

Multi-Objective Evolutionary Optimization of Type-2 Fuzzy Rule-based Systems for Financial Data Classification

Michela Antonelli, Dario Bernardo, Hani Hagrass, *Fellow, IEEE*, Francesco Marcelloni, *Member, IEEE*

Abstract— Classification techniques are becoming essential in the financial world for reducing risks and possible disasters. Managers are interested in not only high accuracy but also in interpretability and transparency. It is widely accepted now that the comprehension of how inputs and output are related to each other is crucial for taking operative and strategic decisions. Furthermore, inputs are often affected by contextual factors and characterized by a high level of uncertainty. In addition, financial data are usually highly skewed towards the majority class. With the aim of achieving high accuracies, preserving the interpretability and managing uncertain and unbalanced data, the paper presents a novel method to deal with financial data classification by adopting type-2 fuzzy rule-based classifiers (FRBCs) generated from data by a multi-objective evolutionary algorithm (MOEA). The classifiers employ an approach, denoted as scaled dominance, for defining rule weights in such a way to help minority classes to be correctly classified. In particular, we have extended PAES-RCS, an MOEA-based approach to learn concurrently the rule and data bases of FRBCs, for managing both interval type-2 fuzzy sets and unbalanced datasets. To the best of our knowledge, this is the first work that generates type-2 FRBCs by concurrently maximizing accuracy and minimizing the number of rules and the rule length with the objective of producing interpretable models of real-world skewed and incomplete financial datasets. The rule bases are generated by exploiting a rule and condition selection (RCS) approach, which selects a reduced number of rules from a heuristically generated rule base and a reduced number of conditions for each selected rule during the evolutionary process. The weight associated with each rule is scaled by the scaled dominance approach on the fuzzy frequency of the output class, in order to give a higher weight to the minority class. As regards the data base learning, the membership function parameters of the interval type-2 fuzzy sets used in the rules are learned concurrently to the application of RCS. Unbalanced datasets are managed by using, in addition to complexity, selectivity and specificity as objectives of the MOEA rather than only the classification rate. We tested our approach, named IT2-PAES-RCS, on eleven financial datasets and compared our results with the ones obtained by the original

PAES-RCS with three objectives and with and without scaled dominance, the fuzzy rule-based classifiers FARC-HD and FURIA, the classical C4.5 decision tree algorithm and its cost-sensitive version. Using non-parametric statistical tests, we will show that IT2-PAES-RCS generates FRBCs with, on average, accuracy statistically comparable to and complexity lower than the ones generated by the two versions of the original PAES-RCS. Further, the FRBCs generated by FARC-HD and FURIA and the decision trees computed by C4.5 and its cost-sensitive version, despite the highest complexity, result to be less accurate than the FRBCs generated by IT2-PAES-RCS. Finally, we will highlight how these FRBCs are easily interpretable by showing and discussing one of them.

Index Terms—type-2 fuzzy rule-based classifiers, multi-objective evolutionary fuzzy systems, financial datasets, unbalanced datasets.

I. INTRODUCTION

THE financial crisis of 2008 demonstrated that lack of good information can lead to disasters. Financial services organizations, customers, and particularly regulators quickly came to understand that clear and relevant information was key to risk reduction. Therefore, we are now witnessing ongoing efforts by regulators to ensure that firms operating in financial services generate comprehensive and comprehensible information. Superficially, the demands of regulators look burdensome. In reality, however, they provide an opportunity for organizations to improve their strategic and operational activities through risk reduction based on well-managed information [1].

Machine learning in financial applications differs from other domains in how the quality of a model is assessed. Whereas in most applications, “accuracy of prediction” is often the only metric used, in financial applications, interpretability and transparency are also important and sometimes a requirement. Within financial applications, the accuracy of the model is not the only crucial issue. There is a growing interest in having high levels of model transparency, which is the ability to provide a clear and understandable explanation of the output result. If advanced analytical techniques are used, there is now an obligation to manage the whole process of creating and using the resulting models. It is no longer enough to create a model, deploy it into production and leave it unattended without any oversight. A set of

Manuscript received XXXXX; revised XXXXX and XXXXX; accepted XXXXXX. Date of publication XXXXX; date of current version XXXXX.

M. Antonelli and F. Marcelloni are with the Dipartimento di Ingegneria dell'Informazione, University of Pisa, Pisa I-56100, Italy (e-mail: michela.antonelli@iet.unipi.it; francesco.marcelloni@iet.unipi.it).

D. Bernardo and H. Hagrass are with the Computational Intelligence Centre, School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester, CO43SQ, UK. (e-mail: hani@essex.ac.uk).

M. Antonelli is also with Translational Imaging Group, Centre for Medical Image Computing (CMIC), University College London, London, UK.

> PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

capabilities and processes are required to ensure that every aspect of model creation, deployment and performance is well understood, managed and documented. This implies additional technology infrastructure and methods, since in large firms the number of models in use might be measured in the thousands. This represents a significant shift to much greater sophistication [1]. Another reason why it is important that we can understand models is trust. A system that can explain why a certain decision was taken is more trustworthy in the eyes of a layman user. This need for transparency is reflected in legislation that forces financial institutions to disclose the reasoning behind their financial decisions and models. Left unchecked, inevitably there will be rogue models that cause financial harm and breach regulatory requirements [2]. Furthermore, transparency of a model is important because it allows users to understand data association by observing why a specific decision has been taken. This process helps users to drill-down into their data, understand it, and extract some useful knowledge that could be a competitive advantage in the market. Ultimately, a transparent model can become not only a tool for foresight and prediction but also for analysis and domain knowledge extraction. As it is often the case, managers in finance face two conflicting demands. On the one hand, they need to employ ever more powerful analytical techniques to remain competitive, while, on the other hand, the models they use must be transparent and relatively easy to explain ([1]-[3]).

Neural networks, Bayesian networks, support vector machines are all considered “black box”. This adjective is applied to systems that, for a given input, are able to output a class label, but without providing a clear explanation of the decision process. Logistic regression can provide some statistic correlations between the inputs and the output, but this is not enough to understand why, for a given input, a given label was chosen, or to gain a deep insight of either the model or the data. On the other hand, “white box” models usually refer to rule-based systems that are able to provide an insight of the data on which the models have been trained, and an explanation of the decision process through their rules. Decision trees can translate their internal state into a set of rules and, like any other rule-based system, are able to provide transparency. Nevertheless in complex real world applications, such as in the financial domain, the number of generated rules can explode. It is debatable that a rule base containing thousands of rules can be considered an understandable and transparent model. Decision trees [4]-[6] and random forests [7] produce associations among sets of data, which are selected to optimize the classification problem. Thus, the produced associations could be meaningless in the context of profiling and knowledge extraction.

