International Journal of Approximate Reasoning 73 (2016) 108–122

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Enhancing evolutionary fuzzy systems for multi-class problems: Distance-based relative competence weighting with truncated confidences (DRCW-TC) ☆

Alberto Fernández [a,*], Mikel Elkano [b], Mikel Galar [b], José Antonio Sanz [b], Saleh Alshomrani [d,f], Humberto Bustince [b,c], Francisco Herrera [e,f]

[a] Department of Computer Science, University of Jaén, 23071, Jaén, Spain
[b] Departamento de Automática y Computación, Universidad Pública de Navarra, 31006, Pamplona, Spain
[c] Institute of Smart Cities (ISC), Universidad Pública de Navarra, 31006, Pamplona, Spain
[d] Faculty of Computing and Information Technology, University of Jeddah, 21589, Jeddah, Saudi Arabia
[e] Department of Computer Science and Artificial Intelligence, University of Granada, 18071, Granada, Spain
[f] Faculty of Computing and Information Technology, King Abdulaziz University, 21589, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 28 August 2015
Received in revised form 21 January 2016
Accepted 9 February 2016
Available online 23 February 2016

Keywords:
Multi-class classification
Pairwise learning
One-vs-One
Classifier selection
Evolutionary fuzzy systems
Fuzzy rule based classification systems

ABSTRACT

Classification problems with multiple classes suppose a challenge in Data Mining tasks. There is a difficulty inherent to the learning process when trying to find the most adequate discrimination functions among the different concepts within the dataset. Using Fuzzy Rule Based Classification Systems in general, and Evolutionary Fuzzy Systems in particular, provide the advantage of describing smoother borderline areas, thanks to the linguistic label-based representation.

In multi-classification, the pairwise learning approach (One-vs-One) has gained a notorious attention. However, there is certain dependence between the goodness of the confidence degrees or scores of binary classifiers, and the final performance shown by the global model. Regarding this fact, the problem of non-competent classifiers is of special relevance. It occurs when a binary classifier outputs a positive score for a couple of classes unrelated with the input example, which may degrade the final accuracy. Precisely, the previously exposed properties of fuzzy classifiers make them more prone to the former condition.

In this paper, we propose an extension of the distance-based combination strategy to overcome this non-competence problem. It is based on the truncation of the confidence degrees of the classes prior to the distance-based tuning. This allows taking advantage of the good classification abilities of Evolutionary Fuzzy Systems, while diminishing the adverse effect of the aforementioned non-competence. Experimental results, using FARC-HD with overlap functions as the fuzzy learning algorithm, show that this new adaptation of the Distance-based Relative Competence Weighting model outperforms both the OVO and standard distance-based approaches, and it is competitive with robust classifiers such as Support Vector Machines.

☆ This paper is part of the virtual special issue on IFSA – EUSFLAT 2015, edited by Oscar Cordón, Luis Magdalena and José M. Alonso.

* Corresponding author. Tel.: +34 953 213016; fax: +34 953 212472.
E-mail addresses: alberto.fernandez@ujaen.es (A. Fernández), mikel.elkano@unavarra.es (M. Elkano), mikel.galar@unavarra.es (M. Galar), joseantonio.sanz@unavarra.es (J.A. Sanz), sshomrani@kau.edu.sa (S. Alshomrani), bustince@unavarra.es (H. Bustince), herrera@decsai.ugr.es (F. Herrera).

http://dx.doi.org/10.1016/j.ijar.2016.02.005
0888-613X/

1. Introduction

Among all Data Mining tasks, classification can be regarded as one of the most studied problems [1]. It consists in learning a mapping function from the attributes of a set of examples whose class is a priori known. The final goal is to automatically determine the class for every new pattern arriving to the system. In a standard framework, the classifier is devoted to learn a dichotomization function. However, when a multi-class problem is considered, the complexity inherent to the higher number of decision boundaries, and the overlapping among these classes, makes this task harder than the binary case [2].

In accordance with the former, a successful way to address this problem is using a divide and conquer methodology. The main idea is to apply a decomposition strategy [3], i.e. to simplify the initial multi-class problems by generating several binary sub-problems. There are several strategies that consider a divide-and-conquer scheme, and most of them can be included within Error Correcting Output Codes framework [4,5]. One of the most common approaches is the One-vs-One (OVO) or pairwise learning scheme [6,7]. This approach divides the original problem in as many pairs of classes as possible, ignoring the examples that do not belong to the related classes. Then, these are learnt in an independent way by the so-called base learners or base classifiers of the ensemble [8]. In fact, even when the base classifiers are able to cope with multi-class problems, it has been shown that the use of binarization techniques allows for the achievement of a more robust system, and thus the improvement of the performance with respect to the standard case [9,10].

The simplification of the learning task implies that more attention must be focused on the combination stage, where the classifiers learned to solve each binary problem should be aggregated to output the final decision over the class label [11, 12]. Specifically, every new example to be classified is queried by all the classifiers. Each classifier outputs a pair of "scores" (confidence degrees) for the two classes that were considered for the training stage. All these score values are collected to build a decision matrix which will be used to determine the output class.[1]

When the confidence degrees generated by the baseline classifiers are good enough, then any aggregation method will be adequate to determine the correct class. In particular, the simplest and most widely used method in pairwise learning the "Weighted Voting" (WV) [13], which considers the maximum vote among the summation of the scores for the binary classifiers associated with the same class.

If we analyze this procedure in depth, we find an inherent problem: all base classifiers will be fired for a given instance, even when they have not been trained to recognize the real class of this particular instance. Therefore, these classifiers could submit an erroneous score that might alter the decision process degrading the accuracy of the final system. This case is better known as the "non-competent classifiers problem" [13], which can mislead the correct labeling of the query example.

It is straightforward to acknowledge that the competence of a classifier in the OVO approach cannot be established a priori. Therefore, the techniques that have been proposed in the specialized literature to cope with this problem are based on dynamic post-processing methodology. They are aimed at adapting the score matrix values prior to the decision step. In [14], authors compute the closest classes to the query instance in order to remove those classifiers that were not related with this query example. A similar approach was developed in [15], known as Distance-based Relative Competence Weighting (DRCW), where confidence degrees of the classifiers were weighted according to the distance computed from the example to the nearest neighbors of each class, instead of directly removing the classifiers.

Among different classification algorithms, Fuzzy Rule Based Classification Systems (FRBCSs) [16] have shown a good behavior when modeling complex problems due to the proper management of the uncertainty achieved by fuzzy sets, as well as their interpretability based on the linguistic variables [17]. Furthermore, FRBCSs can be further enhanced towards more accurate systems by including the learning and adaptation capabilities of evolutionary optimization, leading to Evolutionary Fuzzy Systems (EFSs) [18]. Their success among other Soft Computing techniques, is related to their smoothness when defining the borderline areas in complex problems [19,20].

The previous fuzzy properties stressed for this type of classifiers make them more prone to the non-competence problem. Therefore, our hypothesis is that the confidence degrees computed by the EFS classification algorithms within the OVO approach may have a higher benefit in cooperation with the distance weighting scheme. In accordance with the former, in this research our objective is three-fold:

1. First, we want to analyze experimentally whether the combination of DRCW and EFSs is able to obtain a significant improvement in the performance of the classifier.
2. Then, we will propose a modification of the DRCW-OVO model by truncating the confidence degrees in the score matrix to 0.0 and 1.0, as in "Simple Voting". The objective is to combine the good behavior shown by both schemes. This new aggregation will be noted as DRCW-TC.
3. Finally, we will contrast the performance of this new approach over different paradigms of classifiers, including Decision Trees, Support Vector Machines (SVMs), and of course EFSs. We suggest that our model will achieve a better synergy with fuzzy based classifiers, allowing it to improve the results of the crisp approaches.

[1] Throughout this document, we will refer interchangeably to "confidence degrees" or "scores" or as those values assigned to the pair of classes addressed by each single classifier. In the case of fuzzy classifiers, these values are computed as the pattern classification soundness degree for each class.

In order to do so, we carry out a thorough experimental study with 23 classification datasets selected from the KEEL repository [21]. The validation of the results will be carried out using the classical accuracy rate. Finally, this experimental study will be consolidated by means of the use of statistical tests, as suggested in the specialized literature [22–24].

The outline of this paper is as follows. First, Section 2, presents the preliminaries of this work by introducing the main features FRBCSs, EFSs and the OVO learning approach. Then, Section 3 describes the DRCW-OVO methodology that address the non-competent classifier problem, and our novel DRCW-TC proposal specifically designed for EFSs. Section 4 introduces the experimental framework. In Section 5 the experimental analysis is carried out, where the main findings of this contribution are stressed. Finally, Section 6 summarizes this study and concludes the paper.

