Expert Systems with Applications 41 (2014) 1622–1631

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Filter-based optimization techniques for selection of feature subsets in ensemble systems

Laura Emmanuella A. dos S. Santana, Anne M. de Paula Canuto *

Informatics and Applied Mathematics Department, Federal University of RN, Lagoa Nova, Natal, RN 59072-970, Brazil

ARTICLE INFO

Keywords:
Ensemble systems
Feature selection
Particle swarm optimization
Ant colony optimization
Genetic algorithms

ABSTRACT

Feature selection methods select a subset of attributes (features) of a dataset and it is done based on a defined measure, eliminating the redundant and irrelevant ones. When a feature selection method is applied in a dataset, we aim to improve the quality of the dataset representation. For ensemble systems, feature selection techniques can supply different feature subsets for the individual components, reducing the redundancy that can exist among the features of an input pattern and to increase the diversity level of these systems. This paper proposes the application of three well-known optimization techniques (particle swarm optimization, ant-colony optimization and genetic algorithms), in both mono and bi-objective versions, to choose subsets of features for the individual components of ensembles. The feature selection process was based on two filter-based evaluation criteria that tried to capture the idea of diversity of individual classifiers and group diversity of an ensemble system. In this case, these optimization techniques try to maximize these diversities measures, either individually (mono-objective) or together (bi-objective). An empirical analysis was performed, where all ensemble systems were applied to 11 datasets and we compared both mono and bi-objective versions among each other and with a random subset procedure. Based on the empirical analysis, we will observe that PSO with a bi-objective function will be the most promising direction, when selecting attributes for individual components of ensemble systems.

1. Introduction

One of the most important attempts to increase the accuracy and/or generalization of a classification method is by using ensemble systems, also called machine committees. These systems are composed of parallely organized components (classifiers) whose outputs are combined by a combination process that provides the final answers of an ensemble. However, there is no gain in combining identical components and, therefore, diversity is an important issue to take into consideration when designing ensemble systems (Banfield, Hall, Bowyer, & Kegelmeyer, 2005; Canuto, Abreu, Oliveira, Xavier-Jr, & Santos, 2007; Kuncheva, 2004; Mao, Jiao, Xiong, & Gou, 2011). When using ensemble systems, the design of individual classifiers under different circumstances tends to increase the diversity level of these systems, which are: initial parameter setting, training datasets and/or classification algorithms. For the second case, the use of feature selection techniques tends to increase diversity of the individual classifiers of the ensembles (Yu et al., 2003). It happens because each input pattern is handled redundantly by individual classifiers but these classifiers were designed with different views of the same problem (subsets of features). Therefore, feature selection methods aim to reduce redundancy among features of a dataset and to enhance diversity of an ensemble.

Feature subset selection is described as the search for one or more subsets of attributes, becoming a search problem. This process is defined in the following way: Suppose that X is the original set of attributes, where $\{x_i \in R^d | i = 1, 2, \ldots, n\}$ are the unlabeled examples. We would like to select a subset S, with s features that represent the original data, where $s < d$. In the context of ensembles, we will select c subsets, $S_j | j = 1, \ldots, c$ where c represents the individual classifiers that compounds an ensemble system. Therefore, each subset of S_j has a cardinality $s_i < d$ and $S_j \subseteq X$. For this, a function $f(X)$ is applied as an evaluation criterion and it should be maximized or minimized, according to the chosen problem.

When considering feature subset selection as a search problem, it is possible to use search strategies to solve these problems, applying exhaustive or heuristic approaches. Among the heuristic search algorithms, the most popular technique is genetic algorithm since it has been successfully applied in traditional classification algorithms (Huang & Wang, 2006; Siedlecki & Sklansky, 1989) and, more recently, for ensemble systems (Opitz, 1999; Santana, Silva, Canuto, Pintro, & Vale, 2010). Recently, ant colony optimization (ACO) and Particle swarm algorithms (PSO) have also been used to select attributes, but mainly for single classifiers.

Most of the works related to feature selection methods use a wrapper approach (search process is dependent on a classification

* Corresponding author. Tel.: +55 84 3215 3815.
E-mail addresses: lauraemmanuella@yahoo.com.br (L.E.A.S.), anne@dimap.ufrn.br (A.M.P.).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.08.059

algorithm), where the accuracy of a classifier can be employed as an evaluation measure. This approach is commonly chosen because it is often more robust than the filter approach (search process is independent on a classification method), since the feature selection can be optimized for the specific classification algorithm. Nevertheless, this model has a high computational cost and it can become prohibitive for cases in which the dataset contains a large number of attributes. Furthermore, since the feature selection procedure is strongly coupled to the used classification algorithm, the wrapper approach is less general than filter, having to run again when switching classification methods. In ensemble systems, the main disadvantage of the wrapper approach is enhanced, because an evaluation criterion typically considers the accuracy level of the entire ensemble system. This causes an increase in the complexity of this process. We believe that a feature technique method needs to be simple, robust and efficient. Therefore, filter-based feature selection methods happen to be an interesting (and efficient) option for selecting features in ensembles.