Fuzzy logic extends the concepts of association rule learning by extending the rule antecedent sets to fuzzy concepts. This technique, in conjunction with genetic and evolutionary algorithms, is a powerful approach for creating accurate and interpretable models. Studies such as [8]-[11] have shown that accuracy and interpretability are in a trade-off and it is necessary to sacrifice one in order to increase the

other. It is difficult to define to which extent accuracy or interpretability can be sacrificed in order to gain in the other. Usually different applications and specific situations have different requirements. Multi-objective genetic algorithms are able to provide an evolution through the two competitive objectives: accuracy and interpretability [12] [13]. Such evolutionary algorithms generate a set of solutions, also known as Pareto front, that optimize both objectives at different levels. This feature gives the ability to easily identify the desired level of complexity/accuracy for the specific application. However, the vast majority of fuzzy systems employ the type-1 fuzzy sets, which cannot directly handle the high levels of uncertainty present in financial applications. Indeed, type-1 fuzzy sets are crisp and precise (i.e., their membership functions are supposed to be perfectly known) and do not allow for any uncertainty about membership values, which is a liability for their use. A type-2 fuzzy set is characterized by a fuzzy membership function, i.e., the membership value for each element of this set is itself a fuzzy set defined on the universe [0,1] [14]. The membership functions of the type-2 fuzzy sets are three-dimensional and include a *footprint of uncertainty*. The third dimension and the footprint of uncertainty provide additional degrees of freedom that make it possible to directly model and handle the high level of uncertainty affecting the inputs in financial applications. In addition, it should be noted that using type-2 fuzzy sets to represent the system inputs can result in reduction of the fuzzy classifier rule base and complexity (as it will be shown in Section IV) when compared to using type-1 fuzzy sets. Indeed, the footprint of uncertainty, which characterizes the type-2 fuzzy sets, lets us cover the same range as type-1 fuzzy sets with a smaller number of labels: of course, the rule reduction will be greater when the number of inputs increases [14].

Previous works have already employed type-2 fuzzy classifiers in financial domain [15]-[17] and have shown how these systems outperform their type-1 versions and other state of the art classifiers. However, to date most of the type-2 fuzzy systems reported in the literature have been generated from data by optimizing only the accuracy, while neglecting the complexity [17]-[23]. This aspect is of major importance to the financial domain since offering compact fuzzy classifiers with the same accuracy as their counterparts will help to realize transparent and easy to understand models, which are becoming essential requirements especially after the recent economic crisis.

In financial applications, as in many real-world problems, the data presents challenges that are not often found in traditional academic datasets. Some of these are: size, noise, sparsity and uncertainty. Furthermore, in the vast majority of financial applications, data is highly unbalanced [24]. For example, in credit card applications the number of good customers is much higher than that of bad customers, and in fraud detection the majority of the data is normal transactions with only a few fraudulent transactions. Most classifiers designed for minimizing the global error rate perform poorly on unbalanced datasets, because they misclassify most of the

> PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

data belonging to the class with few examples. To tackle this problem, pre-processing techniques like under-sampling or over-sampling are usually applied, but both of them present problems. On the one hand, under-sampling techniques may increment the noise since they could eliminate some important patterns. On the other hand, over-sampling techniques may add noise for the original input data or violate the inherent geometrical structure of the minority and majority classes. Hence, in financial applications it is not desirable to pre-process or sample the data as this could cause problems. Thus, there is a need for predictive analytics techniques that can handle unbalanced financial data sets to give accurate and interpretable financial models.

In this paper, with the aim of dealing with uncertain and unbalanced data, and generating accurate and interpretable classifiers, we employ PAES-RCS [25][26], an MOEA-based approach to learn concurrently the rule and data bases of fuzzy rule-based classifiers (FRBCs). In PAES-RCS, the learning process is performed by selecting a set of rules from an initial rule base and a set of conditions for each selected rule. This scheme is denoted as rule and condition selection (RCS). During the multi-objective evolutionary process, PAES-RCS generates the rule bases of the classifiers by using the RCS approach and concurrently learns the membership function parameters of the linguistic values used in the rules. The original PAES-RCS is extended so as to manage interval type-2 (IT2) fuzzy sets and unbalanced datasets. We denote this extension as IT2-PAES-RCS in the following. We modified both the inference mechanism and the evolutionary process for coping with the IT2 fuzzy sets. Further, we adopted three objectives, namely false positive rate (FPR), true positive rate (TPR) and complexity. In our previous works [27], we have verified that the use of FPR and TPR as objectives of the evolutionary optimization process has proved to be very effective in managing unbalanced datasets. Indeed, one of the main strengths of IT2-PAES-RCS is that it can be applied to unbalanced datasets without any rebalancing.

We tested IT2-PAES-RCS on eleven financial datasets and compared the results with the ones obtained by the original PAES-RCS, employing FPR and TPR as objectives, with (PAES-RCS-SD) and without scaled dominance, the fuzzy rule-based classifiers FARC-HD [28] and FURIA [29], the classical C4.5 decision tree algorithm [30] and its cost-sensitive version (C4.5-CS) [31]. Using non-parametric statistical tests, we will show that IT2-PAES-RCS generates FRBCs with accuracy statistically comparable to the ones generated by PAES-RCS and PAES-RCS-SD, employing a lower number of rules and a lower number of conditions in the antecedent of the rules. The FRBCs generated by IT2-PAES-RCS results therefore to be less complex and more interpretable. Further, the FRBCs generated by FARC-HD and FURIA, and the decision trees computed by C4.5 and its cost-sensitive version, despite the lowest interpretability, result to be less accurate than the solutions generated by IT2-PAES-RCS.

The paper is organized as follows. In Section II, we provide a basic description of FRBCs based on IT2 fuzzy sets and

introduce some notations. Section III shows the proposed MOEA-based learning approach and includes the details of the initial rule base generation technique, of the chromosome coding and mating operators, and of the adopted MOEA. In Section IV, we illustrate the experimental results and in Section V we draw some final conclusion.

II. INTERVAL TYPE-2 FUZZY RULE-BASED CLASSIFIER

Object classification consists of assigning a class C_j from a predefined set $\{C_1, \dots, C_K\}$ of classes to an object. Each object is considered as an F -dimensional point in a feature space \mathfrak{R}^F . Let $X = \{X_1, \dots, X_F\}$ be the set of features and $U_f, f = 1, \dots, F$, be the universe of the f -th feature. Let $\tilde{P}_f = \{\tilde{A}_{f,1}, \dots, \tilde{A}_{f,T_f}\}$, $f = 1, \dots, F$, be a fuzzy partition with T_f IT2 fuzzy sets of the universe U_f . We recall that an IT2 fuzzy set \tilde{A} is characterized by a fuzzy membership function $\mu_{\tilde{A}}(x)$, that is, the membership value for each element of this set is a fuzzy set [32]. The membership functions of IT2 fuzzy sets include a footprint of uncertainty, which provides additional degrees of freedom that make it possible to directly model and handle uncertainties. In the IT2 fuzzy sets, all the third dimension values are equal to one. More formally, the membership function $\mu_{\tilde{A}}(x)$ of an IT2 fuzzy set \tilde{A} is defined as:

$$\mu_{\tilde{A}}(x) = \int_{x \in X} \left[\int_{u \in [\underline{\mu}_{\tilde{A}}(x), \bar{\mu}_{\tilde{A}}(x)]} 1/u \right] / x \quad (1)$$

where $\bar{\mu}_{\tilde{A}}(x)$ and $\underline{\mu}_{\tilde{A}}(x)$ represent, respectively, the upper and lower membership functions of the IT2 fuzzy set \tilde{A} . In this paper, we use triangular membership functions defined by three points (a, b, c) , where a and c correspond to the endpoints of the support and b to the core. We build the IT2 fuzzy sets by using the following procedure. First, we define the upper triangular membership functions $\bar{\mu}_{\tilde{A}_f}(x)$ through the three points $(\bar{a}_{f,j}, \bar{b}_{f,j}, \bar{c}_{f,j})$. Then, the left endpoints $\underline{a}_{f,j}$ of the supports of the lower membership functions $\underline{\mu}_{\tilde{A}_f}(x)$ are computed as midpoints $\frac{\bar{a}_{f,j} + \bar{b}_{f,j}}{2}$ between the left endpoints $\bar{a}_{f,j}$ of the supports of the upper membership functions $\bar{\mu}_{\tilde{A}_f}(x)$ and their cores $\bar{b}_{f,j}$. Similarly, the right endpoints $\underline{c}_{f,j}$ of the supports of the lower membership functions $\underline{\mu}_{\tilde{A}_f}(x)$ correspond to the mid-points $\frac{\bar{b}_{f,j} + \bar{c}_{f,j}}{2}$ between the right endpoints $\bar{c}_{f,j}$ of the supports of the upper membership functions $\bar{\mu}_{\tilde{A}_f}(x)$ and the cores $\bar{b}_{f,j}$. It follows that $\underline{a}_{f,j} = \underline{c}_{f,j-1}$, for $j = 2, \dots, T_f$. The cores $\underline{b}_{f,j}$ coincide with the cores $\bar{b}_{f,j}$. Fig. 1 shows an example of IT2 fuzzy partition with $T_f = 5$. Here, the upper membership functions (thick lines) are obtained by defining a uniform Ruspini partition with triangular membership functions on the universe U_f .