2. Preliminaries

This section briefly introduces the main features of FRBCS (Subsection 2.1). Then, we will present some basic concepts on EFS in order to describe the FARC-HD algorithm (Subsection 2.2). Afterwards, we recall decomposition strategies, and more specifically, OVO scheme (Subsection 2.3). Finally, we present the most significant combinations used in pairwise learning (Subsection 2.4).

2.1. A short overview on FRBCS

Any classification problem consists of m training patterns $x_p = (x_{p1}, \ldots, x_{pn}, C_p)$, $p = 1, 2, \ldots, m$ from M classes where x_{pi} is the ith attribute value ($i = 1, 2, \ldots, n$) of the p-th training pattern, and C_p the output class.

In this work we use an FRBCSs with a linguistic Rule Base of size L, where fuzzy rules follow the structure below [25,26]:

$$\begin{aligned} &\text{Rule } R_j: \text{ If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \\ &\text{then Class} = C_j \text{ with } RW_j; \quad j \in \{1, \ldots, L\} \end{aligned} \tag{1}$$

where R_j is the label of the jth rule, $x = (x_1, \ldots, x_n)$ is an n-dimensional pattern vector, A_{ji} is an antecedent fuzzy set, C_j is a class label ($C_j \in \{1, \ldots, M\}$), and RW_j is the rule weight [27]. In this work, linguistic labels for the antecedent fuzzy sets are represented as triangular membership functions.

When a new pattern x_p is selected for classification, then the steps of the fuzzy reasoning method are as follows:

1. **Matching degree**, that is, the strength of activation of the if-part for all rules in the Rule Base with the pattern x_p. In order to carry out this computation, a conjunction operator γ shall be applied. This operator is used to combine the membership degrees for every variable of the example, which were obtained by means of the μ function. Traditionally, a T-norm is selected for this purpose, although any aggregation operator can be employed [28]:

$$\begin{aligned} &\mu_{R_j}(x_p) = \gamma(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})), \\ &\quad j \in \{1, \ldots, L\} \end{aligned} \tag{2}$$

2. **Association degree**. To compute the association degree of the pattern x_p with the M classes according to each rule in the Rule Base. To this end, a combination operator h is applied in order to combine the matching degree with the rule weight (RW). In our case, this association degree only refers to the consequent class of the rule.

$$b_j^k = h(\mu_{R_j}(x_p), RW_j^k), \quad j \in \{1, \ldots, L\}; \quad k = j \tag{3}$$

3. **Pattern classification soundness degree for all classes**. We use an aggregation function f, which combines the positive degrees of association calculated in the previous step.

$$\begin{aligned} &Y_k = f(b_j^k, j = 1, \ldots, L \text{ and } b_j^k > 0), \\ &\quad k \in \{1, \ldots, M\} \end{aligned} \tag{4}$$

4. **Classification**. We apply a decision function F over the soundness degree of the system for the pattern classification for all classes. This function will determine the class label l corresponding to the maximum value.

$$F(Y_1, \ldots, Y_M) = \arg\max(Y_k), \quad [k \in \{1, \ldots, M\}] \tag{5}$$

Where L denotes the number of rules in the Rule Base and M the number of classes of the problem.

2.2. Evolutionary fuzzy systems and FARC-HD algorithm

EFSs are a family of approaches that are built on top of FRBSs, whose components are improved by means of an evolutionary learning/optimization process as depicted in Fig. 1. This process is designed for acting or tuning the elements of a fuzzy system in order to improve its behavior in a particular context. Traditionally, this was carried out by means of Genetic

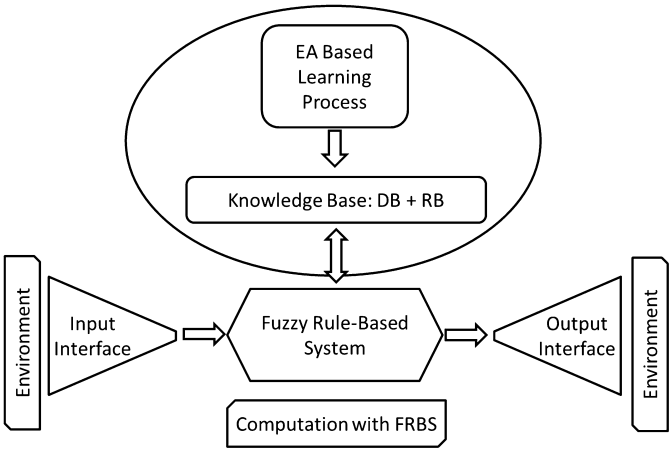

Fig. 1. Integration of an EFS on top of an FRBS.

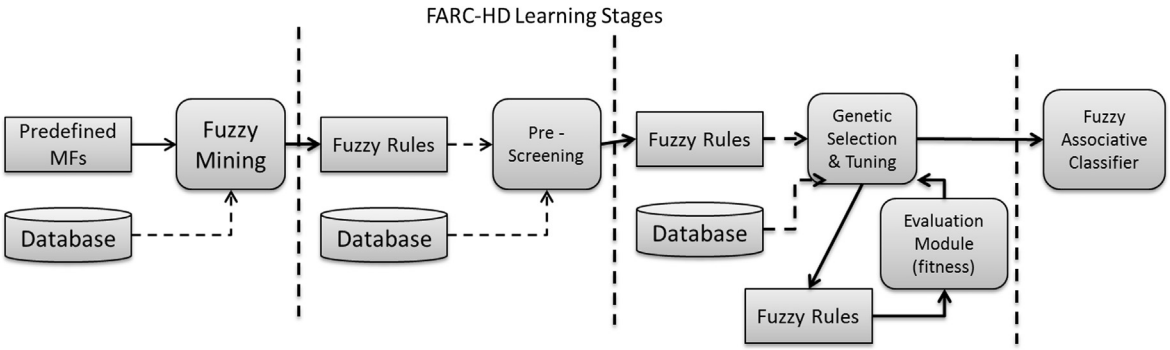

Fig. 2. Learning stages for the FARC-HD algorithm.

Algorithms, leading to the classical term of Genetic Fuzzy Systems [29–32]. More recently, a generalization of the former was consider by means the use of Evolutionary Algorithms in the learning stage [18].

In any Data Mining problem, and especially in classification tasks, the use of any fuzzy sets approach is usually considered when an interpretable system is sought, when the uncertainty involved in the data must be properly managed, or even when a dynamic model is under consideration. Depending on the complexity of the problem it might be solved by a simple FRBCS; however, more sophisticated solutions can be required when we aim at improving the achieved accuracy of the output system.

In these cases, EFSs are a proper approach by learning or tuning the Knowledge Base components of the FRBCS. In this paper we have make use of a robust linguistic FRBCS, i.e. the Fuzzy Association Rule-based Classification for High-Dimensional problems (FARC-HD) approach [33]. This algorithm is based on association discovery, a commonly used technique in data mining for extract interesting knowledge from large datasets [34] by means of finding relationships between the different items in a database [35]. The integration between association discovery and classification leads to precise and interpretable models.

FARC-HD is aimed at obtaining an accurate and compact FRBCS with a low computational cost. In short, this method is based on the following three stages (as depicted in Fig. 2):

Stage 1 *Fuzzy association rule extraction for classification*: A search tree is employed to list all possible frequent fuzzy item sets and to generate fuzzy association rules for classification, limiting the depth of the branches in order to find a small number of short (i.e., simple) fuzzy rules.

Stage 2 *Candidate rule pre-screening*: After the rule generation, the size of the rule set can be too large to be interpretable by the end user. Therefore, a pre-selection of the most interesting rules is carried out by means of a "subgroup discovery" mechanism based on an improved weighted relative accuracy measure (wWRAcc') [36].

Stage 3 *Genetic rule selection and lateral tuning*: Finally, in order to obtain a compact and accurate set of rules within the context of each problem, an evolutionary process will be carried out in a combination for the selection of the rules with a tuning of membership function, as its positive synergy has been shown in previous work on the topic [37,38].

