 6.82 | 10 | 90.91 | **93.18** | 2.27 | 6 |
| *libras* | 52.22 | **65.56** | 13.33 | 12 | 50.00 | **63.33** | 13.33 | 12 | 54.44 | **65.56** | 11.11 | 12 |
| *lymph.* | 81.08 | 81.08 | 0.00 | 2 | 81.08 | **83.78** | 2.70 | 8 | **72.97** | 70.27 | −2.70 | 9 |
| *promoter* | 73.08 | 73.08 | 0.00 | 2 | 73.08 | 73.08 | 0.00 | 1 | 80.77 | 80.77 | 0.00 | 1 |
| *satimage* | 86.05 | **86.25** | 0.20 | 5 | 83.80 | **84.50** | 0.70 | 3 | **85.00** | 83.55 | −1.45 | 2 |
| *sonar* | 67.31 | **74.04** | 6.73 | 11 | 67.31 | **75.00** | 7.69 | 11 | 68.27 | **75.96** | 7.69 | 11 |
| *soybean* | 88.95 | **91.28** | 2.33 | 6 | 94.19 | **95.35** | 1.16 | 5 | 90.70 | **92.44** | 1.74 | 3 |
| *SPECTF* | 66.84 | **69.52** | 2.67 | 7 | **66.31** | 63.64 | −2.67 | 7 | 70.59 | **76.47** | 5.88 | 10 |
| *waveform* | 74.32 | **74.40** | 0.08 | 4 | 85.92 | **86.88** | 0.96 | 4 | **78.88** | 77.04 | −1.84 | 4 |
| | $T = min\{66, 12\} = 12$ | | | | $T = min\{68.5, 9.5\} = 9.5$ | | | | $T = min\{62.5, 15.5\} = 15.5$ | | | |

Table 6 outlines the global accuracy test results of the best CFS-AS and CNS-AS approaches versus based-BF CFS or CNS. The best results, in each data set, are marked in bold. We can assert the following statements in relation to the

**Table 5.** CNS-AS: Accuracy test results and statistical tests

| *Data set* | *C*4.5 | | | | SVM | | | | PART | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $CNS-AS$ | | | | $CNS-AS$ | | | | $CNS-AS$ | | | |
| | h1 | h2 | $Diff.$ | R. | h1 | h2 | $Diff.$ | R. | h1 | h2 | $Diff.$ | R. |
| *batch(gas)* | **96.00** | 95.54 | −0.46 | 3 | 64.49 | **69.15** | 4.66 | 10 | 95.46 | 95.46 | 0.00 | 1.5 |
| *cardiotoc.* | 63.84 | **66.48** | 2.64 | 6 | 61.39 | **64.41** | 3.01 | 9 | 58.76 | **64.22** | 5.46 | 8 |
| *hepatitis* | 84.21 | **86.84** | 2.63 | 5 | **86.84** | 84.21 | −2.63 | 8 | 76.32 | **84.21** | 7.89 | 10 |
| *ionosphere* | 88.64 | **93.18** | 4.55 | 12 | 81.82 | **84.09** | 2.27 | 7 | 89.77 | **95.45** | 5.68 | 9 |
| *libras* | 56.67 | **61.11** | 4.44 | 11 | 67.78 | **70.00** | 2.22 | 6 | 55.56 | **64.44** | 8.89 | 11 |
| *lymph.* | 78.38 | **81.08** | 2.70 | 8 | **91.89** | 83.78 | −8.11 | 11 | **78.38** | 64.86 | −13.51 | 12 |
| *promoter* | **84.62** | 80.77 | −3.85 | 10 | 73.08 | **84.62** | 11.54 | 12 | 76.92 | **80.77** | 3.85 | 5 |
| *satimage* | 84.70 | **84.75** | 0.05 | 2 | **84.50** | 83.70 | −0.80 | 4 | 85.50 | **86.25** | 0.75 | 4 |
| *sonar* | **75.96** | 73.08 | −2.88 | 9 | **75.96** | 75.00 | −0.96 | 5 | 72.12 | **75.96** | 3.85 | 6 |
| *soybean* | 91.86 | 91.86 | 0.00 | 1 | **95.35** | 94.77 | −0.58 | 3 | 92.44 | 92.44 | 0.00 | 1.5 |
| *SPECTF* | **69.52** | 66.84 | −2.67 | 7 | 64.71 | 64.71 | 0.00 | 1 | 65.24 | **69.52** | 4.28 | 7 |
| *waveform* | **76.00** | 74.88 | −1.12 | 4 | 85.12 | **85.36** | 0.24 | 2 | 77.36 | **77.84** | 0.48 | 3 |
| | $T = min\{44.5, 33.5\} = 33.5$ | | | | $T = min\{51.5, 26.5\} = 26.5$ | | | | $T = min\{64.5, 13.5\} = 13.5$ | | | |