The m -th rule R_m ($m=1, \dots, M$) of an IT2 FRBC is typically expressed as:

$$R_m: \mathbf{IF} X_1 \text{ is } \tilde{A}_{1,j_{m,1}} \text{ and } \dots \text{ and } X_F \text{ is } \tilde{A}_{F,j_{m,F}} \quad (2)$$

THEN Y is C_{j_m} with RW_{j_m}

where Y is the classifier output, C_{j_m} is the class label

> PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

associated with the m -th rule, $j_{m,f} \in [1, T_f]$ identifies the index of the IT2 fuzzy set (among the T_f IT2 fuzzy sets of the partition \tilde{P}_f), which has been selected for X_f in rule R_m , and RW_m is the *rule weight*, i.e., a *certainty degree* of the classification in the class C_{j_m} for a pattern that fires the antecedent of the rule.

Let $T = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$ be a training set composed of N input-output (\mathbf{x}_t, y_t) pairs, with $\mathbf{x}_t = [x_{t,1}, \dots, x_{t,F}] \in \mathfrak{R}^F$ and $y_t \in \{C_1, \dots, C_K\}$. The strength of activation $w_m(\mathbf{x}_t)$ (*matching degree* of the rule with the input) of the rule R_m is calculated as:

$$w_m(\mathbf{x}_t) = \frac{w_m(\mathbf{x}_t) + \bar{w}_m(\mathbf{x}_t)}{2} \quad (3)$$

where $w_m(\mathbf{x}_t) = \prod_{f=1}^F \underline{\mu}_{\tilde{A}_f}(x_{t,f})$ and $\bar{w}_m(\mathbf{x}_t) = \prod_{f=1}^F \bar{\mu}_{\tilde{A}_f}(x_{t,f})$ are the lower and upper bounds of the strength of activation computed, respectively, on the lower and upper membership functions. To take the “don’t care” condition into account, a particular IT2 fuzzy set $\tilde{A}_{f,0}$ ($f = 1, \dots, F$) is added to all the F partitions \tilde{P}_f . This fuzzy set is characterized by both the lower and upper membership functions equal to 1 on the overall universe. This means that the condition $X_f \in \tilde{A}_{f,0}$ does not affect the computation of the strength of activation. In other words, for the specific rule, the variable X_f is not taken into account and therefore can be removed. The terms $\tilde{A}_{f,0}$, therefore, allow generating rules, which contain only a subset of the input variables, thus reducing the total rule length and consequently increasing the interpretability of the rules.

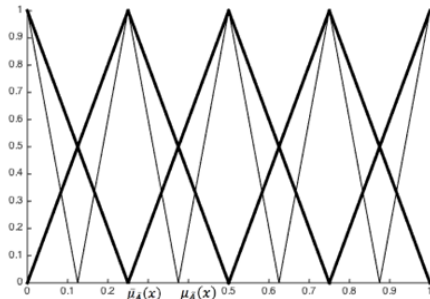


Fig. 1. An example of IT2 fuzzy partitions with $T_f = 5$ IT2 fuzzy sets (the thick and thin lines represent the upper and lower membership functions, respectively).

As we have pointed out in Section I, financial data is usually highly unbalanced. To give minority class a fair chance when competing with majority class, we adopted a new approach to calculate the rule weight that takes the fuzzy frequency of the class into account. The approach is called “scaled dominance”, and has been introduced in [33]-[35]. In the literature, fuzzy rule weights are traditionally calculated as fuzzy extension of the confidence and support. Confidence and support are data mining metrics used in association rule learning. These metrics, in fuzzy rule-based systems, are extended by using fuzzy strength instead of crisp counting of the item-sets. The confidence and support extensions used in

this paper exploit a scaled version w_m^s of the matching degree. For a given rule R_m , having a consequent class C_{j_m} , we scale the matching degree of the rule by dividing the upper and lower bounds of the strengths of activation by the sum of, respectively, the upper $\bar{w}_l(\mathbf{x}_t)$ and lower $w_l(\mathbf{x}_t)$ bounds of the strengths of activation of all the rules R_l , which have C_{j_m} as the consequent class. The scaled upper and lower bounds are therefore computed as follows:

$$\bar{w}_m^s(\mathbf{x}_t) = \frac{\bar{w}_m(\mathbf{x}_t)}{\sum_{l, out=C_{j_m}} \bar{w}_l(\mathbf{x}_t)} \quad (4)$$

$$w_m^s(\mathbf{x}_t) = \frac{w_m(\mathbf{x}_t)}{\sum_{l, out=C_{j_m}} w_l(\mathbf{x}_t)}. \quad (5)$$

In IT2 fuzzy rule-based systems, confidence and support of a rule are determined from the strength of activation and, therefore, defined by upper and lower bounds. From equations (4) and (5), we derive the following scaled upper and lower bounds of the confidence:

$$\bar{c}_m^s(Ant_m \Rightarrow C_{j_m}) = \frac{\sum_{\mathbf{x}_t \in C_{j_m}} \bar{w}_m^s(\mathbf{x}_t)}{\sum_{m=1}^M \bar{w}_m^s(\mathbf{x}_t)} \quad (6)$$

$$c_m^s(Ant_m \Rightarrow C_{j_m}) = \frac{\sum_{\mathbf{x}_t \in C_{j_m}} w_m^s(\mathbf{x}_t)}{\sum_{m=1}^M w_m^s(\mathbf{x}_t)} \quad (7)$$

where M is the number of rules in the rule base and Ant_m is the antecedent of R_m . The confidence can be viewed as a numerical approximation of the conditional probability $P(C_{j_m} | Ant_m)$. The scaled upper and lower bounds of the support are defined as:

$$\bar{s}_m^s(Ant_m \Rightarrow C_{j_m}) = \frac{\sum_{\mathbf{x}_t \in C_{j_m}} \bar{w}_m^s(\mathbf{x}_t)}{M} \quad (8)$$

$$s_m^s(Ant_m \Rightarrow C_{j_m}) = \frac{\sum_{\mathbf{x}_t \in C_{j_m}} w_m^s(\mathbf{x}_t)}{M}. \quad (9)$$

The support can be viewed as a measure of the coverage of training patterns performed by R_m .