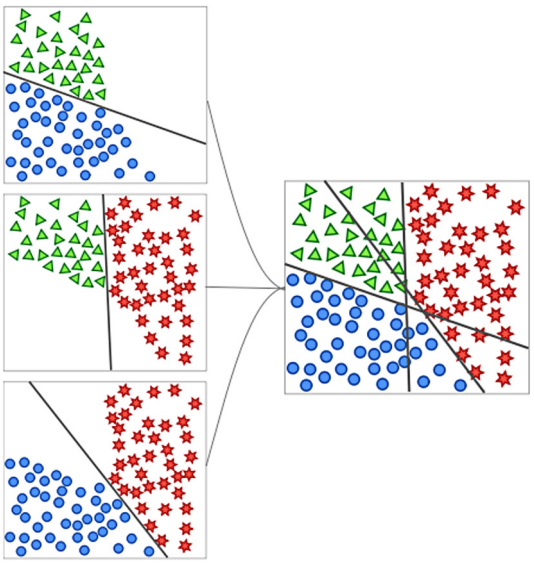

Fig. 3. Example of the OVO binarization technique for a 3-class problem.

In this research, we consider an extension of the FARC-HD algorithm. Particularly, in the first two stages of the inference process, the use of a conjunction and combination operators (functions γ and h in Eqs. (2), (3)) are replaced with an overlap function [39], aiming at enhancing the confidence values provided by the fuzzy reasoning method. Standard overlap functions were originally defined in just two two-dimensions, since they were employed in image processing tasks [40]. Then, in [28] authors extend them to n dimensions, providing some valid examples such as the minimum, product, harmonic mean, or sinusoid functions. The application of these functions allows at obtaining more adequate outputs from the base classifiers for the subsequent aggregation in OVO schemes. Thus, and according to the performance shown in [28], we will make use of the Harmonic Mean (HM). The returned value corresponds the harmonic mean of input values if all of them are different than zero and 0 otherwise (Eq. (6)).

$$O^n(x_1, x_2, \ldots, x_n) = \begin{cases} \frac{n}{\frac{1}{x_1}+\ldots+\frac{1}{x_n}} & if\, x_i \neq 0, \text{ for all } i = 1, \ldots, n \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

2.3. Decomposition strategies: the One-vs-One scheme

The use of decomposition strategies in multi-classification has shown to be of great interest in the research community [41,9]. This scheme is used to address the classification by simplifying the original problem into binary-class subsets, following a divide and conquer paradigm. It is straightforward to acknowledge that the boundaries between two classes are easier to learnt than in the general case, when they are more likely to be highly overlapped. Therefore, the critical step is moved towards the decision process, in which the confidence degrees of all binary-classifiers must be aggregated in order to output a single class.

Among different types of methodologies to carry out the decomposition process, the OVO approach [7] is one of the most extended schemes. Its popularity is mainly due to its simplicity and accuracy, which has made it the default model to handle multi-class problems with SVMs [42] (whose multi-class extension is not established yet [43,44]) in several widely used software tools [45–47]. We must stress that decomposition strategies are not only useful for binary classifiers, but also for improving the performance of classifiers that inherently support multiple classes [48,43,9,49], including of course FRBCSs [50,51,28].

In OVO, also known as pairwise classification, the original m-class problem is divided into $m \cdot (m-1)/2$ two-class problems, one for each possible pair of classes. Each sub-problem is faced by a classifier, which only takes into account instances from both classes it is responsible for. An example of this binarization technique is depicted in Fig. 3.

The way pairwise learning is carried out, causes the so-called *non-competence* problem [52–54,15] in testing phase. The reason is that in this phase every classifier output is considered to decide the final class, although some classifiers would not have been trained with instances of the class to be predicted. In order to do so, it is usual to construct a score-matrix R containing these outputs, which are used to decide the final class via different combination models:

$$R = \begin{pmatrix} - & r_{12} & \cdots & r_{1m} \\ r_{21} & - & \cdots & r_{2m} \\ \vdots & & & \vdots \\ r_{m1} & r_{m2} & \cdots & - \end{pmatrix} \tag{7}$$

where $r_{ij} \in [0, 1]$ represents the confidence of the classifier discriminating classes i and j in favor of the former; whereas the confidence for the latter is computed by $r_{ji} = 1 - r_{ij}$ (if the classifier does not provide it). In the case of FRBCS, these values correspond to the computation of the soundness degree for each class, i.e. values Y_k as pointed out by Eq. (4). We must point out that all scores are then normalized to the range $[0, 1]$.

2.4. Combination strategies for OVO scheme

Several strategies for combining OVO classifiers have been proposed in the literature that intend to achieve the highest accuracy addressing different features of this inference step. In [53], a thorough review and experimental comparison was developed considering the most-recent and well-known techniques. From this study, it was possible to select the better suited combination strategies for different paradigms of classifiers:

- *Simple Voting strategy* (VOTE) (also called binary voting and Max–Wins rule [55]). Each binary classifier gives a vote for the predicted class. The votes received by each class are counted and the class with the largest number of votes is predicted:

$$Class = arg \max_{i=1,\ldots,m} \sum_{1 \le j \ne i \le m} s_{ij} \tag{8}$$

  where s_{ij} is 1 if $r_{ij} > r_{ji}$ and 0 otherwise.
- *Weighted Voting strategy* (WV) [52] uses the confidence of each base classifier in each class to vote for it. The class with the largest total confidence is the final output class:

$$Class = arg \max_{i=1,\ldots,m} \sum_{1 \le j \ne i \le m} r_{ij} \tag{9}$$

- *Wu et al. Probability Estimates by Pairwise Coupling approach* (PE) [56] aims at estimating the posterior probabilities of all the classes starting from the pairwise class probabilities. Therefore, being $r_{ij} = \text{Prob}(Class_i \mid Class_i \text{ or } Class_j)$, the method finds the best approximation of the class posterior probabilities $\mathbf{p} = (p_1, \ldots, p_m)$ according to the pairwise outputs. Finally, the class having the largest posterior probability is predicted:

$$Class = arg \max_{i=1,\ldots,m} \widehat{p}_i \tag{10}$$

  The posterior probabilities ($\mathbf{p}$) are computed solving the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{p}} \sum_{i=1}^{m} \sum_{1 \le j \ne i \le m} (r_{ji} p_i - r_{ij} p_j)^2 \\ &\text{subject to } \sum_{i=1}^{m} p_i = 1, \; p_i \ge 0, \text{ for all } i \in \{1, \ldots, m\}. \end{aligned} \tag{11}$$

3. Addressing the non-competence problem in OVO strategy: a truncated weighting approach

In this section we will first describe the approach developed for addressing the non-competence classifier problem in an OVO environment (Subsection 3.1). Next, we will present our new proposal to take the most advantage from the confidence degrees given by EFS classifiers, which we have named as DRCW-TC (Subsection 3.2).

3.1. DRCW-WV: distance-based relative competence weighting for One-vs-One

In an OVO learning scheme, a classifier is said to be "non-competent" to classify an instance whenever the actual class of this query example does not corresponds with none of the pair of classes learned by the binary classifier [13]. Although this problem implies "noise" in the decision score-matrix, it does not necessarily imply a decrease in the performance if all competent classifiers achieve a high confidence degree. However, when competent classifiers fail, the final decision is influenced by the non-competent ones, which could lead to a misclassification of the instance.

In order to alleviate the negative effect of the non-competent classifiers, the DRCW-WV approach was recently developed [15]. This methodology consists of carrying out a dynamic adaptation of the score-matrix, altering the confidence degrees by

assigning a higher weight to those classifiers whose predicted classes are closer to that of the query instance, assuming that they are more competent than those which are farther. This distance is computed by using the standard k-nearest neighbors approach. In this way, once the score-matrix has been obtained, the operating procedure of DRCW-WV is as follows.

1. Compute the k nearest neighbors of each class for the given instance and store the average distances of the k neighbors of each class in a vector $\mathbf{d} = (d_1, \ldots, d_m)$.
2. A new score-matrix R^w is created where the output r_{ij} of a classifier distinguishing classes i, j are weighted as follows,
$$r_{ij}^w = r_{ij} \cdot w_{ij}, \tag{12}$$
where w_{ij} is the relative competence of the classifier on the corresponding output computed as
$$w_{ij} = \frac{d_j^2}{d_i^2 + d_j^2} \tag{13}$$
being d_i the distance of the instance to the nearest neighbor of class i.
3. Use weighted voting strategy on the modified score-matrix R^w to obtain the final class.

