Enhancing Fuzzy Rule Based Systems in Multi-Classification Using Pairwise Coupling with Preference Relations

Alberto Fernández
Dept. of Computer
Science and Artificial Intelligence,
University of Granada, Spain
alberto@decsai.ugr.es

María Calderón
Edurne Barrenechea
Humberto Bustince
Dept. of Automatic
and Computation,
Public University of Navarra, Spain
Maria.CalderonTurullols@cajanavarra.es
{edurne.barrenechea,bustince}@unavarra.es

Francisco Herrera
Dept. of Computer
Science and Artificial Intelligence,
University of Granada, Spain
herrera@decsai.ugr.es

Abstract

This contribution proposes a technique for Fuzzy Rule Based Classification Systems (FRBCSs) based on a multi-classifier approach using fuzzy preference relations for dealing with multi-class classification. The idea is to decompose the original data-set into binary classification problems using a pairwise coupling approach (confronting all pair of classes), and to obtain a fuzzy system for each one of them. Along the inference process, each FRBCS generates an association degree for its two classes, and these values are encoded into a fuzzy preference relation. The final class of the whole FRBCS will be obtained by decision making following a non-dominance criterium.

We show the goodness of our proposal in contrast with the base fuzzy model with an extensive experimental study following a statistical study for analysing the differences in performance among the algorithms. We will also contrast our results versus the well-known C4.5 decision tree.

Keywords:

Fuzzy Rule-Based Classification Systems, Multi-class Problems, Pairwise Coupling, Fuzzy Preference Relations, Multi-classifiers

1 Introduction

Fuzzy Rule Based Classification Systems (FRBCSs) [10] are a popular tool in Machine Learning because of their advantage on the use of linguistic terms to increase the global interpretability of the output model.

In this work, we focus our attention on the multi-class problem, since this type of data is very common in real-world applications and may represent an additional difficulty for the fuzzy classifier to find the separability in an search space with a high number of classes. In order to simplify the original problem, we propose a methodology based on pairwise coupling (all-pairs) [3, 6], transforming an m-class problem into $m(m-1)/2$ binary problems, one for each pair of classes, and learning a fuzzy classifier for each separate problem. In order to manage the ensemble of classifiers, we will apply a fuzzy preference relation model [13] to determine the output among all predictions of the associated classifiers. This connection between classification learning and fuzzy preference modeling was established by Hüllermeier and Brinker in [8], and was later employed by Hühn and Hüllermeier on a Fuzzy Round Robin Ripper approach based on a decision rule using fuzzy preference relations [7].

We will analyse the effect of this multi-classifier approach over a well-performing FRBCSs, the Fuzzy Hybrid Genetics-Based Machine Learning (FH-GBML) algorithm [12], in order to determine whether our proposal enhances the behaviour of the corresponding base fuzzy model. Furthermore, we will include in the comparative study the C4.5 decision tree [14], as a well-known state-of-the-art rule-based algorithm.

We have selected 16 multi-class data-sets from UCI repository [1] within the experimental framework. The measure of performance is based on accuracy rate and the significance of results is supported by the proper statistical analysis as suggested in the literature [2, 5].

To do so, this work is organised as follows. In Section 2 we present the concept of multi-classification and we introduce our pairwise coupling proposal using fuzzy preference relations, also describing the FH-GBML algorithm selected for our study. Next, Section 3 includes the experimental framework, that is, the benchmark data-sets, parameters, and the statistical tests for the performance comparison. In Section 4 we present our empirical analysis. Finally, Section 5 concludes the paper summarizing all the lessons learned.

2 Fuzzy Rule Based Multi-Classifier Systems

In this work, we aim to improve the performance of FRBCSs introducing a multi-classifier proposal based on pairwise coupling [3, 6]. This methodology enables a better exploration of the domain of the problem by dividing the original data-set into binary sub-problems that are easier to discriminate.

In this section we will first introduce the concept of multi-classification and then, we will present our learning proposal for a fuzzy multi-classifier system using pairwise coupling. Next, we will define the procedure used to obtain the final classification output among all classifiers by means of fuzzy preference relations. Finally, we will describe the FH-GBML algorithm, which will be employed as the base fuzzy model.

2.1 Learning Proposal for a Fuzzy Multi-classifier System

There are a high amount of applications which require multi-class categorization, namely, objects identification, image processing and handwritten digit recognition among others. In order to simplify the boundaries of those type of problems we can proceed by dividing the initial problem into multiple two-class sets that can be solved separately.

