

# Multi-objective Genetic Algorithms to Create Ensemble of Classifiers

Luiz S. Oliveira, Marisa Morita, Robert Sabourin, and Flávio Bortolozzi

<sup>1</sup> Pontifícia Universidade Católica do Paraná, Curitiba, Brazil

<sup>2</sup> Universidade Tuiuti do Paraná, Curitiba, Brazil

<sup>3</sup> Ecole de Technologie Supérieure, Montreal, Canada

soares@ppgia.pucpr.br

**Abstract.** Feature selection for ensembles has shown to be an effective strategy for ensemble creation due to its ability of producing good subsets of features, which make the classifiers of the ensemble disagree on difficult cases. In this paper we present an ensemble feature selection approach based on a hierarchical multi-objective genetic algorithm. The algorithm operates in two levels. Firstly, it performs feature selection in order to generate a set of classifiers and then it chooses the best team of classifiers. In order to show its robustness, the method is evaluated in two different contexts: supervised and unsupervised feature selection. In the former, we have considered the problem of handwritten digit recognition while in the latter, we took into account the problem of handwritten month word recognition. Experiments and comparisons with classical methods, such as Bagging and Boosting, demonstrated that the proposed methodology brings compelling improvements when classifiers have to work with very low error rates.

## 1 Introduction

Ensemble of classifiers has been widely used to reduce model uncertainty and improve generalization performance. Developing techniques for generating candidate ensemble members is a very important direction of ensemble of classifiers research. It has been demonstrated that a good ensemble is one where the individual classifiers in the ensemble are both accurate and make their errors on different parts of the input space [7]. In other words, an ideal ensemble consists of good classifiers (not necessarily excellent) that disagree as much as possible on difficult cases.

The literature has shown that varying the feature subsets used by each member of the ensemble should help to promote this necessary diversity [6, 15, 18]. Traditional feature selection algorithms aim at finding the best trade-off between features and generalization. On the other hand, ensemble feature selection has the additional goal of finding a set of feature sets that will promote disagreement among the component members of the ensemble. The Random Subspace Method (RMS) proposed by Ho in [6] was one early algorithm that constructs an ensemble by varying the subset of features. Strategies based on genetic algorithms (GAs) also have been proposed [5, 15]. All these strategies claim better results than those produced by traditional methods for creating ensembles such as Bagging and Boosting. In spite of the good results brought by GA-based methods, they still can be improved in some aspects, e.g., avoiding classical methods such as

the weighted sum to combine multiple objective functions. It is well known that when dealing with this kind of combination, one should deal with problems such as scaling and sensitivity towards the weights.

It has been demonstrated that feature selection through multi-objective genetic algorithm (MOGA) is a very powerful tool for finding a set of good classifiers [4, 14], since GA is quite effective in rapid global search of large, non-linear and poorly understood spaces [17]. Besides, it can overcome problems such as scaling and sensitivity towards the weights. Kudo and Sklansky [8] have compared several algorithms for feature selection and concluded that GAs are suitable when dealing with large-scale feature selection (number of features is over 50). This is the case of most of the problems in handwriting recognition, which is the test problem in this work.

In this light, we propose an ensemble feature selection approach based on a hierarchical MOGA. The underlying paradigm is the “overproduce and choose” [16]. The algorithm operates in two levels. The former is devoted to generate a set of good classifiers by minimizing two criteria: error rate and number of features. The latter combines these classifiers in order to find an ensemble by maximizing the following two criteria: accuracy of the ensemble and a measure of diversity. We demonstrated through experimentation that using diversity jointly with performance to guide selection can avoid overfitting during the search.

In order to show robustness of the proposed methodology, it was evaluated in two different contexts: supervised and unsupervised feature selection. In the former, we have considered the problem of handwritten digit recognition and used three different feature sets and multi-layer perceptron (MLP) neural networks as classifiers. In the latter, we took into account the problem of handwritten month word recognition and used three different feature sets and hidden Markov models (HMM) as classifiers. We demonstrate that it is feasible to find compact clusters and complementary high-level representations (codebooks) in subspaces without using the recognition results of the system. Experiments and comparisons with classical methods, such as Bagging and Boosting, demonstrated that the proposed methodology brings compelling improvements when classifiers have to work with very low error rates.

## 2 Methodology Overview

In this section we outline the hierarchical approach proposed. As stated before, it is based on an “overproduce and choose” paradigm where the first level generates several classifiers by conducting feature selection and the second one chooses the best ensemble among such classifiers. Figure 1 depicts the proposed methodology. Firstly, we carry out feature selection by using a MOGA. It gets as inputs a trained classifier and its respective data set. Since the algorithm aims at minimizing two criteria during the search<sup>1</sup>, it will produce at the end a 2-dimensional Pareto-optimal front, which contains a set of classifiers (trade-offs between the criteria being optimized). The final step of this first level consists in training such classifiers.

---

<sup>1</sup> Error rate and number of features in the case of supervised feature selection and a clustering index and the number of features in the case of unsupervised feature selection.