Finally, the decision of the output class is carried out by means of the Weighted Voting method since its robustness has been both theoretically [13] and experimentally [9] proved. Considering this combination, steps 2) and 3) are merged obtaining the output class as follows,
$$Class = \arg \max_{i=1,\ldots,m} \sum_{1 \le j \ne i \le m} r_{ij} \cdot w_{ij} \tag{14}$$

We must point out that the distance with respect to the k nearest neighbors of each class are used, that is, $k \cdot m$ neighbors are used and hence, taking $k = 1$ is not the same as using 1NN classifier, because a neighbor for each class is obtained. In order to understand the behavior of DRCW-WV, an illustrative example for its working procedure can be found in [15].

3.2. DRCW-TC: a novel approach for EFS based on truncating scores

When EFSs are embedded into a pairwise learning scheme, the scores obtained by means of the confidence degree are not well-suited for the WV aggregation [28]. Instead, they show a better behavior with both "Simple Voting", which truncates to 1.0 and 0.0 the maximum and minimum values for each position (i, j) and (j, i) of the score-matrix. This fact is predictably given by the fuzzy inference mechanism of this kind of rules. In particular, we must stress two points:

1. Both output classes of the binary classifiers are likely to have a positive activation degree, especially in areas of high overlapping.
2. Non-competent classifiers will erroneously obtain a high confidence due to the high coverage and generalization of rules with fuzzy labels.

Regarding the former issues, our hypothesis is that the behavior of EFS within an OVO learning approach can be improved following a simple but effective two-step methodology:

1. Truncating the confidence degrees to values $\{0.0, 1.0\}$ in the score matrix. In this sense, values r_{ij} will be updated as follows:
$$r_{ij} = \begin{cases} 1 & \text{if } r_{ij} > r_{ji} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$
The objective is to limit the influence of the non-competent classifiers at least in one of the two related classes.
2. Then, using the DRCW technique to carry out the combination from the "truncated" score matrix. In this way, the goodness shown by the Binary Voting model and the dynamic decision process are unified. The ultimate goal is to reduce as much as possible the non-competence problem.

In order to clarify the whole process, next we present a simple example for the use of DRCW-TC:

Suppose that an instance x, whose real class is known to be c_1 has to be classified. After submitting it to all base classifiers, the following score-matrix R is obtained:
$$R(x) = \begin{pmatrix} & c_1 & c_2 & c_3 & c_4 & c_5 \\ c_1 & - & 0.55 & 0.45 & 0.80 & 0.90 \\ c_2 & 0.45 & - & 0.55 & 1.00 & 0.80 \\ c_3 & 0.55 & 0.45 & - & 0.45 & 0.40 \\ c_4 & 0.20 & 0.00 & 0.55 & - & 0.10 \\ c_5 & 0.10 & 0.20 & 0.60 & 0.90 & - \end{pmatrix} \tag{16}$$

Either applying a "Simple Voting" or a "WV" strategy, class c_2 will be chosen instead of real class c_1 due to the high confidence degrees associated to the former class (see Eq. (17)). Therefore, the source of failure is related to the non-competent classifiers.

$$R(x) = \left(\begin{array}{c|ccccc|cc} & c_1 & c_2 & c_3 & c_4 & c_5 & SV & WV \\ c_1 & - & 0.55 & 0.45 & 0.80 & 0.90 & 3 & 2.70 \\ c_2 & 0.45 & - & 0.55 & 1.00 & 0.80 & 3 & \mathbf{2.80} \\ c_3 & 0.55 & 0.45 & - & 0.45 & 0.40 & 1 & 1.85 \\ c_4 & 0.20 & 0.00 & 0.55 & - & 0.10 & 1 & 0.85 \\ c_5 & 0.10 & 0.20 & 0.60 & 0.90 & - & 2 & 1.80 \end{array}\right) \tag{17}$$

In order to fix this issue, we try to apply the standard DRCW-WV method. For this task, the distances to the k nearest neighbors of each class (d) are computed, such as $d = (0.9, 0.9, 0.7, 1.2, 1.6)$. Based on these distances, the weight-matrix W is computed to represent all w_{ij} for $i, j = 1 \ldots, m$:

$$W(x) = \left(\begin{array}{c|ccccc} & c_1 & c_2 & c_3 & c_4 & c_5 \\ c_1 & - & 0.50 & 0.38 & 0.64 & 0.76 \\ c_2 & 0.50 & - & 0.38 & 0.64 & 0.76 \\ c_3 & 0.62 & 0.62 & - & 0.75 & 0.84 \\ c_4 & 0.36 & 0.36 & 0.25 & - & 0.64 \\ c_5 & 0.24 & 0.24 & 0.16 & 0.36 & - \end{array}\right) \tag{18}$$

When this weight-matrix W is applied to the initial score-matrix R, we obtain a modified score-matrix R^w (Eq. (19)). Then, the output class is computed using the WV method, but the initial confidences associated to class c_2 were low in contrast to that of c_1, and the classification cannot been corrected yet.

$$R^w(x) = \left(\begin{array}{c|ccccc|c} & c_1 & c_2 & c_3 & c_4 & c_5 & WV \\ c_1 & - & 0.28 & 0.17 & 0.51 & 0.68 & 1.64 \\ c_2 & 0.22 & - & 0.21 & 0.64 & 0.61 & \mathbf{1.68} \\ c_3 & 0.34 & 0.28 & - & 0.34 & 0.34 & 1.29 \\ c_4 & 0.07 & 0.00 & 0.14 & - & 0.06 & 0.28 \\ c_5 & 0.02 & 0.05 & 0.10 & 0.32 & - & 0.49 \end{array}\right) \tag{19}$$

To address this problem, we can apply our novel DRCW-TC procedure. In this way, the initial score matrix R is first truncated to R^{tc} as shown in (Eq. (20)):

$$R^{tc}(x) = \left(\begin{array}{c|ccccc} & c_1 & c_2 & c_3 & c_4 & c_5 \\ c_1 & - & 1.00 & 0.00 & 1.00 & 1.00 \\ c_2 & 0.00 & - & 1.00 & 1.00 & 1.00 \\ c_3 & 1.00 & 0.00 & - & 0.00 & 0.00 \\ c_4 & 0.00 & 0.00 & 1.00 & - & 0.00 \\ c_5 & 0.00 & 0.00 & 1.00 & 1.00 & - \end{array}\right) \tag{20}$$

The last step is to combine this new score matrix R^{tc} with the weight matrix W. Now, we can observe how the confidence degrees for class c_2 have been positively boosted by both the truncation and the distance tuning. This results into a correct classification for the query instance (see Eq. (21)).

$$R^{tc-w}(x) = \left(\begin{array}{c|ccccc|c} & c_1 & c_2 & c_3 & c_4 & c_5 & WV \\ c_1 & - & 0.50 & 0.00 & 0.64 & 0.76 & \mathbf{1.90} \\ c_2 & 0.00 & - & 0.38 & 0.64 & 0.76 & 1.78 \\ c_3 & 0.62 & 0.00 & - & 0.00 & 0.00 & 0.62 \\ c_4 & 0.00 & 0.00 & 0.25 & - & 0.00 & 0.25 \\ c_5 & 0.00 & 0.00 & 0.16 & 0.36 & - & 0.52 \end{array}\right) \tag{21}$$

4. Experimental framework

In this section we first provide details of the real-world multi-class problems chosen for the experiments, and the validation procedure (Subsection 4.1). Then, we will present the configuration parameters for the classification algorithms used in the study (Subsection 4.2). Finally, we present the statistical tests applied to compare the results obtained with the different classifiers (Subsection 4.3).

Table 1
Summary description of data-sets.

Data-set	id	#Ex.	#Atts.	#Num.	#Nom.	#Cl.
Balance	bal	625	4	4	0	3
Contraceptive	con	1473	9	9	0	3
Hayes-roth	hay	132	4	4	0	3
Iris	iri	150	4	4	0	3
New-thyroid	new	215	5	5	0	3
Tae	tae	151	5	5	0	3
Thyroid	thy	7200	21	21	0	3
Wine	win	178	13	13	0	3
Vehicle	veh	846	18	18	0	4
Cleveland	cle	297	13	13	0	5
Page-blocks	pag	5472	10	10	0	5
Shuttle	shu	58 000	9	9	0	5
Autos	aut	159	25	15	10	6
Glass	gla	214	9	9	0	7
Satimage	sat	6435	36	36	0	7
Segment	seg	2310	19	19	0	7
Ecoli	eco	336	7	7	0	8
Penbased	pen	10 992	16	16	0	10
Yeast	yea	1484	8	8	0	10
Texture	tex	5500	40	40	0	11
Vowel	vow	990	13	13	0	11
Wine-quality-red	wqr	1599	11	11	0	11
Wine-quality-white	wqw	4898	11	11	0	11

4.1. Benchmark data

We have carried out a selection of 23 different benchmark problems from KEEL repository [21], so that the same data partitions can be used by other researchers. The characteristics of these datasets are shown in Table 1, which is ordered with respect to their number of classes. They comprise a number of situations, from totally balanced data-sets to highly imbalanced ones, besides the different number of classes. We must stress that datasets include a low proportion of nominal attributes (only "autos" problem), so that this issue will not affect the inference model, thus having the highest advantage of the fuzzy partitions.[2]

Additionally, instead of the standard stratified cross-validation, we will use a recently published partitioning procedure called Distribution Optimally Balanced Cross Validation [57]. This allows at correcting the dataset shift [58–60] (when training and test data do not follow the same distribution), which may hinder the results obtained in the experimental analysis. This validation procedure is carried out using 10 folds, aiming at having a representative number of different cases that support a robust average performance value. Finally, we will consider the accuracy rate to evaluate the quality of the classifiers.