We will employ the pairwise coupling approach [3, 6], dividing the original training set into $m(m-1)/2$ subsets, where m stands for the number of classes of the problem, in order to obtain $m(m-1)/2$ different fuzzy classifiers. Every subset contains the examples for a different pair of classes and thus, the trained classifiers are devoted to discriminate between two specific classes of the initial data-set. At classification time, a query instance is submitted to all binary models, and the predictions of these models are combined into an overall classification.

Each one of these fuzzy classifiers will have its own Knowledge Base (KB), composed by a local Data Base (DB) and Rule Base (RB). We decided to focus on the precision of the model by contextualizing the fuzzy partitions for each sub-problem separately, taking as universe of discourse the range of the values for the variables of the subset of examples selected for each training subset.

The RB for each classifier is learnt using an ad-hoc procedure, which can be selected among the different approaches of the specialized literature, making our proposal independent of the fuzzy rule learning method used.

Once all KBs have been learnt, we proceed to the final inference step. When a new input pattern arrives to the system, each FRBCS is fired in order to define the output degree for its pair of associated classes.

The next part of this section is devoted to describe how these association degrees are combined in order to provide the final class label.

2.2 Non-Dominance Classification Process with Fuzzy Preference Relations

We will consider the classification problem as a decision making problem, and we will define a fuzzy preference relation R [13] with the corresponding outputs of the FRBCSs, treating them as fuzzy degrees of preference. In this manner, the computation of each fuzzy degree of preference is based on the aggregation function that combines the positive degrees of association between the fuzzy rules and the input pattern, which is known as Fuzzy Reasoning Method (FRM).

We consider the maximum matching FRM, where every new pattern x_p is classified as the consequent class of a single winner rule ($Class(x_p) = C_w$) which is determined as

$$\mu_{A_w}(x_p) \cdot RW_w = \max\{\mu_{A_q}(x_p) \cdot RW_q, R_q \in RB\} \quad (1)$$

where $\mu_{A_q}(x_p)$ is the membership degree of the pattern example $x_p = (x_{p1}, \ldots, x_{pn})$ with the antecedent of the rule R_q and RW_q is the rule weight [9].

$R(i, j)$ (the fuzzy degree of preference between classes i and j) is the maximum association degree for all rules in RB that concludes class i. $R(i, j)$ will be normalized to $[0, 1]$ having the relation $R(i, j) = 1 - R(j, i)$.

$$R = \begin{bmatrix} - & r_{1,2} & \ldots & r_{1,m} \\ r_{2,1} & - & \ldots & r_{2,m} \\ \vdots & \ddots & \ddots & \vdots \\ r_{m,1} & r_{m,1} & \ldots & - \end{bmatrix} \quad (2)$$

From the fuzzy preference relation we must extract a set of non-dominated alternatives (classes) as the solution of the fuzzy decision making problem and thus, our classification output. Specifically, the maximal non-dominated elements of R are calculated by means of the following operations, according to the non-dominance criterium proposed by Orlovsky [13]:

- First, we compute the fuzzy strict preference relation R' which is equal to:

$$R'(i, j) = \begin{cases} R(i, j) - R(j, i), & \text{when } R(i, j) > R(j, i) \\ 0, & \text{otherwhise.} \end{cases} \quad (3)$$

- Then, we compute the non-dominance degree of each class ND_i, which is simply obtained as:

$$ND_i = 1 - \sup_{j \in C}[R'(j, i)] \quad (4)$$

  This value represents the degree to which the class i is dominated by no one of the remaining classes. C stands for the set of total classes in the data-set. The output class is computed as the index of the maximal non-dominance value:

$$Class(x_p) = \arg \max_{i=1,\ldots,m} \{ND_i\} \quad (5)$$

In order to clarify this procedure, we have selected a pattern from the vehicle data-set, which is shown in Table 1.

Table 1: Vehicle pattern

Compactness = 95.0,
Circularity = 48.0,
Distancecircularity = 83.0,
Radiusratio = 178.0,
Praxisaspectratio = 72.0,
Maxlengthaspectratio = 10.0,
Scatterratio, 162.0,
Elongatedness = 42.0,
Praxisrectangular = 20.0,
Lengthrectangular = 159.0,
Majorvariance = 176.0,
Minorvariance = 379.0,
Gyrationradius = 184.0,
Majorskewness = 70.0,
Minorskewness = 6.0,
Minorkurtosis = 16.0,
Majorkurtosis = 187.0,
Hollowsratio = 197.0;
Class = Van {Van, Saab, Bus, Opel}.