The rule weight is then calculated as product of the scaled confidence and support. It follows that the rule weight RW_m in (2) becomes a closed interval bounded by the upper \bar{RW}_m and \underline{RW}_m endpoints, calculated as:

$$\bar{RW}_m = \bar{c}_m^s \cdot \bar{s}_m^s \quad (10)$$

$$\underline{RW}_m = c_m^s \cdot s_m^s \quad (11)$$

The *association degree* with the class C_{j_m} will be, in its turn, a closed interval bounded by the upper $\bar{h}_m(\mathbf{x}_t)$ and lower $\underline{h}_m(\mathbf{x}_t)$ endpoints, which are computed as follows:

> PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

$$\bar{h}_m(x_t) = \bar{w}_m^s(x_t) \cdot \overline{RW}_m = \bar{w}_m^s(x_t) \cdot \frac{\sum_{x_t \in C_{j_m}} \bar{w}_m^s(x_t)}{\sum_{m=1}^M \bar{w}_m^s(x_t)} \cdot \frac{\sum_{x_t \in C_{j_m}} \bar{w}_m^s(x_t)}{M} = \frac{\sum_{x_t \in C_{j_m}} \bar{w}_m^s(x_t)}{M} \cdot \frac{(\sum_{x_t \in C_{j_m}} \bar{w}_m^s(x_t))^2}{\sum_{m=1}^M \bar{w}_m^s(x_t)} \quad (12)$$

$$\underline{h}_m(x_t) = \underline{w}_m^s(x_t) \cdot \underline{RW}_m = \underline{w}_m^s(x_t) \cdot \frac{\sum_{x_t \in C_{j_m}} \underline{w}_m^s(x_t)}{\sum_{m=1}^M \underline{w}_m^s(x_t)} \cdot \frac{\sum_{x_t \in C_{j_m}} \underline{w}_m^s(x_t)}{M} = \frac{\sum_{x_t \in C_{j_m}} \underline{w}_m^s(x_t)}{M} \cdot \frac{(\sum_{x_t \in C_{j_m}} \underline{w}_m^s(x_t))^2}{\sum_{m=1}^M \underline{w}_m^s(x_t)} \quad (13)$$

We adopt the maximum matching method as *reasoning method*: an input pattern is classified into the class corresponding to the rule with the maximum association degree calculated for the pattern. In the case of tie, we randomly classify the pattern. The association degree for rule R_m is computed as:

$$h_m(x_t) = \frac{\bar{h}_m(x_t) + \underline{h}_m(x_t)}{2} \quad (14)$$

Once fixed the number T_f of IT2 fuzzy sets for each linguistic variable, we adopt an MOEA-based approach to learn rules and membership function parameters so as to generate a set of IT2 FRBCs with different trade-offs between accuracy and rule base complexity.

III. THE PROPOSED THREE OBJECTIVE EVOLUTIONARY OPTIMIZATION OF IT2 FUZZY RULE-BASED CLASSIFIERS

MOEAs have been applied in several different domains to search for optimal solutions to problems characterized by multiple performance criteria in competition with each other [36]. MOEAs do not generate a unique solution, but rather a set of equally valid solutions, where each solution tends to fulfill a criterion to a higher extent than another. Comparison between different solutions is performed by using the notion of Pareto dominance. A solution x , associated with a performance vector \mathbf{u} , dominates a solution y , associated with a performance vector \mathbf{v} , if and only if, $\forall i \in \{1, \dots, I\}$, with I the number of criteria, u_i performs better than, or equal to, v_i and $\exists i \in \{1, \dots, I\}$, such that u_i performs better than v_i , where u_i and v_i are the i -th elements of vectors \mathbf{u} and \mathbf{v} , respectively. The set of solutions, which are not dominated by any other possible solution, is denoted as Pareto front. The objective of any MOEA is, therefore, to search for a set of solutions that are a good approximation of the Pareto front. In the last years, in designing fuzzy rule-based systems, developers have not only considered accuracy, but also interpretability as a crucial requirement. Since accuracy and interpretability are objectives in competition with each other, MOEAs have been so extensively applied that the term multi-objective evolutionary fuzzy system has been coined to identify fuzzy rule-based systems generated by MOEAs [12] [13] [37] [38]. While the accuracy objective has been typically measured in terms of classification rate and approximation error for, respectively, classification and regression problems, a number of specific measures have been proposed for evaluating the

interpretability, taking the rule base complexity and the data base integrity into account [39] [40]. A large number of contributions have been recently published under the framework of multi-objective evolutionary fuzzy systems, with application mostly to regression [41]-[51] and classification [52]-[58] problems. Recently, some taxonomies of the main contributions have been also introduced in [12] [13].

In this paper, we extend PAES-RCS, a multi-objective evolutionary fuzzy system that has been recently proposed by some of the authors of this paper in [25][26]. PAES-RCS has proved to be very effective and efficient in classification problems [26]. The original PAES-RCS learns concurrently the rule and data bases of type-1 FRBCs by exploiting the RCS approach, which selects a reduced number of rules from a heuristically generated rule base and a reduced number of conditions for each selected rule during the evolutionary process. Thus, RCS can be considered a sort of rule learning in a search space constrained by the heuristically generated rule base. The membership function parameters of the type-1 fuzzy sets are learned concurrently to the application of RCS. This requires an appropriate chromosome coding and properly defined mating operators. In particular, chromosome C consists of two parts (C_{RB} , C_{DB}), which define the rule base and the membership function parameters of the input variables, respectively. Both crossover and mutation operators are applied to each part of the chromosome independently. The objectives used in PAES-RCS are classification rates and complexity measured in terms of the total number of antecedent conditions of the rules in the rule base.

In this paper, we extend PAES-RCS along three directions. First of all, we employ IT2 fuzzy sets rather than type-1 fuzzy sets. This has required the adoption of a different inference mechanism. Second, in order to cope with unbalanced datasets, we split the accuracy into two objectives, namely True Positive Rate (TPR) and False Positive Rate (FPR). We recall that TPR and FPR coincide, respectively, with the sensitivity and the complement to 1 of the specificity. As experimented in [27] and [37] using rule learning, this approach allows achieving high accuracies when dealing with unbalanced datasets without needing to re-balance the dataset. Third, we use an approach denoted as scaled dominance, which was introduced in [33]-[35], to handle unbalanced data by trying to give minority classes a fair chance when competing with a majority class. This improvement further contributes to manage unbalanced data.

In the following subsections, we will discuss the method to generate the initial rule base and summarize the RCS approach and the membership function parameter learning used in IT2-PAES-RCS.

A. The initial rule base generation

We generate the initial rule base by first transforming each continuous variable into a categorical and ordered variable. Then, we apply the well-known C4.5 algorithm to the transformed dataset for generating a decision tree. Finally, we extract the initial rule base from the decision tree.

More precisely, for each continuous variable X_f , first we generate an IT2 fuzzy partition $\tilde{P}_f = \{\tilde{A}_{f,1}, \dots, \tilde{A}_{f,T_f}\}$ of T_f IT2 fuzzy sets as shown in Fig. 1. The number T_f of IT2 fuzzy sets can be different from an input variable to another. For the sake of simplicity, in our experiments, we have used the same number of IT2 fuzzy sets for all the variables X_f . Then, we compute the α -cut, with $\alpha=0.5$, of the fuzzy sets defined by the upper membership functions $\bar{\mu}_{\tilde{A}_{f,j}}$ of the IT2 fuzzy sets $\tilde{A}_{f,j}$, $j=1, \dots, T_f$. The corresponding contiguous intervals, shown in Fig. 2, are used to discretize the universe U_f of each variable X_f before applying the C4.5 algorithm. For simplicity, we will denote the intervals with the index of the corresponding IT2 fuzzy set, which the α -cut is applied to. For instance, interval 1 denotes the interval corresponding to the α -cut of the fuzzy set defined by $\bar{\mu}_{\tilde{A}_{f,1}}$. Then, each input value of the input-output pairs in the training set is replaced by the interval, which contains it. Thus, the overall training set is transformed so as to contain exclusively categorical values. Finally, we apply the classical C4.5 algorithm to the transformed training set. We extract the initial rule base from the decision tree generated by the C4.5 algorithm. Rules are extracted from each path from the root to a leaf node. Each splitting criterion along a given path is logically ANDed to form the rule antecedent (“IF” part). The leaf node holds the class prediction, forming the rule consequent (“THEN” part). Since each branch is identified by one of the intervals determined by the discretization process and an input variable is involved in just one node in a path, the rules extracted from the decision tree are expressed as in (2). Each rule is identified by an integer from 1 to M_{C45} , where M_{C45} is the number of rules extracted from the tree and included in the initial rule base.