4.2. Algorithms selected and configuration parameters

In order to show the usefulness of our DRCW-TC approach to OVO combination in general, and for EFS in particular, we have selected several well-known Machine Learning algorithms as base learners. We must point out that all these algorithms are available within the KEEL software tool [45].

- **FARC-HD-Ov** – It has been selected as a robust and representative *EFS* [33]. As introduced in Section 2.2 the extension of the FARC-HD algorithm using overlap functions, noted as **FARC-HD-Ov**, has been selected. We must recall that this approach is better suited for pairwise learning than the original classifier, as suggested in [28].
- **C4.5** – This *decision tree* classifier [61] must be regarded as the state-of-the-art for rule learning algorithms.
- **SVM** – *Support Vector Machines* [42] has been shown to be one of the most precise classifiers. In this study we use SMO [62] algorithm to train the SVM base classifiers.

The selected parameters for these learning algorithms are detailed in Table 2. There, we also show the aggregation schemes that will be used as basis for the OVO approach on each classifier (as was stressed in Section 2.4).

The majority of the combination methods in OVO make use of the confidence degrees of the outputs of each base classifier. These confidence degrees are obtained for each classifier as follows:

- **FARC-HD-Ov –** Soundness degree for each class, as computed by Eq. (4).
- **SVM –** Probability estimates obtained by the SVM logistic model [63].

[2] In this case we will use as fuzzy singletons for different values of the nominal attributes.

Table 2
Parameter specification for the base learners employed in the experimentation.

Algorithm	Parameters

FARC-HD-Ov	# Labels = 5, Aggregation operator = Harmonic mean
	Inference = Additive combination. Min. support = 0.05,
	Min. confidence = 0.8, Max. tree depth = 3, k for pre-screening = 2
	OVO Aggregation = VOTE

C4.5	Prune = True, Confidence level = 0.25 Minimum number of item-sets per leaf = 2 **OVO Aggregation** = WV
SVM	C = 100.0, Tolerance Parameter = 0.001, Epsilon = 1.0E−12 Kernel Type = RBFN, RBFNKernel $\gamma = 0.01$ Fit Logistic Models = True Attribute normalization = Standard $\in [0, 1]$ **OVO Aggregation** = PE

- **C4.5 –** Accuracy of the leaf making the prediction, i.e., correctly classified train examples divided by the total number of covered train instances in that leaf.

In some of the combination strategies ties might occur. As usual, in those cases the majority class is predicted. If the tie continues, the class is selected randomly.

Finally, as distance metric for the DRCW procedure we make use of the Heterogeneous Value Difference Metric (HVDM) [64], as it can be applied for both continuous and nominal variables.

4.3. Statistical tests for performance comparison

In this paper we use the hypothesis testing techniques to provide statistical support for the analysis of the results [24]. Specifically, we will use non-parametric tests, due to the fact that the initial conditions that guarantee the reliability of the parametric tests may not be satisfied, causing the statistical analysis to lose credibility with these types of tests [65,22]. Any interested reader can find additional information on the Website http://sci2s.ugr.es/sicidm/.

First of all, we consider the method of aligned ranks of the algorithms in order to show at a first glance how good a method is with respect to its partners. In order to compute this ranking, the first step is to obtain the average performance of the algorithms in each dataset. Next, we compute the subtractions between the accuracy of each algorithm minus the average value for each dataset, which is computed regarding the output results for all algorithms considered in the comparison. Then, we rank all these differences in a descending way and, finally, we average the rankings obtained by each algorithm. In this manner, the algorithm which achieves the **lowest average ranking** is the best one.

The Friedman Aligned test [24] will be used to check whether there are significant differences among the results, and the Holm post-hoc test [66] in order to find which algorithms reject the hypothesis of equality with respect to a selected control method in a $1 \cdot n$ comparison. We will compute the adjusted p-value (APV) associated with each comparison, which represents the lowest level of significance of a hypothesis that results in a rejection. This value differs from the standard p-value in the sense that it determines univocally whether the null hypothesis of equality is rejected at a significance level α.

Regarding pairwise comparisons, we will make use of Wilcoxon signed-rank test [67] to find out whether significant differences exist between a pair of algorithms. This procedure computes the differences between the performance scores of the two classifiers on each one of the available datasets (N_{ds}). The differences are ranked according to their absolute values, from smallest to largest, and average ranks are assigned in case of ties. We call R^+ the sum of ranks for the datasets on which the second algorithm outperformed the first, and R^- the sum of ranks for the opposite. Let T be the smallest of the sums, $T = min(R^+, R^-)$. If T is less than or equal to the value of the distribution of Wilcoxon for N_{ds} degrees of freedom (Table B.12 in [68]), the null hypothesis of equality of means is rejected.

5. Experimental study

This section contains the experimental analysis which shows the robustness of our new DRCW-TC approach. In order to do so, we divide our study into two different parts:

1. First, we carry out a intra-family comparison between DRCW-TC and the baseline methodologies, including the standard classifier for multiple-class (EFS and C4.5), the pairwise learning approach, and DRCW-WV [15] (Subsection 5.1).
2. Next, we perform a inter-family comparison to contrast the results among the different classifiers. Our aim is to excel the better synergy of DRCW-TC in combination with the EFS model with respect to the classifiers selected from the state-of-the-art (Subsection 5.2).

5.1. Analysis of DRCW-TC for pairwise learning

As previously pointed out, in this part of the study our objective is to show the superior performance achieved by our proposed DRCW-TC in EFSs in particular, as well as a robust behavior when applied in synergy with different classifiers. Regarding this fact, we must point out that, despite this methodology was initially designed for EFS, we aim at determining whether it is also suitable for different paradigms of classifiers. Specifically, we will contrast whether the improvement with respect to baseline models are greater in the case of DRCW-WV or DRCW-TC. In this way, Table 3 shows the complete experimental results for the three different classifiers, namely FARC-HD-Ov, C4.5 and SVM.

Observing these results in detail, we stress that in all three cases DRCW-TC achieves the highest average result overall. Just focusing on the level of absolute performance, the achieved values may seem close between DRCW-WV and DRCW-TC. For this reason, we complement the analysis by carrying out statistical tests, which are shown in Table 4. Additionally, we carry out a Wilcoxon test (Table 5) for contrasting directly DRCW-WV and DRCW-TC for addressing the non-competence.

Focusing on the global comparison among all models, the quality of the proposed DRCW-TC stands out, as it achieves the highest ranking in all case studies. Specifically, we observe that significant differences are obtained with respect to both the baseline algorithm and the OVO approach. Finally, regarding the output given by the Wilcoxon test, both for FARC-HD-Ov and C4.5 there is a better behavior, both in the ranking and the number of "wins". For SVM both approaches have a similar ranking. This analysis stress the significance of the truncated approach for EFS rather than for crisp approaches. We have shown that DRCW-TC has a good behavior in synergy with rule-based classifiers. In this type of algorithms, the non-competence problem is supposed to appear with a higher frequency in contrast to that of SVMs, due to the higher generalization of the generated rules. Additionally, we have shown that the most notorious improvement is associated to the case study of the EFS.

5.2. On the positive synergy of DRCW-TC and EFSs

In the first part of our analysis, we have emphasized the good properties of DRCW-TC for enhancing the classification abilities of different paradigms of classifiers, since in all cases a significant improvement in the performance values is shown. In this section, we aim to excel the positive synergy between our proposal and EFSs in particular. We must point out that, since the truncation of the confidence degrees in the score-matrix is well-suited for this type of classifiers [28], then it must obtain the highest benefit from the application of this methodology. In this sense, we suggest that the use of DRCW-TC will allow EFSs to be the best performing approach overall.

