Knowledge-Based Systems 54 (2013) 3–21

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Multiobjective genetic classifier selection for random oracles fuzzy rule-based classifier ensembles: How beneficial is the additional diversity?

Krzysztof Trawiński [a,*], Oscar Cordón [a,b], Arnaud Quirin [c], Luciano Sánchez [d]

[a] European Centre for Soft Computing, Edificio de Investigación, Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Asturias, Spain
[b] Department of Computer Science and Artificial Intelligence (DECSAI) and Research Center on Information and communication Technologies (CITIC-UGR), University of Granada, 18071 Granada, Spain
[c] Galician Research and Development Center in Advanced Telecommunications (GRADIANT), Communications Area, Edif. CITEXVI, Local 14, 36310 Vigo, Spain
[d] Computer Science Department, University of Oviedo, Campus de Viesques, 33203 Gijón, Asturias, Spain

ARTICLE INFO

Article history:
Available online 23 August 2013

Keywords:
Fuzzy rule-based classifier ensembles
Random oracles
Bagging
Genetic classifier selection
Diversity measures
Evolutionary multiobjective optimization
NSGA-II
High complexity datasets

ABSTRACT

Recently we proposed the use of the Random Linear Oracles classical classifier ensemble (CE) design methodology in a fuzzy environment. It derived fuzzy rule-based CEs obtaining an outstanding performance. Random Oracles introduce an additional diversity into the base classifiers improving the accuracy of the entire CE. Meanwhile, the overproduce-and-choose strategy leads to a good accuracy-complexity trade-off. It is based on the generation of a large number of component classifiers and a subsequent selection of the best cooperating subset of them. The current contribution has a twofold aim: (1) Introduce a new Random Oracles approach into the fuzzy rule-based CEs design; (2) Incorporate an evolutionary multi-objective overproduce-and-choose strategy to our approach analyzing the influence of this additional diversity in the final CE performance (focusing on the accuracy). To do so, firstly, we incorporate the two Random Oracle variants into the fuzzy rule-based CE framework. Then, we use NSGA-II to provide a specific component classifier selection driven by three different criteria. Exhaustive experiments are carried out over 29 UCI and KEEL datasets with high complexity (considering both the number of attributes as well as the number of examples) showing the good performance of the proposed approach.

1. Introduction

Classifier ensembles (CEs), also called multiclassifiers, are well-recognized tools in the machine learning community and more recently in the soft computing community. They are able not only to outperform a single classifier but also to deal with complex and high dimensional classification problems [1].

In a preceding contribution [2], we incorporated Random Linear Oracles (RLOs) [3], a classical CE design methodology, into a previously proposed CE framework [4] to derive fuzzy rule-based classifier ensembles (FRBCEs). Thanks to the additional diversity introduced by RLOs into the robust FURIA-based fuzzy classifiers [5,6], the obtained FRBCEs were able to achieve an outstanding performance in terms of accuracy, outperforming RLO combined with the classical base classifiers.

Nevertheless, the performance of FRBCEs can still be improved. It has been theoretically and empirically shown that smaller ensembles can outperform larger ones [7–9]. Thus, selecting a subset of classifiers is a natural way to follow. In our previous contributions, we used the well known *overproduce-and-choose strategy* [10] (OCS) to reduce the CE dimensionality, while improving its accuracy. OCS is a classifier selection method based on the generation of a large number of component classifiers and a subsequent selection of the best cooperating subset of them.

Therefore, OCS helps to obtain a good accuracy-complexity trade-off in the CE design as well as in many cases it also improves the accuracy of the final CE. In fact, these characteristics were exhibited in [11] for FRBCEs using an OCS strategy based on NSGA-II [12]. NSGA-II, which is a state-of-the-art evolutionary multi-objective (EMO) algorithm [13], generated a set of CE designs with different accuracy-complexity trade-offs in a single run.

In this contribution, we introduce two novel aspects to our FRBCE design methodology in [2] in order to improve the CE accuracy, while reducing its complexity:

* Corresponding author. Tel.: +34 985456545.
E-mail addresses: krzysztof.trawinski@softcomputing.es (K. Trawiński), oscar.cordon@softcomputing.es, ocordon@decsai.ugr.es (O. Cordón), aquirin@gradiant.org (A. Quirin), luciano@uniovi.es (L. Sánchez).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.08.006

1. To keep a high diversity in the set of classifiers as well as high performance, we incorporate a new Random Oracle (RO) approach, namely the Random Spherical Oracle (RSO) [14], into the FRBCE framework. Opposite to RLO, RSO uses an oracle based on a random hypersphere to divide the feature space into two regions in order to feed two subclassifiers, which both compose the final RSO. We expect to improve the performance of the FRBCEs by combining the RSO randomness and its oracle shapes with the "soft boundaries" provided by the FURIA-based component classifier.
2. To reduce the complexity, we design a specific EMO-based OCS strategy for RO-based FRBCEs from our previous proposal in [11]. Since RO is composed of two base classifiers, this approach offers a tremendous advantage over bagging FURIA-based component classifiers because each classifier can be independently selected within each pair component. A higher degree of freedom is achieved during the selection procedure, while still having the potential of drastically reducing the complexity.