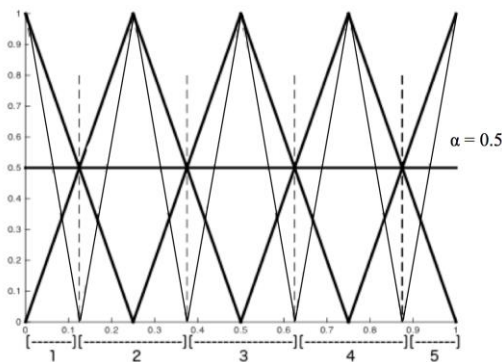


Fig. 2. Discretization of the universe U_f based on an IT2 fuzzy partition (the thick and thin lines represent the upper and lower membership functions, respectively; the dashed lines denote the boundaries of the intervals generated by the α -cut).

Figure 3 shows an example of a decision tree generated by the C4.5 algorithm from a training set characterized by six input variables and two classes (C_1, C_2). Each input variable X_f , $f = 1, \dots, 6$, has been partitioned with $T_f = 5$ fuzzy sets. We observe that only three out of the six original input variables are included in the decision tree. This is due to the well-known characteristic of the C4.5 algorithm that can select features

during the generation of the tree. Figure 4 shows the rule base extracted from the decision tree of Figure 3. We note that the rule base consists of thirteen rules, which correspond to the thirteen possible paths from the root to the leaf nodes.

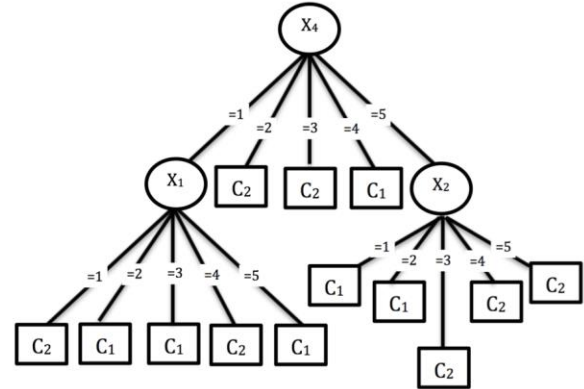


Figure 3. An example of decision tree generated by the C4.5 algorithm applied to the transformed training set.

- | |
|--|
| <p>R_1: IF X_4 is $A_{4,1}$ and X_1 is $A_{1,1}$ THEN Y is C_2</p> <p>R_2: IF X_4 is $A_{4,1}$ and X_1 is $A_{1,2}$ THEN Y is C_1</p> <p>R_3: IF X_4 is $A_{4,1}$ and X_1 is $A_{1,3}$ THEN Y is C_1</p> <p>R_4: IF X_4 is $A_{4,1}$ and X_1 is $A_{1,4}$ THEN Y is C_2</p> <p>R_5: IF X_4 is $A_{4,1}$ and X_1 is $A_{1,5}$ THEN Y is C_1</p> <p>R_6: IF X_4 is $A_{4,2}$ THEN Y is C_2</p> <p>R_7: IF X_4 is $A_{4,3}$ THEN Y is C_2</p> <p>R_8: IF X_4 is $A_{4,4}$ THEN Y is C_1</p> <p>R_9: IF X_4 is $A_{4,5}$ and X_2 is $A_{2,1}$ THEN Y is C_1</p> <p>R_{10}: IF X_4 is $A_{4,5}$ and X_2 is $A_{2,2}$ THEN Y is C_1</p> <p>R_{11}: IF X_4 is $A_{4,5}$ and X_2 is $A_{2,3}$ THEN Y is C_2</p> <p>R_{12}: IF X_4 is $A_{4,5}$ and X_2 is $A_{2,4}$ THEN Y is C_2</p> <p>R_{13}: IF X_4 is $A_{4,5}$ and X_2 is $A_{2,5}$ THEN Y is C_2</p> |
|--|

Figure 4. The fuzzy rule base extracted from the decision tree shown in Fig. 3.

B. Rule and condition selection

The C_{RB} part of the chromosome is a vector of M_{\max} pairs $p_m = (k_m, v_m)$, where k_m identifies the index of the rule in the set of M_{C45} rules extracted from the decision tree, and $v_m = [v_{m,1}, \dots, v_{m,F}]$ is a binary vector, which indicates, for each condition in the rule, if the condition is present ($v_{m,f} = 1$) or corresponds to a “don’t care” ($v_{m,f} = 0$). Rule bases generated by the C4.5 algorithm could include a high number of rules, especially when dealing with large and high dimensional training sets. With the aim of obtaining compact and interpretable FRBCs, we have set an upper bound M_{\max} to the number of rules that can be contained in any rule base generated during the evolutionary process. In the experiments, we have set $M_{\max} = 50$. In our previous works [26], we have verified that this value permits us to generate FRBCs with reasonable accuracy,

> PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

maintaining the complexity at an adequate level. Let M_{C45} be the number of rules extracted from the decision tree. If $M_{C45} < M_{\max}$, then $M_{\max} = M_{C45}$. During the evolutionary process, the MOEA can generate rule bases, which contain a number of rules lower than M_{\max} . Indeed, if $k_m = 0$, then the m^{th} rule is not included in the rule base. Further, the number of conditions can be lower than the number F of features. Indeed, if $v_{m,f} = 0$, then the f^{th} condition of the m^{th} rule is replaced by a “don’t care” condition and, therefore, is not considered in the inference process. Whenever a condition selection is performed on the rule, the rule weight associated with the rule is re-computed.

As an example, given a two input fuzzy model, let us assume that the C4.5 algorithm has generated the following four rules:

- R_1 : **IF** X_1 **is** $\tilde{A}_{1,1}$ **and** X_2 **is** $\tilde{A}_{2,1}$ **THEN** Y **is** C_1
 R_2 : **IF** X_1 **is** $\tilde{A}_{1,2}$ **and** X_2 **is** $\tilde{A}_{2,2}$ **THEN** Y **is** C_2
 R_3 : **IF** X_1 **is** $\tilde{A}_{1,5}$ **and** X_2 **is** $\tilde{A}_{2,3}$ **THEN** Y **is** C_1
 R_4 : **IF** X_2 **is** $\tilde{A}_{2,1}$ **THEN** Y **is** C_1

Let us suppose that, during the evolutionary process executed with $M_{\max} = 3$, the C_{RB} chromosome part shown in Figure 5 is generated.

k_1	$v_{1,1}$	$v_{1,2}$	k_2	$v_{2,1}$	$v_{2,2}$	k_3	$v_{3,1}$	$v_{3,2}$
2	1	1	3	1	0	0	1	0

Figure 5. An example of the C_{RB} part of a chromosome.

The first gene of the chromosome selects rule R_2 (k_1 is equal to 2) with all the conditions (both $v_{1,1}$ and $v_{1,2}$ are equal to 1). The second gene selects rule R_3 (k_2 is equal to 3) with only the first condition ($v_{2,1}$ is equal to 1, while $v_{2,2}$ is equal to 0). The third gene selects no rule (k_3 is equal to 0).

The rule base corresponding to the chromosome in Fig. 5 will therefore be:

- R_2 : **IF** X_1 **is** $\tilde{A}_{1,2}$ **and** X_2 **is** $\tilde{A}_{2,2}$ **THEN** Y **is** C_2
 R_3 : **IF** X_1 **is** $\tilde{A}_{1,5}$ **THEN** Y **is** C_1

We note that, even though $M_{\max} = 3$, only two rules have been selected in the final rule base. Furthermore, for the third rule, only the first condition has been selected.