In order to study our previous hypothesis, we show in Table 6 the average results of the selected models. Specifically, we include the results for the baseline classifier, the OVO approach and DRCW-TC for the three selected classifiers, i.e. FARC-HD-Ov, C4.5 and the SVM.

From these performance results, we can conclude the high level of robustness achieved by the combination between FARC-HD-Ov and DRCW-TC. First, because it obtains the highest average value overall. Additionally, and referring to Table 3 from the previous section, this approach shows a better behavior with respect to the models of comparison as the number of classes in the dataset increases. Finally, contrasting the results versus the original DRCW-WV methodology, we may observe that it allows an enhancement in both the training and test partitions. Specifically, in the main case study with the EFS, the number of datasets in which our DRCW-TC obtains a superior performance is more than two thirds of the total.

In order to complement the analysis, the average ranks (computed by means of the Friedman aligned method) and a Holm post-hoc test are shown in Table 7.

In accordance with the quality of the experimental results, together with the support given by the statistical test, we can determine the goodness of the DRCW-TC model in conjunction with FARC-HD-Ov, as it has excelled as the classifier with the best ranking, showing significant differences versus the baseline OVO methods (C4.5 and SVM). Regarding the comparison with the state-of-the-art classifiers using DRCW-TC, we observe that the EFS achieves a competitive performance, beating both schemes with respect to both average value of accuracy and ranking. Finally, to carry out a detailed study, Table 8 shows a statistical analysis among these three algorithms.

As a result of the former analysis, we have been able to determine a superiority for the DRCW-TC model in synergy with the EFS method FARC-HD-Ov. When the new DRCW-TC aggregation scheme is used as an extension to the OVO learning, the fuzzy classifier reaches a leap of quality that allows it to be competitive over traditionally more robust classifiers, such as SVMs.

6. Concluding remarks

In this work we have proposed a novel approach to address the non-competence problem in a pairwise learning environment. This methodology is based on DRCW, which weights the confidence degrees in the score matrix depending on the distance of the input example to each of the classes of the problem. Our new aggregation scheme, named as DRCW-TC, modifies this process by truncating the scores to 1.0 and 0.0 prior to the weighting step. Our hypothesis in this case is that this approach was especially suggested for EFS due to the properties related to the confidence degrees computed by this paradigm of classifiers.

Table 3
Complete experimental results in training and test with the standard accuracy metric. From the leftmost to the rightmost column we show the results for the EFS (FARC-HD-Ov), the decision tree (C4.5) and the SVM. Results are grouped into three parts, i.e. the standard classification algorithm (Baseline), the pairwise learning approach (OVO), and our proposed method (DRCW-TC). The highest average value per dataset is stressed in boldface. The highest average per classifier is underlined.

Dataset	#Cl	FARC-HD-Ov								C4.5								SVM					
		Baseline		OVO		DRCW-WV		DRCW-TC		Baseline		OVO		DRCW-WV		DRCW-TC		OVO		DRCW-WV		DRCW-TC	
		Tr	Tst	Tr	Tst	Tr	Tst	Tr	Tst	Tr	Tst	Tr	Tst	Tr	Tst	Tr	Tst	Tr	Tst	Tr	Tst	Tr	Tst
bal	3	92.12	88.81	92.07	86.57	89.17	86.88	92.05	87.68	89.80	78.09	46.22	45.61	80.41	75.53	82.03	76.80	91.68	91.70	88.20	87.06	92.20	**92.18**
con	3	62.53	53.15	66.36	**54.37**	56.02	52.48	64.33	53.69	73.44	53.08	69.78	53.70	64.23	53.70	67.37	54.10	53.23	51.72	51.76	51.25	53.21	51.86
hay	3	91.60	77.97	90.56	76.20	85.83	79.27	90.49	78.52	89.10	**82.33**	89.10	**82.33**	88.68	**82.33**	89.03	82.33	54.66	53.13	57.58	56.19	54.73	53.13
iri	3	99.41	96.00	99.41	95.33	96.00	95.33	99.41	95.33	98.07	94.67	98.07	94.67	98.00	94.67	98.07	94.67	97.63	**96.67**	97.11	94.67	97.63	**96.67**
new	3	99.54	96.26	99.95	96.69	97.16	**97.66**	99.95	97.16	98.45	94.39	98.86	95.35	98.76	95.35	98.86	95.80	97.16	96.71	97.11	96.73	97.16	96.71
tae	3	77.78	56.82	82.42	**62.72**	60.72	56.95	74.17	57.85	83.59	55.30	70.64	53.51	61.43	51.42	62.17	51.87	54.81	53.19	50.10	50.63	51.65	49.75
thy	3	93.13	93.06	93.18	93.10	93.35	93.33	93.56	93.53	99.88	99.56	99.93	**99.58**	99.80	99.49	99.82	99.50	97.07	96.96	96.90	96.82	97.00	96.97
win	3	99.75	92.15	99.94	93.84	99.06	96.20	99.94	96.07	98.88	93.29	99.25	94.50	99.44	95.61	99.25	95.61	99.75	**97.74**	99.75	**97.74**	99.75	**97.74**
veh	4	82.60	71.03	88.11	71.16	78.92	74.44	88.06	72.69	90.87	71.83	82.25	71.39	80.95	71.70	82.92	72.56	79.41	76.57	76.49	75.03	79.62	**77.28**
cle	5	87.17	**59.62**	93.94	57.14	77.21	55.27	88.03	54.64	83.81	52.55	82.75	53.17	77.70	51.17	78.44	50.86	68.01	56.45	64.72	56.58	65.99	56.11
pag	5	96.13	95.43	96.66	96.04	95.46	95.43	97.16	96.60	98.55	96.95	98.42	97.06	98.09	**97.31**	98.16	97.22	94.75	94.68	96.53	96.53	95.99	96.05
shu	5	99.58	99.58	99.69	99.68	99.21	99.21	99.88	99.87	99.99	99.97	100.00	**99.98**	99.98	99.97	99.98	99.97	97.31	97.30	99.78	99.79	99.78	99.78
aut	6	98.61	81.32	100.00	79.50	89.37	82.73	93.57	82.80	92.86	81.15	88.54	78.72	90.78	**87.85**	90.71	87.22	96.79	79.30	92.32	83.21	91.69	83.98
gla	7	84.16	70.68	88.37	71.99	78.02	73.35	85.98	70.95	93.46	66.42	91.33	72.78	88.58	**74.80**	88.58	74.78	66.92	62.23	76.11	73.89	73.52	71.44
sat	7	64.29	63.81	88.88	86.83	88.51	88.49	89.78	87.99	97.62	86.75	97.87	86.90	97.97	89.17	97.98	89.02	86.07	85.89	90.59	**90.52**	87.75	87.77
seg	7	94.43	92.73	97.88	95.67	97.03	96.45	98.21	96.84	99.25	97.10	99.13	97.40	99.17	98.18	99.24	**98.35**	93.82	93.77	96.61	96.54	95.60	95.58
eco	8	92.33	82.16	95.64	82.94	85.68	82.27	91.70	**84.06**	91.26	79.69	74.44	71.67	82.98	81.33	83.48	80.72	82.80	79.37	82.77	82.04	83.63	82.58
pen	10	97.29	96.42	99.53	97.86	99.19	99.08	99.66	99.18	99.28	96.46	99.40	97.00	99.69	99.13	99.68	99.10	99.21	98.96	99.55	99.50	**99.50**	99.45
yea	10	64.14	59.22	70.10	59.36	59.34	59.35	65.38	60.15	80.88	55.80	70.58	58.34	64.85	57.85	66.35	59.12	60.83	59.90	58.76	59.74	59.92	**60.22**
tex	11	90.66	89.71	98.43	95.36	98.41	98.24	99.12	98.11	99.03	93.09	99.39	94.67	99.65	98.09	99.65	98.02	99.55	99.27	99.64	99.49	99.65	**99.51**
vow	11	73.83	66.36	99.70	91.72	98.89	98.28	99.71	**99.09**	96.89	80.10	97.46	82.63	99.51	97.78	99.44	97.47	82.05	77.98	99.03	98.69	98.53	97.88
wqr	11	65.60	61.15	69.68	61.47	62.11	60.40	69.05	**62.97**	88.80	62.11	61.09	47.69	53.58	45.15	55.13	46.97	61.84	60.09	61.97	60.90	63.39	62.15
wqw	11	55.38	53.38	58.52	55.03	56.54	56.13	59.91	57.13	90.02	**58.85**	72.88	56.50	69.59	57.93	70.97	58.10	54.48	53.69	56.87	56.42	57.06	56.71
avg	–	85.31	78.12	89.96	80.89	84.40	81.62	88.66	**81.87**	92.77	79.54	86.41	77.61	86.69	80.67	87.27	80.88	81.30	78.84	82.18	80.87	82.39	80.93

Table 4
Average ranks (Friedman aligned) and APVs (Holm test) for all classifiers. Control method is pointed out with asterisks. Symbol * implies significant differences at 95%, whereas symbol + sets the confidence degree at 90%.