On the one hand, we aim to obtain a good accuracy-complexity trade-off when dealing with high complexity datasets. While the main goal in the design of CEs is to obtain an accurate system, the complexity is an interesting secondary objective allowing us to obtain simpler and quicker CEs. On the other hand, we aim to analyze whether the additional diversity induced by ROs is beneficial for the EMO OCS-based FRBCEs. That is, our goal is to check if the OCS-based selection leads to more accurate results when applied on RO-based FURIA fuzzy CEs than on bagging fuzzy CEs thanks to the additional freedom degrees resulting from the RO design. For that purpose, we use a novel NSGA-II design with a three-objective fitness function including an advanced accuracy measure as well as complexity and diversity indices for the component classifier selection. Specifically, we propose a special binary coding for NSGA-II in order to take advantage of the additional degrees of freedom offered by the RO base classifiers, and test two different mutation operator settings to look for the best performance.

To perform the experimental analysis, we carry out exhaustive experiments on 29 high complexity datasets from the UCI machine learning [15] and the KEEL dataset [16] repositories.

This paper is set up as follows. In the next section, the preliminaries required for a good understanding of our work are reviewed. Section 3 presents RLOs, RSOs, both RLO- and RSO-based FRBCEs, and a set of experiments focused on the comparison of different RO-based strategies for the combination of the component classifiers. Then, Section 4 introduces our NSGA-II proposal for RSO component fuzzy classifier selection incorporating a three-objective fitness function and the analysis of the experiments performed. Finally, Section 5 concludes this contribution with some future research lines.

2. Preliminaries

This section explores the current literature related to the generation of a FRBCE. The techniques used to generate CEs and fuzzy CEs are described in Sections 2.1 and 2.2, respectively. Some ways to reduce the size of the ensembles are described in Section 2.3. The use of genetic algorithms (GAs) within the OCS strategy is explored in Section 2.4. Finally, we briefly introduce evolutionary fuzzy systems in Section 2.5.

2.1. Classifier ensembles design methodologies

A CE is the result of the combination of the outputs of a group of individually trained classifiers in order to get a system that is usually more accurate than any of its single components [1]. These kinds of methods have gained a large acceptance in the machine learning community during the last two decades due to their high performance. Decision trees are the most common classifier structure considered and much work has been done in the topic [17,18], although CEs can be used with any other type of classifiers (neural networks are also very extended, see for example [19]).

There are different ways to design a classifier ensemble. On the one hand, there is a classical group of approaches considering *data resampling* to obtain different training sets to derive each individual classifier. In *bagging* [7], they are independently learnt from resampled training sets ("bags"), which are randomly selected with replacement from the original training data set. *Boosting* methods [20] sequentially generate the individual classifiers (weak learners) by selecting the training set for each of them based on the performance of the previous classifier(s) in the series. Opposed to bagging, the resampling process gives a higher selection probability to the incorrectly predicted examples by the previous classifiers.

On the other hand, a second group can be found comprised by a more diverse set of approaches which induce the individual classifier diversity using some ways different from resampling [21]. Feature selection plays a key role in many of them where each classifier is derived by considering a different subset of the original features [22,23]. *Random subspace* [24], where each feature subset is randomly generated, is one of the most representative methods of this kind.

Finally, there are some advanced proposals that can be considered as a combination of the two groups, such as *random forests* [25] and more recently *rotation forest* [26] and *fuzzy random forest* [27].

The interested reader is referred to [18,19] for two surveys for the case of decision tree (both) and neural network ensembles (the latter), including exhaustive experimental studies.

2.2. Related work on fuzzy classifier ensembles

Focusing on fuzzy CEs, only a few contributions for bagging fuzzy classifiers have been proposed considering fuzzy neural networks (together with feature selection) [28], neuro-fuzzy systems [29], and fuzzy decision trees [27,30] as component classifier structures.

Especially worth mentioning is the contribution of Bonissone et al. [27]. This approach hybridizes Breiman's idea of random forests [25] with fuzzy decision trees [31]. Such resulting fuzzy random forest combines characteristics of CEs with randomness and fuzzy logic in order to obtain a high quality system joining robustness, diversity, and flexibility to not only deal with traditional classification problems but also with imperfect and noisy datasets. The results show that this approach obtains good performance in terms of accuracy for all the latter kind of classification problems.

Some advanced Evolutionary Fuzzy System-based contributions should also be remarked. On the one hand, a fuzzy rule-based classifier system (FRBCS) ensemble design technique is proposed in [32] considering feature selection methods based on some niching GA [33] to generate the diverse component classifiers, and another GA for classifier fusion by learning the combination weights. On the other hand, another interval and fuzzy the rule-based ensemble design method using a single- and multiobjective genetic selection process is introduced in [34,35]. In this case, the coding scheme allows an initial set of either interval or fuzzy rules, considering the use of different features in their antecedents, to be distributed among different component classifiers trying to make them as diverse as possible by means of two accuracy and one entropy measures. Besides, the same authors presented a previous proposal in [36], where an EMO algorithm generated a Pareto set