The C_{DB} part of the chromosome codifies the upper membership functions of each variable X_f . Since the lower membership functions are built, as described in Section II, from the upper membership functions, the C_{DB} part codifies exclusively these functions. Since we adopt strong fuzzy partitions for defining the upper membership functions with, for $j = 2, \dots, T_f - 1$, $\bar{b}_{f,j} = \bar{c}_{f,j-1}$ and $\bar{b}_{f,j} = \bar{a}_{f,j+1}$, each triangular fuzzy set $(\bar{a}_{f,j}, \bar{b}_{f,j}, \bar{c}_{f,j})$ of the partition is completely defined by fixing the positions of the cores $\bar{b}_{f,j}$ along the universe U_f of the f^{th} variable (we normalize each variable in $[0,1]$). Since $\bar{b}_{f,1}$ and \bar{b}_{f,T_f} coincide with the lower and upper extremes of universe U_f , the partition of each linguistic variable X_f is completely defined by $T_f - 2$ parameters

$\{\bar{b}_{f,2}, \dots, \bar{b}_{f,T_f-1}\}$, which define the positions of the cores of the upper membership functions defined on X_f . As shown in Fig. 6, the C_{DB} chromosome part, therefore, consists of F vectors of $T_f - 2$ real numbers. A good level of integrity, in terms of order, coverage and distinguishability, of the partitions is ensured by, $\forall j \in [2, T_f - 1]$, forcing $\bar{b}_{f,j}$ to vary in the definition interval $[\bar{b}_{f,j} - \frac{\bar{b}_{f,j} - \bar{b}_{f,j-1}}{2}, \bar{b}_{f,j} + \frac{\bar{b}_{f,j+1} - \bar{b}_{f,j}}{2}]$.

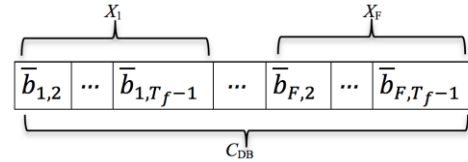


Figure 6. The C_{DB} part of a chromosome.

C. The genetic operators

Both crossover and mutation operators are employed to generate the offspring population. In particular, we apply the one-point crossover to the C_{RB} part and the BLX- α crossover, with $\alpha = 0.5$, to the C_{DB} part. In applying the one-point crossover, the common gene between the two mating chromosomes s_1 and s_2 is determined by extracting randomly a number in $[1, \rho_{MAX}]$, where ρ_{MAX} is the maximum number of rules in s_1 and s_2 . The crossover point is always chosen between two rules and not within a rule.

As regards mutation, two operators are applied to the C_{RB} part. Both the operators randomly choose a pair p_m , i.e. a rule, in the chromosome. Then, the first operator replaces the rule in p_m with another rule by setting k_m to an integer value randomly generated in $[1, M_{C45}]$. The second operator modifies the rule in p_m by complementing each gene $v_{m,f}$ with a probability equal to P_{cond} ($P_{cond} = 2/f$ in the experiments).

The mutation operator applied to C_{DB} , first, randomly chooses an input variable X_f , $f \in [1, F]$, and a fuzzy set $j \in [2, T_f - 1]$ and then replaces the value of $\bar{b}_{f,j}$ with a value randomly chosen within the definition interval of $\bar{b}_{f,j}$.

If, after applying the crossover, the rule base contains one or more pairs of equal rules, we simply eliminate one of the rules from each pair setting the corresponding k_m to zero.

D. Multi-objective evolutionary algorithm

The MOEA used in this paper is the (2+2)M-PAES algorithm proposed in [41] and adopted in [26]. The application scheme of the crossover and mutation operators employed in (2+2)M-PAES for generating the offspring solutions o_1 and o_2 from the current solutions s_1 and s_2 is shown in Figure 7. Here, P_{CRB} , P_{CDB} , P_{MRB1} and P_{MRB2} represent the probabilities of applying the crossover operators to C_{RB} and C_{DB} parts and the first and the second mutation operators to C_{RB} , respectively. P_{MDB} represents the probability of applying the mutation operator to C_{DB} . Unlike classical (2+2)PAES, which maintains the current solutions s_1 and s_2 until they are not replaced by solutions with particular characteristics, we

> PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

observe that in (2+2)M-PAES s_1 and s_2 are randomly extracted at each iteration.

At the beginning, we generate two current solutions s_1 and s_2 . While the genes of the C_{DB} part and the k_m values of the C_{RB} part of s_1 and s_2 are randomly generated, all the values $v_{m,f}$ of the conditions of all the rules are set to 1. An offspring solution o_x is added to the archive only if it is dominated by no solution contained in the archive; possible solutions in the archive dominated by o_x are removed. If the archive is full and no solution in the archive can be removed, then the offspring solution o_x is inserted into the archive and the solutions (possibly o_x itself), which belong to the region with the highest crowding degree, are removed. If the region contains more than one solution, then, the solution to be removed is randomly chosen. (2+2)M-PAES concurrently optimizes three objectives, namely false positive rate (FPR), true positive rate (TPR) and complexity. The complexity is measured as the sum of the conditions, which compose the antecedents of the rules in the rule base. This number is denoted as total rule length (TRL). Low values of TRL correspond to rule bases characterized by a low number of rules and a low number of input variables really used in each rule.

```

//Generate two new solutions
[s1, s2] = random_selection(archive)
o1 = s1
o2 = s2
if (rand() < P_CRB)
    [o1.CRB, o2.CRB] = crossover_CRB(s1.CRB, s2.CRB);
else
    P_MRB1 = 1;
endif
if (rand() < P_CDB)
    [o1.CDB, o2.CDB] = crossover_CDB(s1.CDB, s2.CDB);
endif

loop i=1,2
    if (rand() < P_MRB1)
        o1.CRB = first_mutation_operator(o1.CRB);
    endif
    if (rand() < P_MRB2)
        o1.CRB = second_mutation_operator(o1.CRB);
    endif
    if (rand() < P_MDB)
        o1.CDB = DB_mutation_operator(o1.CDB);
    endif
endloop

```

Figure 7. Application scheme of the genetic operators.

IV. EXPERIMENTS AND RESULTS

We analysed eleven financial datasets. For each dataset, we performed a ten-fold cross-validation and executed three trials for each fold with different seeds for the random function generator (30 trials in total). We fixed 50,000 evaluations as stopping criteria.

In the following, we first describe the financial datasets. Then, we show the results obtained by IT2-PAES-RCS, PAES-RCS, PAES-RCS-SD, FARC-HD, FURIA, C4.5 and its cost-

sensitive version C4.5-CS. Finally, we analyse the results along accuracy and interpretability dimensions.

A. The Financial Datasets

In financial applications, as in many real-world problems, the data is highly unbalanced. For example, in a credit card application, the number of good customers is much higher than that of bad customers; in fraud detection, the majority of the data are normal transactions whereas a few fraudulent transactions are usually present. Most classifiers designed for minimizing the global error rate perform poorly on unbalanced datasets because they misclassify most of the data belonging to the class represented by few examples. Hence, in our experiments, in order to evaluate the proposed system for various financial applications, we have chosen eleven datasets with various sizes and different levels of imbalance ratios between the minority and majority classes. The chosen datasets cover different financial applications, including credit card and loan authorization, stock market related predictions, insurance, fraud detection and investment banking.