Method	FARC-HD-Ov		C4.5		SVM	
	Rank (position)	APV (Holm)	Rank (position)	APV (Holm)	Rank (position)	APV (Holm)
Baseline	65.6087 (4)	.000056*	55.2609 (4)	.002966*	–	–
OVO	48.1957 (3)	.076782+	59.7174 (3)	.012754*	46.8478 (3)	.001526*
DRCW-WV	40.3043 (2)	.285314	37.2391 (2)	.660677	31.2174 (2)	.469131
DRCW-TC	31.8913 (1)	******	33.7826 (1)	*******	26.9348 (1)	********

Table 5
Wilcoxon test to compare DRCW-TC [R^+] and the standard DRCW-WV [R^-]. Symbol * implies significant differences at 95%, whereas symbol + sets the confidence degree at 90%. Wins/Ties/Losses values are computed with respect to the novel DRCW-TC approach.

Classifier	DRCW-TC (R^+)	DRCW-WV (R^-)	p-value	W/T/L
FARC-HD-Ov	185.0	91.0	.148539	15/0/8
C4.5	167.0	109.0	.36959	14/0/9
SVM	127.0	126.0	.974101	12/1/12

Table 6
Average experimental results in training and test with the accuracy metric. From the leftmost to the rightmost column we show the results for the standard classification algorithm (Baseline), the pairwise learning approach (OVO), the standard weighting procedure for OVO (DRCW-WV), and our proposed method (DRCW-TC). The highest average value per classifier is stressed in boldface.

Algorithm	Baseline		OVO		DRCW-WV		DRCW-TC	
	Tr	Tst	Tr	Tst	Tr	Tst	Tr	Tst
FARC-HD-Ov	85.31	78.12	88.65	80.37	84.40	81.62	88.66	**81.87**
C4.5	92.77	79.54	86.41	77.61	86.69	80.67	87.27	**80.88**
SVM	–	–	81.3	78.84	82.18	80.87	82.39	**80.93**

Table 7
Average ranks (Friedman aligned) and APVs (Holm test) for the analysis of all methodologies. Control method is pointed out with asterisks. Symbol * implies significant differences at 95%, whereas symbol + sets the confidence degree at 90%.

Method	Rank (position)	APV (Holm)
FARC-HD-Ov	107.6522 (6)	.056023+
FARC-HD-Ov-OVO	89.1304 (4)	.486336
FARC-HD-OV-DRCW-TC	67.1739 (1)	*******
C4.5	108.000 (7)	.056023+
C4.5-OVO	112.000 (8)	.030207*
C4.5-DRCW-TC	81.5652 (3)	.719003
SVM-OVO	103.8261 (5)	.078446+
SVM-DRCW-TC	70.6522 (2)	.824730

Table 8
Average ranks (Friedman aligned) and APVs (Holm test) for DRCW-TC. Control method is pointed out with asterisks.

Method	Rank (position)	APV (Holm)
FARC-HD-Ov-DRCW-TC	32.3043 (1)	********
C4.5-DRCW-TC	38.2609 (3)	.628027
SVM-DRCW-TC	34.4348 (2)	.718765

From this study, we have observed that the DRCW-TC model have shown a significant robustness over different classification algorithms, namely C4.5 and SVM, outperforming the initial results of the simple OVO scheme. In all case studies with different baseline classifiers, the comparison between the standard DRCW-WV model and our new DRCW-TC was in favor of the latter scheme, implying a higher support for the current proposal. Another important issue in the application of the new DRCW-TC, is that for those datasets with a higher number of classes, the performance of FARC-HD (with overlap functions) grows over the one achieved by applying the standard DRCW-WV model.

Finally, after the analysis carried out, we can determine the good synergy existing between EFS and the DRCW-TC procedure. Indeed, the combination between both models have allowed to reach higher quality levels in classification tasks, in contrast with more precise state-of-the-art classifiers such as C4.5 and SVM. In summary, our proposal allows a significant leap in the results over the base FARC-HD model and its OVO extension.

Acknowledgements

This work was partially supported by the Spanish Ministry of Science and Technology under projects TIN2011-28488, TIN-2012-33856, TIN2013-40765-P; the Andalusian Research Plans P12-TIC-2958, P11-TIC-7765 and P10-TIC-6858; and both the University of Jaén and Caja Rural Provincial de Jaén under project UJA2014/06/15.