The complete process is depicted in Table 2.

For non fuzzy classifiers, or Machine Learning algorithms that do not have associated a certainty degree for each output, i.e. the C4.5 decision tree [14], we define a binary preference relation in which:

Table 2: Example of the classification process by means of the use of the fuzzy preference relation

Step1. Obtain R:

$$R = \begin{bmatrix} 1.0 & 1.0 & 0.727 & 0.789 \\ 0.0 & 1.0 & 0.005 & 0.0 \\ 0.273 & 0.995 & 1.0 & 0.478 \\ 0.210 & 1.0 & 0.522 & 1.0 \end{bmatrix}$$

Step2. Transform R to R':

$$R' = \begin{bmatrix} 0.0 & 1.0 & 0.455 & 0.579 \\ 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.989 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.044 & 0.0 \end{bmatrix}$$

Step3. Compute ND:

ND = {1.0, 0.0, 0.545, 0.421}

Step4. Get class index:

$Class = \arg\max_{i=1,\ldots,4}\{ND_i\} = 1\ (van)$

$$R(i,j) = \begin{cases} 1, & \text{if } Class(i,j) = i \\ 0, & \text{otherwhise.} \end{cases} \qquad (6)$$

where $Class(i,j)$ stands for the classification output computed by the classifier associated to classes i and j. The remaining of the process follows the same scheme defined in this section, which is summarized in Algorithm 1.

Algorithm 1 Procedure for the multi-classifier learning proposal

1. Divide the training set into $m(m-1)/2$ subsets for all pair of classes.
2. For each training subset i:
   - 2.1. Build a fuzzy classifier composed by a local DB and an RB generated with any rule learning procedure
3. For each input test pattern:
   - 3.1. Build a fuzzy preference relation R as:
     - For each class i, $i = 1,\ldots,m$
       - For each class j, $j = 1,\ldots,m$, $j \neq i$
         - The preference degree for $R(i,j)$ is the normalized association degree for the classifier associated to classes i and j. $R(j,i) = 1 - R(i,j)$
   - 3.2. Transform R to the fuzzy strict preference relation R'.
   - 3.3. Compute the degree of non-dominance for all classes.
   - 3.4. The input pattern is assigned to the class with maximum non-dominance value.

2.3 Fuzzy Hybrid Genetic Based Machine Learning Rule Generation Algorithm

The FH-GBML method [12] consists in a Pittsburgh approach where each rule set is handled as an individual. It also contains a Genetic Cooperative-Competitive learning approach (an individual represents an unique rule), which is used as a kind of heuristic mutation for partially modifying each rule set.

This method uses standard fuzzy rules with rule weights [11] where each input variable x_i is represented by a linguistic term or label. The system defines 14 possible linguistic terms for each attribute as well as a special "do not care" set.

In the learning process, N_{pop} rule sets are created by selecting randomly N_{rule} training patterns. Then, a fuzzy rule from each of the selected training patterns is generated by choosing probabilistically an antecedent fuzzy set from the 14 candidates ($P(B_k) = \frac{\mu_{B_k}(x_{pi})}{\sum_{j=1}^{14}\mu_{B_j}(x_{pi})}$) and each antecedent fuzzy set of the generated fuzzy rule is replaced with *don't care* using a pre-specified probability $P_{don't\ care}$.

N_{pop} -1 rule sets are generated by selection, crossover and mutation in the same manner as the Pittsburgh-style algorithm. Next, with a pre-specified probability, a single iteration of the Genetic Cooperative-Competitive-style algorithm is applied to each of the generated rule sets.

Finally, the best rule set is added in the current population to the newly generated (N_{pop} -1) rule sets to form the next population and, if the stopping condition is not satisfied, the genetic process is repeated again. Classification is performed following the FRM of the winning rule.

3 Experimental Framework

In this section we will first provide details of the real-world multi-class problems chosen for the experimentation and the configuration parameters of the FRBCSs (subsections 3.1 and 3.2 respectively). Next, we will present the statistical tests applied to compare the results obtained along the experimental study (subsection 3.3).