We have used eleven real-world datasets from various financial domains. Table I summarizes the main characteristics of these datasets. For each dataset, we report the name, the number of instances (#Instances), the number of attributes (#Attributes), and the imbalance ratio (IR). We recall that IR is defined as the ratio between the number of instances of the majority class and of the minority class. The datasets are sorted for increasing IRs. We do not show the number of classes because all the datasets represent two classes problems.

TABLE I
FINANCIAL DATASETS USED IN THE EXPERIMENTS
(SORTED FOR INCREASING IRs)

Dataset	#Instances	#Attributes	IR
BLA	1747	42	1.47
CARD	176463	66	1.59
AF	1894	121	2.50
ARB	1641	7	3.09
COMM	16102	83	3.34
SL	35798	63	4.16
LEN	24772	20	4.50
DPKG	72983	23	7.20
BAN	45211	13	7.54
GIV	150000	10	13.96
COI	9823	85	15.79

In the following, we shortly describe each financial dataset.

- **BLA:** the dataset is related to the prediction of good (profitable) or bad (non-profitable) customers for bank loan authorization.
- **CARD:** the dataset is used to evaluate if a customer is going to default on a credit card or no.
- **AF:** the dataset is related to investment banking and is used to predict if customers are going to pay back their loans or if they will default on the given loan.
- **ARB:** the dataset is used for spotting arbitrage opportunities in the London International Financial Futures Exchange (LIFFE) market. The dataset was developed in

> PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

[4]-[6] to identify arbitrage situations by analyzing option and futures prices in the LIFFE market.

- COMM: the dataset is used for the evaluation of customers (Fraud or No Fraud customer) for commercial loans applications.
- SL: the dataset is used for the evaluation of customers (good or bad customers) for personal small loans applications where there is no knowledge on the customer full credit history.
- LEN: the dataset is used for evaluation of small companies (good or bad customer) for business loans applications when the customer full credit history is known.
- DPKG: the dataset is used to predict whether in an auction, the customer will be real or fraud.
- BAN: the dataset is used to predict if a customer is eligible for increasing the credit limits on her/his credit cards.
- GIV: the dataset is used to predict whether an applicant is eligible to give her/him extra credit on her/his existing loan or not.
- COI: the dataset is used to predict whether a customer will buy a caravan insurance or not.

B. The Classifiers

In this section, we shortly describe the classifiers applied to the financial datasets. IT2-PAES-RCS was widely discussed in Section III. The PAES-RCS algorithm used in this paper is slightly different from the original version. Indeed, to manage unbalanced datasets, we use three objectives as in IT2-PAES-RCS, but generate type-1 FRBCs. PAES-RCS-SD is the version of PAES-RCS with three objectives and with the scaled domain approach.

FARC-HD (Fuzzy Association Rule-Based Classification Model for High Dimensional Datasets) was introduced in [28] and is a single objective evolutionary fuzzy classifier, which exploits association rules mining for generating FRBCs. FARC-HD is based on three stages. First, it mines all possible fuzzy association rules building a search tree to list all frequent fuzzy item sets, limiting the depth of the branches in order to find a small number of short fuzzy rules. Second, it uses a pattern weighting scheme to reduce the number of candidate rules, preselecting the most interesting rules, in order to decrease the computational costs for the third step. Finally, a single-objective genetic algorithm, namely CHC, is used to select and tune a compact set of fuzzy association rules. FRBCs generated by FARC-HD use the certainty factor and the additive combination [59] as rule weight and reasoning method, respectively.

FURIA (Fuzzy Unordered Rules Induction Algorithm) is an extension of the RIPPER algorithm [60]. Given a classification problem with K classes, prior to the learning process, RIPPER sorts the training data by class label in ascending order according to the corresponding class frequencies. Then, rules are learned for the first $K - 1$ classes, starting with the least frequent. Once a rule has been generated, the instances covered by that rule are removed from the training data, and this is repeated until no instance from the target class is left. The

algorithm then proceeds with the next class. Finally, when RIPPER finds no more rules to learn, a default rule (with empty antecedent) is added for the last (and hence most frequent) class. To learn each rule the training set is split into a growing set and a pruning set: the former is used to specialize the rule by adding antecedents, while the latter is used to generalize the rule by removing antecedents. FURIA extends RIPPER along three directions: i) the use of fuzzy rather than crisp rules, employing fuzzy intervals with trapezoidal membership functions instead of crisp intervals, ii) the exploitation of unordered rather than ordered rule sets, and iii) the introduction of a novel rule stretching method in order to manage uncovered examples.

C4.5 builds decision trees from a set of training data using the concept of information entropy. At each node of the tree, the C4.5 algorithm chooses one attribute of the training set that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is chosen to make the decision. The cost-sensitive version of C4.5, denoted as C4.5-CS, exploits an instance weighting method similar to the one adopted in the boosting decision tree approach developed by Quinlan [61]. C4.5-CS changes the class distribution so that the induced tree is in favour of the class with high weight/cost. Thus, this version of the C4.5 is less likely to commit errors with high costs.

Before applying FARC-HD, FURIA and C4.5, the datasets are pre-processed by using the Synthetic Minority Oversampling Technique (SMOTE) [62]. In SMOTE, the minority class is oversampled by taking each minority class sample and introducing synthetic examples along the line segments joining any or all of the k minority class nearest neighbours. Depending upon the amount of oversampling required, neighbours from the k -nearest neighbours are randomly chosen.

Table II shows the parameters used for IT2-PAES-RCS, PAES-RCS and PAES-RCS-SD. The values of the parameters come, on the one side, from the long experience we matured in the application of (2+2)M-PAES for generating fuzzy rule-based systems since our initial paper on this subject [41]. On the other side, we performed a number of experiments with different values of these parameters using the datasets in Table I and realized that the parameters in Table II are effective also for these datasets. For the other algorithms, we adopted the implementation in Keel [63] and the default parameters.

C. Analysis of the results

The execution of IT2-PAES-RCS, PAES-RCS and PAES-RCS-SD generates a set of solutions with different trade-offs among the three objectives. At the end of each execution of the algorithms, we verified that the archive of (2+2)M-PAES is always full for each dataset in Table I. Thus, each execution of the three algorithms generates 128 different FRBCs. In order to analyse the results of IT2-PAES-RCS, PAES-RCS and PAES-RCS-SD, each three-dimensional Pareto front approximation is

projected onto the FPR-TPR plane: each FRBC of the Pareto front approximation is therefore represented as a point corresponding to the pair (FPR, TPR). We recall that one classifier in the FPR-TPR plane is better than (dominates) another if it is located more north-west (higher TPR and/or lower FPR) than the other [64]. For this reason, in order to select a set of potentially optimal FRBCs, we extract the non-dominated solutions obtained on the training set in the FPR-TPR plane. Since we do not assume to use any cost function for selecting a single optimal classifier, we consider all the non-dominated solutions in the FPR-TPR plane. With the aim of comparing the outputs of the three multi-objective evolutionary approaches among them and with the other algorithms, for each non-dominated solution, we calculate the Area under the Curve (AUC), defined as $AUC = \frac{100+TPR-FPR}{2}$, and select the solution with the highest AUC on the training set. The highest AUC identifies the most north-west solution in the FPR-TPR plane. Thus, for each comparison algorithm, we consider just one classifier and compare these classifiers in terms of AUC computed on the test set.