**Table 6.** Global accuracy test results of CFS-BF and CNS-BF versus CFS-AS and CNS-AS

| *Data set* | *C*4.5 | | SVM | | PART | |
|---|---|---|---|---|---|---|
| | $CFS-BF$ | $CFS-AS$ | $CFS-BF$ | $CFS-AS$ | $CNS-BF$ | $CNS-AS$ |
| | | h2 | | h2 | | h2 |
| *batch(gas sensor)* | 95.92 | 96.87 | 83.04 | 78.21 | **98.30** | 95.46 |
| *cardiotoc.* | 61.58 | 61.58 | **67.80** | **67.80** | 61.02 | 64.22 |
| *hepatitis* | 84.21 | **89.47** | 86.84 | **89.47** | 84.21 | 84.21 |
| *ionosphere* | 92.05 | 87.50 | 88.64 | 89.77 | 88.64 | **95.45** |
| *libras* | 61.11 | **65.56** | 57.78 | 63.33 | 55.56 | 64.44 |
| *lymph.* | 81.08 | 81.08 | 81.08 | **83.78** | 67.57 | 64.86 |
| *promoter* | 73.08 | 73.08 | 73.08 | 73.08 | 69.23 | **80.77** |
| *satimage* | 85.60 | **86.25** | 83.85 | 84.50 | 85.45 | **86.25** |
| *sonar* | 73.08 | 74.04 | 75.00 | 75.00 | **75.96** | **75.96** |
| *soybean* | 93.02 | 91.28 | 94.77 | **95.35** | 93.02 | 92.44 |
| *SPECTF* | 66.84 | 69.52 | **73.26** | 63.64 | 66.31 | 69.52 |
| *waveform* | 74.40 | 74.40 | **86.88** | **86.88** | 76.16 | 77.84 |
| *Wins by pairs* | 2 | 6 | 2 | 6 | 3 | 7 |
| *Global wins* | 0 | 3 | 3 | 5 | 2 | 4 |
| *Averages* | | | | | | |
| *Accuracy* | 78.50 | 79.22 | 79.33 | 79.23 | 76.79 | 79.29 |
| *Selected feat.*(%) | 27.65 | 25.81 | 27.65 | 25.81 | 20.58 | 34.41 |

achieved results. First, the comparison between pairs of FSS methods for each classifier and data sets points out that: a) C4.5 with CFS-AS gets better results 6 times, b) SVM with CFS-AS wins in 6 problems, and c) PART classifier with CNS-AS 7 times. Second, a global analysis from a qualitative point of view means that (SVM, CFS-AS) pair reaches the best results 5 times, followed by (PART, CNS-AS) with 4 wins. Third, the percentage of selected attributes in (SVM, CFS-AS) pair is close to 25, while with (PART, CNS-AS) is slightly greater and takes a value near 35.

## 6 Conclusions

CFS-AS and CNS-AS were presented. Experiments revealed that ACO-based search via AS in feature subset selection with InfoGain heuristic is better, and in some cases with significant differences, than the pure versions of ACO-based filters (that is, a concrete subset attribute evaluator with the homonymous heuristic evaluator, e.g. CFS-AS with h1). It is very important to stress that ACO-based feature subset selector with the proper heuristic evaluator is better in more than the half of the problems that the traditional Best First search in CFS or CNS. The two preferred classifier-FSS pairs, bearing in mind the performance regarding the accuracy and number of selected attributes are, in this order, (SVM, CFS-AS) and (PART, CNS-AS).

**Acknowledgments.** This work has been partially subsidized by TIN2007-68084-C02-02 and TIN2011-28956-C02-02 projects of the Spanish Inter-Ministerial Commission of Science and Technology (MICYT), FEDER funds and P11-TIC-7528 project of the "Junta de Andalucía" (Spain).

## References

1. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley (1991)
2. Dash, M., Liu, H.: Feature selection for classification. Intelligent Data Analysis 1(3), 131–156 (1997)
3. Dash, M., Liu, H.: Consistency-based search in feature selection. Artificial Intelligence 151(1), 155–176 (2003)
4. Dorigo, M., Di Caro, G., Gambardella, L.M.: Ant algorithms for discrete optimization. Artificial Life 5(2), 137–172 (1999)
5. Dorigo, M., Maniezzo, V., Colorni, A.: Ant system: optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 26(1), 29–41 (1996)
6. Dorigo, M., Stützle, T.: Ant colony optimization: overview and recent advances. In: Handbook of Metaheuristics, pp. 227–263. Springer (2010)
7. Gaschig, J.: Performance measurement and analysis of certain search algorithms. Technical report, DTIC Document (1979)
8. Hall, M.A.: Correlation-based feature selection for machine learning. PhD thesis, The University of Waikato (1999)
9. Jensen, R., Shen, Q.: Finding rough set reducts with ant colony optimization. In: Proceedings of the 2003 UK Workshop on Computational Intelligence, vol. 1 (2003)
10. Langley, P.: Selection of relevant features in machine learning. Defense Technical Information Center (1994)
11. Tallón-Ballesteros, A.J., Hervás-Martínez, C., Riquelme, J.C., Ruiz, R.: Feature selection to enhance a two-stage evolutionary algorithm in product unit neural networks for complex classification problems. Neurocomputing 114, 107–117 (2013)