3.1 Data-sets

Table 3 summarizes the properties of the selected data-sets. It shows, for each data-set, the number of examples (#Ex.), the number of attributes (#Atts.), and the number of classes (#Cl.). The *penbased* and *page-blocks* data-sets have been stratified sampled at 10% in order to reduce their size for training. In the case of missing values (*cleveland*) we have removed those instances from the data-set.

3.2 Parameters

We will use the following configuration for the FRBCS approach: product T-norm as conjunction operator, together with the Penalized Certainty Factor heuristic [11] for the rule weight and the winning rule approach for the FRM. Furthermore, for the genetic process of the FH-GBML method, we consider the following values for the parameters:

- Number of fuzzy rules: 5 $\cdot d$ rules.
- Number of rule sets: 200 rule sets.
- Crossover probability: 0.9.
- Mutation probability: $1/d$.

Table 3: Summary Description of the Data-Sets

id	Data-set	#Ex.	#Atts.	#Cl.
bal	balance scale	625	4	3
cle	cleveland	297	13	5
con	contraceptive method choice	1,473	9	3
eco	ecoli	336	7	8
fla	solar flare	1,389	10	6
gla	glass identification	214	9	7
iri	iris	150	4	3
led	led7digit	500	7	10
new	new-thyroid	215	5	3
pag	page-blocks	548	10	5
pen	pen-based recognition	1,099	16	10
shu	shuttle	2,175	9	7
veh	vehicle	846	18	4
vow	vowel	990	13	11
win	wine	178	13	3
yea	yeast	1,484	8	10

- Number of replaced rules: All rules except the best-one (Pittsburgh-part, elitist approach), number of rules / 5 (GCCL-part).
- Total number of generations: 1,000 generations.
- Don't care probability: 0.5.
- Probability of the application of the GCCL iteration: 0.5.

where d stands for the dimensionality of the problem (number of variables).

3.3 Statistical tests for performance comparison

In this paper, we use the hypothesis testing techniques to provide statistical support to the analysis of the results [4, 15]. 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, making the statistical analysis to lose credibility with these type of tests [2, 5]. Specifically, we will use the Wilcoxon signed-rank test [16] as non-parametric statistical procedure for performing pairwise comparisons between two algorithms.

4 Experimental Analysis

Our experimental analysis is focused to determine the goodness of our proposal by con-

Table 4: Average accuracy results for FH-GBML and C4.5. Basic approaches and multi-classifier proposal (noted with suffix -M)

	FH-GBML		FH-GBML-M		C4.5		C4.5-M	
Data-Set	Train	Test	Train	Test	Train	Test	Train	Test
bal	85.64	82.24	87.04	84.96	89.72	77.28	84.24	76.32
cle	63.05	50.84	75.42	56.56	83.41	51.82	82.32	52.53
con	48.78	45.21	57.03	53.43	73.54	51.93	69.98	52.34
eco	79.76	76.19	92.34	81.55	91.74	78.28	85.27	75.59
fla	68.83	68.10	78.52	74.29	79.08	74.48	76.17	73.92
gla	70.44	60.29	82.48	68.25	91.94	68.73	91.94	73.81
iri	99.33	93.33	99.50	94.67	97.83	93.33	97.83	93.33
led	63.60	60.40	80.15	72.60	77.25	70.60	77.15	71.20
new	96.74	91.16	99.53	95.81	98.37	91.16	98.37	93.49
pag	95.67	94.53	98.17	95.62	98.95	95.07	98.90	95.61
pen	71.93	69.82	96.34	91.09	97.82	89.36	98.55	89.36
shu	95.52	95.22	98.45	97.70	99.72	99.54	99.94	99.63
veh	62.44	58.15	74.53	66.67	89.92	71.87	82.74	71.63
vow	28.79	23.94	89.60	77.68	96.29	79.49	96.87	79.80
win	97.61	92.70	100.00	96.08	99.02	94.90	99.02	91.56
yea	53.84	51.22	65.40	58.90	82.18	55.80	70.25	58.96
Mean	73.87	69.58	85.91	**79.12**	90.42	77.73	88.10	**78.07**

trasting the results of the basic and multi-classifier FRBCS approaches against the well-known C4.5 algorithm, also considering a multi-classifier model in this case.

Estimates of accuracy rate were obtained by means of a 5-fold cross-validation. That is, we split the data set into 5 folds, each one containing the 20% of the patterns of the data-set. For each fold, the algorithm was trained with the examples contained in the remaining folds and then, tested with the current fold.