This paper presents an analysis of three well-known heuristic-based (metaheuristic) optimization techniques, in both mono and bi-objective versions, for selecting subsets of attributes in ensemble systems. In order to do that, we will consider diversity of both individual classifiers and ensemble systems as evaluation criteria for the optimization techniques. Thus, we can describe the major contributions of this paper as follows.

1. The use of two filter-based evaluation criteria as a guide of the optimization techniques to select subsets of features in ensemble systems. In Santana (2012), we have conducted an empirical analysis and showed that these filter-based evaluation criteria had a similar performance of some elaborate wrapper-based techniques.
2. The use of particle swarm optimization (PSO), ant colony optimization (ACO) and genetic algorithm (GA) mainly in their bi-objective versions, to select feature subsets for ensemble systems. In our previous work (Santana, Canuto, & Silva, 2011), we applied these three optimization techniques as feature selectors in ensemble systems. However, we used only the mono-objective version of these algorithms. Furthermore, we used fewer datasets and configurations that we are using in this study;
3. A comparative analysis of mono and multi-objective versions of all three optimization techniques. This analysis will be done in terms of two important parameters in ensemble systems, combination method and ensemble size.

2. Recent studies in feature selection and ensemble systems

Several studies have analyzed the application of feature selection techniques for classification algorithms, including ensemble systems (Correa, Freitas, & Johnson, 2008; Huang & Wang, 2006; Inza, Larraaga, Etxeberria, & Sierra, 2000; Kittler, 1978; Kohavi & John, 1997; Robbins, Zhang, Bertrand, & Rekaya, 2007; Santana et al., 2010; Siedlecki & Sklansky, 1989). In the literature, these feature selection techniques are divided into two main approaches, as follows.

- Filter: No classification algorithm is used during the process of selecting features. Specifically, the process of selecting features is not guided by the performance of a classification method. The filter approach is usually more efficient and less robust than the wrapper approach;
- Wrapper: This process of selecting features is guided by the accuracy of a classification algorithm. A subset of features is chosen based on the performance of a specific classification method. Two different classification algorithms lead to the choice of different feature subsets. The major drawback of this approach is related to the huge computational complexity to evaluate the obtained feature subsets by running a specific classification algorithm on a dataset for all subsets.

In traditional feature selection methods (using individual algorithms), in which the robustness (accuracy) of these methods is the most important parameter to be optimized, the use of wrapper-based approach has better performance than filter-based approaches. However, in the context of ensembles, other important issues should be considered when choosing a feature selection technique and diversity can be considered as one of them. As ensemble systems use a two-step procedure as a decision making method (the base classifiers as well as the fusion phase), the relation between the chosen feature subsets and the classification algorithms tends to decrease. Therefore, the accuracy of an individual classifier is no longer the only parameter that is important to characterize the accuracy level of an ensemble system. In addition, the poor performance of one classifier can be overcome by a different classifier with high performance. In this case, diversity is considered as an important aspect for the efficiency of an ensemble system. Additionally, as mentioned previously, the major disadvantage of wrapper approach is enhanced in these systems, since the objective function usually uses the accuracy of the entire system, making this approach even more complex than when using only one classification algorithm.

In some studies, as in Oliveira, Morita, and Sabourin (2006) and Ko, Sabourin, and de Souza Britt2006,2006, the objective functions take into consideration the accuracy of individual classifiers. Based on this, different feature subsets are chosen and distributed over the individual components. Nevertheless, they did not consider the accuracy level of the entire ensemble. In this case, the objective functions usually did not reveal the real situation of these ensembles. In Santana (2012), we performed a comparative analysis using a classical wrapper-based approach with the filter-based approaches used in this paper and they were not proved to be statistically different. Therefore, the idea of using filter-based approach to choose feature subsets for the individual components in an ensemble system became an important option.

One technique that has been widely applied in the automatic search to find the optimal feature subsets is the optimization technique. For instance, some studies have analyzed the use of genetic algorithms (GA) for designing ensemble systems (Lee et al., 2008,; Nascimento & Canuto, 2012,; Oliveira et al., 2006; Opitz, 1999). In Opitz (1999), genetic algorithms were used to choose different subsets of features for an ensemble system. The idea was to create different attribute subsets, one for each individual classifier. An empirical analysis was conducted using ensembles of neural networks. Additionally, they compared the results obtained by their approach with two well-known techniques for building ensembles, which are: Adaboosting and Bagging.

Ant colony optimization (ACO) has been basically applied to select attributes in single classifiers, such as in Robbins et al. (2007). For ensemble systems, in Santana et al. (2010), mono-objective ACO has been used to select feature subsets and its performance was compared with a genetic algorithm (GA). In accordance with the authors, ensembles using ACO provided higher accuracy rates than GA-based ensembles, when few individual classifiers are considered (three individual classifiers). However, when increasing the number of individual classifiers, GA-based ensemble systems delivered higher accuracy rate than ACO-based ensembles.

The use of particle swarm optimization (PSO) to select attributes of datasets started recently, mainly for single classifiers (Correa et al., 2008). In the context of ensembles, an investigation was done in Santana et al. (2011), comparing mono-objective versions of PSO with ACO and GA. The authors showed that PSO delivered