



## An Ant Colony Optimization Based Approach for Feature Selection

Ahmed Al-Ani

Faculty of Engineering, University of Technology, Sydney  
PO Box 123, Broadway, NSW 2007, Australia  
ahmed@end.uts.edu.au

### Abstract

This paper presents a new feature subset selection algorithm based on the Ant Colony Optimization (ACO). ACO is a metaheuristic inspired by the behaviour of real ants in their search for the shortest paths to food sources. It looks for optimal solutions by utilizing distributed computing, local heuristics and previous knowledge. We modified the ACO algorithm so that it can be used to search for the best subsets of features. A new pheromone trail update formula is presented, and the various parameters that lead to better convergence are tested. Results on speech classification problem show that the proposed algorithm achieves better performance than both greedy and genetic algorithm based feature selection methods.

**Keywords:** Feature selection, Ant colony optimization, Ant system, Pattern recognition.

### 1. Introduction

Pattern recognition is an important and multi-disciplinary field of research with wide range of applications that include handwriting recognition, speech recognition, medical diagnosis, fingerprint verification and face recognition. Among the several parameters that affect the performance of pattern recognition systems, feature representation of patterns can be considered the most important. In some applications, it might be sufficient to use simple features that are previously known. However, in other applications, unique feature sets that are necessary and sufficient to the classification task do not exist. Moreover, the assumption that more features can offer complementary information about the patterns to be classified is not always valid. It has been found that including more features will make the classification and analysis more difficult, time consuming and may even lead to poorer generalization on unseen data. This makes feature selection and reduction in feature set dimensionality very desirable.

Ideally, the best subset of features can be found by evaluating all the possible subsets, which is known as

exhaustive search. If we have a set of  $N$  features, then there will be  $2^N - 2$  candidate subsets. This procedure may be practically impossible even for a moderate-size feature set, e.g. for  $N = 20$ , there will be 1,048,574 subsets. On the other hand, examining features individually may not be sufficient, as it is important to take into consideration the interaction between features [1]. A more practical approach would be computationally feasible and aims at achieving optimal or “semi-optimal” solutions. Hence, several search procedure methods have been developed, which basically differ in their computational cost and the optimality of the subsets they find. In addition to the search procedure, a feature subset evaluation measure is needed to evaluate the importance of subsets. The existing evaluation measures can be broadly divided into two main groups: filters and wrappers. Filters operate independently of any learning algorithm, where undesirable features are filtered out of the data before learning begins. On the other hand, performance of classification algorithms is used to select features for wrapper methods [2].

In recent years, population-based optimization algorithms have attracted a lot of attention. Such methods attempt to achieve better solutions by utilizing knowledge from previous iterations. One of the promising population-based algorithms is the Ant Colony Optimization (ACO) [3]. The ant algorithm was inspired by the real ants’ behaviour in their search of food, and targets discrete optimization problems. The coordination of a population of ants takes place through indirect communication, which is mediated by laying an odorous substance on food paths. This will increase the probability that other ants will follow those specific marked paths.

In this paper, we propose a new feature selection algorithm that searches the feature space using a modified ACO algorithm. In the next section, we give a brief description of some of the well-known feature subset search algorithms. Section three explains the ACO metaheuristic. The proposed feature subset search algorithm is presented in section four. Experimental

results are presented in section five and a conclusion is given in section six.

## 2. Available Feature Subset Search Algorithms

Due to the importance of feature selection, many feature subset search algorithms have been proposed in the literature. Two of the most famous approaches are stepwise search and Genetic Algorithms (GA).

The stepwise (or greedy) search adds/removes a single feature to/from the current subset [4]. It considers local changes to the current feature subset. Often, a local change is simply the addition or deletion of a single feature from the subset. The stepwise, which is also called the Sequential Forward Selection (SFS)/Sequential Backward Selection (SBS) is probably the simplest search procedure and is generally sub-optimal and suffers from the so-called “nesting effect”. It means that the features that were once selected/deleted cannot be later discarded/re-selected. To overcome this problem, Pudil et al. [5] proposed a method to flexibly add and remove features, which they called “floating search”.

Another famous search approach is based on the Genetic Algorithm (GA). The GA is a combinatorial search technique based on both random and probabilistic measures. Subsets of features are evaluated using a fitness function and then combined via cross-over and mutation operators to produce the next generation of subsets [6]. The GA employ a population of competing solutions, evolved over time, to converge to an optimal solution. Effectively, the solution space is searched in parallel, which helps in avoiding local optima. A GA-based feature selection solution would typically be a fixed length binary string representing a feature subset, where the value of each position in the string represents the presence or absence of a particular feature. According to [7, 8], the GA was able to achieve better performance than other conventional methods.

We propose in this paper an ant system approach for feature subset selection that aims at achieving similar or better results than GA-based feature selection.

## 3. Ant Colony Optimization

There has been an increasing interest in studying the behaviour of animals and insects, and in particular how they interact to achieve their goal. Among the various animals and insects, ants have attracted a lot of attention.

Scientists have found that an odorous substance, known as pheromone, is used as an indirect communication medium between ants. When a source of food is found, ants lay some pheromone to mark the path between the nest and the food source. The quantity of the laid pheromone depends upon the distance, quantity and quality of the food source. While an isolated ant that moves at random detects a laid pheromone, it is very likely that it will decide to follow its path. This ant will

itself lay a certain amount of pheromone, and hence enforce the pheromone trail of that specific path. Accordingly, the path that has been used by more ants will be more attractive to follow. This process is hence characterized by a positive feedback loop [3].

A number of experiments have been conducted by Deneubourg et al. [9] to study the behaviour of ants when they are forced to choose between paths that vary in their lengths. In one experiment a bridge of two branches was placed between the ants’ nest and the food source, where one branch was twice as long as the other branch (see Figure. 1.a). Initially, when ants arrived at the bridge, they randomly chose between the two branches. However, it is obvious that ants that chose the short branch would reach the food and start their journey back to the nest faster than those that chose the long branch. Accordingly, pheromone would start to accumulate faster on the short branch and hence influence more ants to choose the short branch. After a certain period of time, the vast majority of the ants chose the short branch.

In another experiment, only the long branch was offered and after 30 minutes, the short branch was added (see Figure 1.b). Because ants were only using the long branch for the first 30 minutes, a large quantity of pheromone was laid on this branch. Hence, even after adding the short branch, ants were still choosing the long branch, because they were influenced by the high pheromone concentration.

It is important to mention that if ants stop laying pheromone on a certain path, then the pheromone intensity on that path will decrease over time. This process favours exploration of new paths, and is known as pheromone evaporation. Note that in the second experiment, the slow pheromone evaporation rate could not allow ants to disregard the long path and search for a more optimal and shorter path.

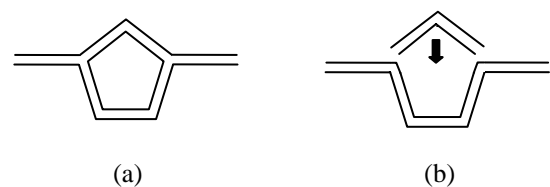


Figure 1. Two bridge experiments aim at measuring the usage of short and long paths

Dorigo et. al. [10] adopted the concept of ants’ foraging behaviour and proposed an artificial colony of ants algorithm. The algorithm was called the Ant Colony Optimization (ACO) metaheuristic, and aimed at solving hard combinatorial optimization problems. The ACO was originally applied to solve the classical traveling salesman problem [3], where it was shown to be an effective tool in finding good solutions. The ACO has also been successfully applied to other optimization problems including data mining and telecommunications networks [11, 12].