References

[1] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, 3rd Edition, Morgan Kaufmann, San Mateo, CA, USA, 2011.
[2] A.C. Lorena, A.C. Carvalho, J.M. Gama, A review on the combination of binary classifiers in multiclass problems, Artif. Intell. Rev. 30 (1–4) (2008) 19–37.
[3] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, F. Herrera, A review on ensembles for class imbalance problem: bagging, boosting and hybrid based approaches, IEEE Trans. Syst. Man Cybern., Part C, Appl. Rev. 42 (4) (2012) 463–484.
[4] E.L. Allwein, R.E. Schapire, Y. Singer, Reducing multiclass to binary: a unifying approach for margin classifiers, J. Mach. Learn. Res. 1 (2000) 113–141.
[5] T.G. Dietterich, G. Bakiri, Solving multiclass learning problems via error-correcting output codes, J. Artif. Intell. Res. 2 (1995) 263–286.
[6] S. Knerr, L. Personnaz, G. Dreyfus, Single-layer learning revisited: a stepwise procedure for building and training a neural network, in: F. Fogelman Soulié, J. Hérault (Eds.), Neurocomputing: Algorithms, Architectures and Applications, in: NATO ASI Series, vol. F68, Springer-Verlag, 1990, pp. 41–50.
[7] T. Hastie, R. Tibshirani, Classification by pairwise coupling, Ann. Stat. 26 (2) (1998) 451–471.
[8] J. Fürnkranz, Round Robin classification, J. Mach. Learn. Res. 2 (2002) 721–747.
[9] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, F. Herrera, An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes, Pattern Recognit. 44 (8) (2011) 1761–1776.
[10] M. Galar, A. Fernandez, E. Barrenechea, F. Herrera, Empowering difficult classes with a similarity-based aggregation in multi-class classification problems, Inf. Sci. 264 (2014) 135–157.
[11] L.I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, 1st edition, Wiley-Interscience, 2004.
[12] M. Wozniak, M. Graña, E. Corchado, A survey of multiple classifier systems as hybrid systems, Inf. Fusion 16 (2014) 3–17.
[13] E. Hüllermeier, S. Vanderlooy, Combining predictions in pairwise classification: an optimal adaptive voting strategy and its relation to weighted voting, Pattern Recognit. 43 (1) (2010) 128–142.
[14] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, F. Herrera, Dynamic classifier selection for one-vs-one strategy: avoiding non-competent classifiers, Pattern Recognit. 46 (12) (2013) 3412–3424.
[15] M. Galar, A. Fernandez, E. Barrenechea, F. Herrera, DRCW-OVO: distance-based relative competence weighting combination for one-vs-one strategy in multi-class problems, Pattern Recognit. 48 (1) (2015) 28–42.
[16] H. Ishibuchi, T. Nakashima, M. Nii, Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining, Springer–Verlag, 2004.
[17] M. Gacto, R. Alcala, F. Herrera, Interpretability of linguistic fuzzy rule-based systems: an overview of interpretability measures, Inf. Sci. 181 (20) (2011) 4340–4360.
[18] A. Fernandez, V. Lopez, M.J. del Jesus, F. Herrera, Revisiting evolutionary fuzzy systems: taxonomy, applications, new trends and challenges, Knowl.-Based Syst. 80 (2015) 109–121.
[19] S. Alshomrani, A. Bawakid, S.-O. Shim, A. Fernandez, F. Herrera, A proposal for evolutionary fuzzy systems using feature weighting: dealing with overlapping in imbalanced datasets, Knowl.-Based Syst. 73 (2015) 1–17.
[20] J.A. Sanz, A. Fernandez, H. Bustince, F. Herrera, IVTURS: a linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection, IEEE Trans. Fuzzy Syst. 21 (3) (2013) 399–411.
[21] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera, KEEL data–mining software tool: data set repository, integration of algorithms and experimental analysis framework, J. Mult.-Valued Log. Soft Comput. 17 (2–3) (2011) 255–287.
[22] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.
[23] S. Garcia, F. Herrera, An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons, J. Mach. Learn. Res. 9 (2008) 2677–2694.
[24] S. García, A. Fernández, J. Luengo, F. Herrera, Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power, Inf. Sci. 180 (10) (2010) 2044–2064.
[25] H. Ishibuchi, K. Nozaki, H. Tanaka, Distributed representation of fuzzy rules and its application to pattern classification, Fuzzy Sets Syst. 52 (1) (1992) 21–32.
[26] O. Cordon, M.J. del Jesus, F. Herrera, A proposal on reasoning methods in fuzzy rule-based classification systems, Int. J. Approx. Reason. 20 (1) (1999) 21–45.
[27] H. Ishibuchi, T. Yamamoto, Rule weight specification in fuzzy rule-based classification systems, IEEE Trans. Fuzzy Syst. 13 (2005) 428–435.
[28] M. Elkano, M. Galar, J. Sanz, A. Fernandez, E. Barrenechea, F. Herrera, H. Bustince, Enhancing multi-class classification in FARC-HD fuzzy classifier: on the synergy between n-dimensional overlap functions and decomposition strategies, IEEE Trans. Fuzzy Syst. 23 (5) (2015) 1562–1580.
[29] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, L. Magdalena, Ten years of genetic fuzzy systems: current framework and new trends, Fuzzy Sets Syst. 141 (1) (2004) 5–31.
[30] O. Cordon, F. Herrera, F. Hoffmann, L. Magdalena, Genetic Fuzzy Systems. Evolutionary Tuning and Learning of Fuzzy Knowledge Bases, World Scientific, Singapore, Republic of Singapore, 2001.
[31] O. Cordon, A historical review of evolutionary learning methods for mamdani-type fuzzy rule-based systems: designing interpretable genetic fuzzy systems, Int. J. Approx. Reason. 52 (6) (2011) 894–913.
[32] F. Herrera, Genetic fuzzy systems: taxonomy, current research trends and prospects, Evol. Intel. 1 (2008) 27–46.
[33] J. Alcalá-Fdez, R. Alcalá, F. Herrera, A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning, IEEE Trans. Fuzzy Syst. 19 (5) (2011) 857–872.
[34] J. Han, M. Kamber, Data Mining. Concepts and Techniques, 2nd edition, Morgan Kaufmann, 2006.
[35] C. Zhang, S. Zhang, Association Rule Mining, Models and Algorithms, Lecture Notes in Computer Science, vol. 2307, Springer, 2002.
[36] B. Kavsek, N. Lavrac, Apriori-sd: adapting association rule learning to subgroup discovery, Appl. Artif. Intell. 20 (7) (2006) 543–583.
[37] J. Casillas, O. Cordon, M.J. del Jesus, F. Herrera, Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction, IEEE Trans. Fuzzy Syst. 13 (1) (2005) 13–29.
[38] R. Alcala, J. Alcala-Fdez, F. Herrera, A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection, IEEE Trans. Fuzzy Syst. 15 (4) (2007) 616–635.
[39] H. Bustince, J. Fernandez, R. Mesiar, J. Montero, R. Orduna, Overlap functions, Nonlinear Anal.: Theory, Methods Appl. 72 (3–4) (2010) 1488–1499.
[40] D. Paternain, J. Fernández, H.B. Sola, R. Mesiar, G. Beliakov, Construction of image reduction operators using averaging aggregation functions, Fuzzy Sets Syst. 261 (2015) 87–111.

[41] A.C. Lorena, A.C. Carvalho, J.M. Gama, A review on the combination of binary classifiers in multiclass problems, Artif. Intell. Rev. 30 (1–4) (2008) 19–37.
[42] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[43] C.W. Hsu, C.J. Lin, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Netw. 13 (2) (2002) 415–425.
[44] R. Rifkin, A. Klautau, In defense of one-vs-all classification, J. Mach. Learn. Res. 5 (2004) 101–141.
[45] J. Alcalá-Fdez, L. Sánchez, S. García, M.J. del Jesus, S. Ventura, J.M. Garrell, J. Otero, C. Romero, J. Bacardit, V.M. Rivas, J.C. Fernández, F. Herrera, KEEL: a software tool to assess evolutionary algorithms for data mining problems, Soft Comput. 13 (2009) 307–318.
[46] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The weka data mining software: an update, SIGKDD Explor. Newsl. 11 (2009) 10–18.
[47] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (2011) 27, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[48] J. Fürnkranz, Round Robin classification, J. Mach. Learn. Res. 2 (2002) 721–747.
[49] J.A. Sáez, M. Galar, J. Luengo, F. Herrera, Analyzing the presence of noise in multi-class problems: alleviating its influence with the one-vs-one decomposition, Knowl. Inf. Syst. 38 (1) (2014) 179–206.
[50] A. Fernandez, M. Calderon, E. Barrenechea, H. Bustince, F. Herrera, Solving multi-class problems with linguistic fuzzy rule based classification systems based on pairwise learning and preference relations, Fuzzy Sets Syst. 161 (23) (2010) 3064–3080.
[51] S. Elhag, A. Fernandez, A. Bawakid, S. Alshomrani, F. Herrera, On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on intrusion detection systems, Expert Syst. Appl. 42 (1) (2015) 193–202.
[52] E. Hüllermeier, S. Vanderlooy, Combining predictions in pairwise classification: an optimal adaptive voting strategy and its relation to weighted voting, Pattern Recognit. 43 (1) (2010) 128–142.
[53] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, F. Herrera, An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes, Pattern Recognit. 44 (8) (2011) 1761–1776.
[54] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, F. Herrera, Dynamic classifier selection for one-vs-one strategy: avoiding non-competent classifiers, Pattern Recognit. 46 (12) (2013) 3412–3424.
[55] J.H. Friedman, Another approach to polychotomous classification, Tech. rep., Department of Statistics, Stanford University, 1996, http://www-stat.stanford.edu/~jhf/ftp/poly.ps.Z.
[56] T.F. Wu, C.J. Lin, R.C. Weng, Probability estimates for multi-class classification by pairwise coupling, J. Mach. Learn. Res. 5 (2004) 975–1005.
[57] J.G. Moreno-Torres, J.A. Sáez, F. Herrera, Study on the impact of partition-induced dataset shift on k-fold cross-validation, IEEE Trans. Neural Netw. Learn. Syst. 23 (8) (2012) 1304–1313.
[58] J.G. Moreno-Torres, T. Raeder, R. Aláiz-Rodríguez, N.V. Chawla, F. Herrera, A unifying view on dataset shift in classification, Pattern Recognit. 45 (1) (2012) 521–530.
[59] J.Q. Candela, M. Sugiyama, A. Schwaighofer, N.D. Lawrence, Dataset Shift in Machine Learning, The MIT Press, 2009.
[60] V. Lopez, A. Fernandez, F. Herrera, On the importance of the validation technique for classification with imbalanced datasets: addressing covariate shift when data is skewed, Inf. Sci. 257 (2014) 1–13.
[61] J.R. Quinlan, C4.5: Programs for Machine Learning, 1st edition, Morgan Kaufmann Publishers, San Mateo–California, 1993.
[62] J.C. Platt, Fast Training of Support Vector Machines Using Sequential Minimal Optimization, MIT Press, Cambridge, MA, USA, 1999.
[63] J.C. Platt, Probabilistic outputs for support vector machines and comparison to regularized likelihood methods, in: A. Smola, P. Bartlett, B. Schölkopf, D. Schuurmans (Eds.), Advances in Large Margin Classifiers, Cambridge, MA, 2000.
[64] D. Wilson, T. Martinez, Improved heterogeneous distance functions, J. Artif. Intell. Res. 6 (1997) 1–34.
[65] S. García, F. Herrera, An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons, J. Mach. Learn. Res. 9 (2008) 2607–2624.
[66] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. 6 (1979) 65–70.
[67] F. Wilcoxon, Individual comparisons by ranking methods, Biom. Bull. 1 (6) (1945) 80–83.
[68] J.H. Zar, Biostatistical Analysis, Prentice Hall, Upper Saddle River, New Jersey, 1999.