The whole experimental results were presented in Table 4. We must stress that in all cases the best performance is associated to the multi-classifier approach, both for the FRBCS and for C4.5, which implies the goodness of our proposed model.

In order to support this affirmation, we have carried out a Wilcoxon test, shown in Table 5, where we can observe significant differences between the basic FRBCS and the pairwise coupling model in favour of the latter approach. We have also performed another Wilcoxon test (Table 6) that suggests that the application of the multi-classifier approach makes the FRBCS to became very competitive in contrast to the C4.5 decision tree. We observe that the FH-GBML by itself is outperformed by C4.5, whereas with our proposal obtains a better behaviour (refer to the higher ranking and the low p-value obtained in the comparison).

Table 5: Wilcoxon test to compare the basic FH-GBML and C4.5 methods versus their multi-classifier version. R^+ corresponds to the basic approach and R^- to the multi-classifier proposal

Comparison	R^+	R^-	p-value
FH-GBML vs. FH-GBML-M	0.0	136.0	0.000
C4.5 vs. C4.5-M	52.5	83.5	0.470

Table 6: Wilcoxon test to compare FH-GBML and C4.5, both in their basic approach and multi-classifier version. R^+ corresponds to the FH-GBML and R^- to C4.5

Comparison	R^+	R^-	p-value
FH-GBML vs. C4.5	10.5	125.5	0.004
FH-GBML-M vs. C4.5-M	90.0	46.0	0.255

5 Conclusions

In this contribution we have proposed a learning methodology for FRBCSs in order to improve the behaviour of simple fuzzy models. The proposal is based on the application of a pairwise coupling scheme for building a fuzzy multi-classifier system oriented to discriminate between pairs of classes and to obtain a better decision boundary.

In order to aggregate the output for every single classifier, we have made use of a fuzzy preference relation translating the classification problem into a simple decision making problem. The final output class is obtained following the maximal non-dominance criterium.

Our experimental results have shown the high improvement achieved by this model, which outperforms the initial FRBCS. This enhancement enables the fuzzy system to be very competitive even showing a better behaviour than the C4.5 decision tree.

Acknowledgements

This work had been supported by the Spanish Ministry of Science and Technology under Projects TIN2008-06681-C06-01 and TIN2007-65981.

References

[1] A. Asuncion and D. Newman. UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences, 2007. `http://www.ics.uci.edu/~mlearn/MLRepository.html`

[2] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Rechearch*, 7:1–30, 2006.

[3] J. Fürnkranz. Round robin classification. *Journal of Machine Learning Research*, 2:721–747, 2002.

[4] S. García, A. Fernández, J. Luengo, and F. Herrera. A study of statistical techniques and performance measures for genetics–based machine learning: Accuracy and interpretability. *Soft Computing*, 13(10):959–977, 2009.

[5] S. García and F. Herrera. An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. *Journal of Machine Learning Research*, 9:2607–2624, 2008.

[6] T. Hastie and R. Tibshirani. Classification by pairwise coupling. *The Annals of Statistics*, 26(2):451–471, 1998.

[7] J. C. Hühn and E. Hüllermeier. FR3: A fuzzy rule learner for inducing reliable classifiers. *IEEE Transactions on Fuzzy Systems*, 17(1):138–149, 2009.

[8] E. Hüllermeier and K. Brinker. Learning valued preference structures for solving classification problems. *Fuzzy Sets and Systems*, 159(18):2337–2352, 2008.

[9] H. Ishibuchi and T. Nakashima. Effect of rule weights in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*, 9(4):506–515, 2001.

[10] H. Ishibuchi, T. Nakashima, and M. Nii. *Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining.* Springer-Verlag, 2004.

[11] H. Ishibuchi and T. Yamamoto. Rule weight specification in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*, 13:428–435, 2005.

[12] H. Ishibuchi, T. Yamamoto, and T. Nakashima. Hybridization of fuzzy GBML approaches for pattern classification problems. *IEEE Transactions on System, Man and Cybernetics B*, 35(2):359–365, 2005.

[13] S. Orlovsky. Decision-making with a fuzzy preference relation. *Fuzzy Sets and Systems*, 1:155–167, 1978.

[14] J. R. Quinlan. *C4.5: Programs for Machine Learning.* Morgan Kaufmann Publishers, San Mateo–California, 1993.

[15] D. Sheskin. *Handbook of parametric and nonparametric statistical procedures.* Chapman & Hall/CRC, second edition, 2006.

[16] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics*, 1:80–83, 1945.