



# Integration of graph clustering with ant colony optimization for feature selection



Parham Moradi\*, Mehrdad Rostami

Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran

## ARTICLE INFO

### Article history:

Received 18 August 2014

Received in revised form 16 January 2015

Accepted 6 April 2015

Available online 9 April 2015

### Keywords:

Feature selection

Ant colony optimization

Filter method

Graph clustering

## ABSTRACT

Feature selection is an important preprocessing step in machine learning and pattern recognition. The ultimate goal of feature selection is to select a feature subset from the original feature set to increase the performance of learning algorithms. In this paper a novel feature selection method based on the graph clustering approach and ant colony optimization is proposed for classification problems. The proposed method's algorithm works in three steps. In the first step, the entire feature set is represented as a graph. In the second step, the features are divided into several clusters using a community detection algorithm and finally in the third step, a novel search strategy based on the ant colony optimization is developed to select the final subset of features. Moreover the selected subset of each ant is evaluated using a supervised filter based method called novel separability index. Thus the proposed method does not need any learning model and can be classified as a filter based feature selection method. The proposed method integrates the community detection algorithm with a modified ant colony based search process for the feature selection problem. Furthermore, the sizes of the constructed subsets of each ant and also size of the final feature subset are determined automatically. The performance of the proposed method has been compared to those of the state-of-the-art filter and wrapper based feature selection methods on ten benchmark classification problems. The results show that our method has produced consistently better classification accuracies.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

In recent years, with the advance of science and technology, the amount of data has been growing rapidly and thus pattern recognition methods often deal with samples consisting of thousands of features. This problem is called curse of dimensionality and reduction of the datasets' dimensionality becomes crucial to make them tractable [7,45,61]. Dimensionality reduction methods provide a way of understanding the data better, improving prediction performance and reducing the computation time in pattern recognition applications. As a general rule, for a classification problem with  $D$  dimensions and  $C$  classes, a minimum of  $10 \times D \times C$  training samples are required [9]. While it is practically impossible to acquire the required number of training samples, reducing features reduces the size of the training sample required and consequently helps to improve the overall performance of the classification algorithm.

\* Corresponding author. Tel.: +98 8733668513.

E-mail addresses: [p.moradi@uok.ac.ir](mailto:p.moradi@uok.ac.ir) (P. Moradi), [me.rostami@gmail.com](mailto:me.rostami@gmail.com) (M. Rostami).

A common way to deal with such problems is the feature selection technique. The feature selection methods can be classified into four categories including filter, wrapper, embedded, and hybrid models [51]. The filter approach requires the statistical analysis of the feature set without utilizing any learning algorithm. In contrast, wrapper-based feature selection methods apply a learning algorithm to evaluate the quality of feature subsets in the search space iteratively. In the embedded model the feature selection procedure is considered as a part of the process of building models. On the other hand, the goal of the hybrid-based methods is to use the computational efficiency of the filter model and the proper performance of the wrapper model.

In recent years many evolutionary and swarm-based methods such as Genetic Algorithm (GA) [8,56], ant colony optimization (ACO) [1,11,17,54,60], particle swarm optimization (PSO) [27,58] and Artificial Bee Colony (ABC) [36,44] and Harmony Search algorithm (HSA) [62] have been utilized to tackle the feature selection problem. Among the swarm intelligence-based methods, ACO has been successfully used in the feature selection area of research. The ACO is a metaheuristic algorithm for solving hard combinatorial optimization problems [35]. This algorithm has been successfully applied to a large number of difficult combinatorial

problems such as vehicle routing [47], graph coloring [16], and communication network [68]. ACO is a multi-agent system and it has some advantages such as positive feedback, the use of a distributed long-term memory, nature implementation in a parallel way, functions similar to those of reinforcement learning schemata, and a good global and local search capability due to stochastic and greedy components in the algorithm. Moreover, ACO has been successfully applied for feature selection problem [1,11,12,17,26,31,32,37,52,54,67]. Although, ACO has been shown as an effective approach to finding optimal (or near optimal) feature subsets while it suffers from several shortcomings which are listed as follows:

1. *Graph representation*: In many ACO-based feature selection methods the problem space is represented by a fully connected graph except a work [11] in which the problem space was represented by a directed graph with only  $2n$  arcs where  $n$  denotes the number of features. In the case of fully connected graphs, in each step (i.e., step  $t$ ) each ant should compute the probability rule for unselected features (i.e.,  $n - t + 1$ , where  $n$  denotes the number of features) which leads to increase the time complexity of the algorithm. For example if the ant needs to traverse  $m$  numbers of nodes in the graph,  $(n)!/(n - m)!$  computations are needed; therefore one can reduce these computations.
2. *Updating pheromone*: Most ACO-based feature selection methods employed a learning model in their search process to evaluate a constructed feature subset, and thus they are classified as the wrapper model [1,37,60]. While, the wrapper based methods need high computational time especially on the datasets with a large number of features. On the other hand, only in three cases instead of the learning model, information theoretic measures are used to update the pheromone [32,48,52,54].
3. *Selecting redundant features*: In most of ACO-based feature selection methods, the possible dependency between the features is ignored in the search process [1,11,60]. These methods assume that the features are conditionally independent and thus, while the ant selects the next feature, the dependency of the feature on previously selected ones is ignored in the computations. Therefore, the constructed subset may contain the redundant features, which reduces the classifier performance.
4. *Final subset size*: The number of selected features which defines traverse path length of the ants, imposes another challenge on ACO-based methods. In most of the ACO-based feature selection methods the number of traversed nodes should be pre-determined before the ant starts their search processes [48,52,54]. Moreover the accuracy of these methods depends on optimally defining the size of feature subset.

To overcome the mentioned shortcomings, in this paper we propose a novel filter-based feature selection method based on ACO algorithm. The method attempts to select high-quality features within a reasonable time. The proposed algorithm which is called Graph Clustering based ACO feature selection method, in short GCACO, works in three steps. In the first step, the problem space is represented as a graph in which each node denotes a feature and the edges weights are similarities between features. In the second step, features are divided into several clusters by employing an efficient community detection algorithm [59]. Finally in the third step, a novel ACO-based search strategy is proposed for selecting the final feature subset. In this strategy an ant which is placed on a randomly selected cluster, in each step, decides to select the next position in the current cluster or move to another cluster. In the case of remaining in the same cluster, the probability values are computed only for the features of this cluster. In contrast for the other cases the probability values are computed for

the features of the next selected cluster. This process is continued until all of the clusters are visited. Therefore, the number of features which are selected by each ant in each cycle and also the final feature subset can be automatically determined based on the number of clusters in the problem space.

This approach is quite different from those of the existing schemes [48,52,54], where the size of the constructed subset is defined by a fixed number. Furthermore, the aim of using community-based representations of the problem space is to group highly correlated features into the same cluster. Therefore the ACO-based search process is guided in such a way that relatively less correlated features are injected in a high proportion with respect to more correlated features to the consecutive iteration. Besides, the similarity between features is considered in computation of feature relevance, which minimizes the redundancy between selected features. Therefore, the clustering-based strategy of the proposed method has a high probability of identifying a subset of useful and independent features. Moreover, clustering the features in the problem space results in reduction of the computation complexity of probability values because when an ant is placed in a given cluster, the probability value is computed only for features in the current cluster. Furthermore, unlike most of the existing ACO-based feature selection methods which use a learning algorithm [1,11,37] to evaluate the constructed subsets, in this paper a feature subset is evaluated by means of a separability index matrix without using any learning models. Therefore the proposed method can be classified as a filter-based approach and thus it will be computationally efficient for high-dimensional datasets.

The rest of this paper is organized as follows. Section 2 reviews related works on feature selection. A detailed description of our proposed method, including the complexity analysis of the different steps, is presented in detail in Section 3. In Section 4 we compare the proposed algorithm with other existing feature selection methods. Finally, Section 5 summarizes the present study.

## 2. Related works

The main idea behind feature selection is to choose a subset of available features, by eliminating irrelevant features with little or no predictive information, as well as redundant features that are strongly correlated. To find the optimal feature subset one needs to enumerate and evaluate all the possible subsets of the features. The entire search space contains all the possible subsets of features, meaning that the search space size is  $2^n$  where  $n$  is the dimensionality of the problem (i.e., the number of original features). Therefore, the problem of finding the optimal feature subset is NP-hard [10,19]. Since evaluating the entire feature subsets is computationally expensive, time consuming and also impractical even for moderate sized feature sets, the final solution should be found in a feasible computational time with a reasonable trade-off between the quality of the found solution and time-space cost. Therefore, many feature selection algorithms involve heuristic or random search strategies to find the optimal or near optimal subset of features in order to reduce the computational time. Feature selection is a fundamental research topic in machine learning with a long history since the 1970s, and there are a number of attempts to review the feature selection methods [10,51].

The feature selection methods can be classified into four categories including filter, wrapper, embedded, and hybrid models. The filter approach requires only a statistical analysis on a feature set for solving the feature selection task without utilizing any learning algorithms. Therefore, the methods in this approach are typically fast. The filter-based feature selection methods can be classified into univariate and multivariate methods. In the univariate methods, the informativeness of each feature is evaluated

Download English Version:

<https://daneshyari.com/en/article/402288>

Download Persian Version:

<https://daneshyari.com/article/402288>

[Daneshyari.com](https://daneshyari.com)