TABLE II
VALUES OF THE PARAMETERS USED IN THE EXPERIMENTS FOR IT2-PAES-RCS, PAES-RCS AND PAES-RCS-SD

AS	(2+2)M-PAES archive size	128
T_f	Number of fuzzy sets for each variable $X_f, f=1, \dots, F$	5
M_{max}	Maximum number of rules in a rule base	50
P_{CRB}	Probability of applying the crossover operator to C_{RB}	0.4
P_{CDB}	Probability of applying the crossover operator to C_{DB}	0.5
P_{MRB1}	Probability of applying the first mutation operator to C_{RB}	0.1
P_{MRB2}	Probability of applying the second mutation operator to C_{RB}	0.6
P_{MDB}	Probability of applying the mutation operator to C_{DB}	0.2

Table III shows, for each dataset, the average AUC, FPR and TPR on both the training and the test sets, the average number of rules and the average TRL for the classifiers with the highest AUC on the training set generated by IT2-PAES-RCS, PAES-RCS and PAES-RCS-SD, and for the classifiers generated by FARC-HD, FURIA, C4.5 and C4.5-CS. For each dataset, we have shown in bold the best values. We can observe that C4.5 and C4.5-CS suffer very much from overtraining. Indeed, the value of the AUC is very high on the training set, but is quite low on the test set. Although it is less evident than for C4.5 and C4.5-CS, also FURIA suffers from overtraining: the AUC computed on the test set is at least for some datasets much lower than on the training set. IT2-PAES-RCS, PAES-RCS and PAES-RCS-SD do not suffer from overtraining and show similar performance, thus testifying the validity of the three objective approach.

To statistically verify these observations, we apply non-parametric statistical tests for multiple comparisons. First, for each approach, we generate a distribution consisting of the average values of the AUCs on the test set. Then, we apply the Friedman test in order to compute a ranking among the distributions [65], and the Iman and Davenport test [66] to evaluate whether there exist statistically relevant differences

among the distributions. If there exists a statistical difference, we apply a post-hoc procedure, namely the Holm test [67]. This test allows detecting effective statistical differences between the control approach, i.e. the one with the lowest Friedman rank, and the remaining approaches.

Table IV shows the results of the non-parametric statistical tests: for each algorithm, we show the Friedman rank and the Iman and Davenport p -value. If the p -value is lower than the level of significance α (in the experiments $\alpha = 0.05$), we can reject the null hypothesis and affirm that there exist statistical differences between the multiple distributions associated with each approach. Otherwise, no statistical difference exists among the distributions and therefore the solutions are statistically equivalent. We observe that the Iman and Davenport statistical hypothesis of equivalence is rejected and so statistical differences among the six approaches are detected. Thus, we have to apply the Holm post-hoc procedure considering the PAES-RCS-SD as control algorithm (associated with the lowest rank and in bold in the Table). In the part of the table corresponding to the results obtained by the application of the Holm post-hoc procedure, the algorithms are sorted by decreasing Friedman ranks. Index i denotes the position of the algorithm in the sorted list ($i = 1$ and $i = 6$ correspond to the lowest and highest Friedman ranks, respectively). The Holm post-hoc procedure computes the z -values and p -values shown in the table: if the p -value of the algorithm in position i is lower than the adjusted α value (α / i), then the null hypothesis is rejected.

The Holm post-hoc procedure states that the AUCs on the test set of IT2-PAES-RCS and PAES-RCS are statistically equivalent to the AUC of PAES-RCS-SD. The null hypothesis is rejected for all the other algorithms. Thus, we can conclude that the three versions of PAES-RCS with three objectives obtain classifiers, which outperform the ones obtained by the other approaches in terms of AUCs. Also, this result is obtained without rebalancing the datasets. Further, if we analyze the Friedman ranks, we realize that the two algorithms with the highest ranks are just PAES-RCS-SD and IT2-PAES-RCS. Further, both PAES-RCS-SD and IT2-PAES-RCS obtain this result with classifiers characterized by a low number of rules. To verify this observation, we have also applied the non-parametric statistical tests for multiple comparisons to the number of rules and to the TRL values.

Tables V and VI show the results. Since the null hypothesis is rejected for both the tests, we apply the Holm post-hoc procedure by using IT2-PAES-RCS as control algorithm. The procedure states that, in terms of average number of rules (see Table V), the classifiers generated by PAES-RCS, PAES-RCS-SD and FURIA are statistically equivalent to the ones generated by IT2-PAES-RCS. On the contrary, the null hypothesis is rejected for FARC-HD, C4.5 and C4.5-CS. As regards TRL, the Holm post-hoc procedure concludes that the most accurate classifiers generated by IT2-PAES-RCS result to be characterized by an average TRL value statistically equivalent to the most accurate classifiers generated by PAES-RCS and to the classifiers generated by FURIA. On the

> PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

contrary, the null hypothesis is rejected for PAES-RCS-SD, FARC-HD, C4.5 and C4.5-CS.

Among the classifiers used for comparison, only FURIA shows a complexity comparable to the three versions of PAES-RCS. We have to highlight, however, that the interpretability of the classifiers generated by FURIA is limited by the membership functions computed by the method. Indeed, these membership functions are hardly describable using linguistic terms. On the contrary, thanks to the constraints imposed on

the membership function learning during the evolutionary process, the partitions generated by IT2-PAES-RCS, PAES-RCS and PAES-RCS-SD can be easily described by linguistic terms. Just to provide a glimpse of this interpretability, we consider one of the datasets in Table I, namely ARB. Table VII describes in detail the meaning of the attributes for the ARB dataset. We recall that the output here is spotting arbitrage opportunities in the LIFFE market.

TABLE III

AVERAGE AUC, FPR AND TPR ON BOTH THE TRAINING AND THE TEST SETS, AVERAGE TRL AND NUMBER OF RULES FOR THE CLASSIFIERS WITH THE HIGHEST AUC ON THE TRAINING SET GENERATED BY IT2-PAES-RCS, PAES-RCS, PAES-RCS-SD, AND FOR THE CLASSIFIERS GENERATED BY FARC-HD, FURIA, C4.5 AND C4.5-CS

		AUC _{Tr}	FPR _{Tr}	TPR _{Tr}	AUC _{ts}	FPR _{ts}	TPR _{ts}	TRL	#Rules
B L A	IT2-PAES-RCS	65.36	44.95	75.68	59.35	50.13	68.82	123.7	28.8
	PAES-RCS	68.97	41.10	79.03	59.97	48.78	68.73	159.6	34.0
	PAES-RCS-SD	69.11	41.11	79.33	58.70	50.22	67.62	192.6	37.1
	FARC-HD	68.37	36.04	72.78	59.90	42.27	62.07	598.8	203.0
	FURIA	64.39	35.76	64.53	58.37	40.16	56.89	18.4	7.6
	C4.5	90.43	10.40	91.26	57.89	38.04	53.82	445.6	223.8
	C4.5-CS	88.71	14.38	91.81	56.24	39.77	52.26	348.8	175.4
C A R D	IT2-PAES-RCS	68.57	25.74	62.87	68.51	25.74	62.77	458.1	27.4
	PAES-RCS	68.45	26.90	63.79	68.34	27.00	63.69	485.5	28.8
	PAES-RCS-SD	68.57	26.80	63.95	68.51	26.87	63.89	485.6	27.7
	FARC-HD	68.41	24.65	61.59	66.66	24.60	58.04	1852.6	628.0
	FURIA	70.96	25.29	67.20	68.00	25.36	61.35	22.4	5.6
	C4.5	94.43	4.69	93.56	65.78	28.98	60.53	33396.8	16699.4
	C4.5-CS	92.90	9.41	95.20	66.26	30.25	62.77	26145.2	13073.6
A F	IT2-PAES-RCS	62.27	28.44	52.98	54.41	32.77	41.59	364.7	33.0
	PAES-RCS	66.56	36.98	70.09	52.72	45.06	50.50	501.1	38.8
	PAES-RCS-SD	67.34	30.24	64.92	54.79	36.50	46.08	592.5	41.3
	FARC-HD	68.53	26.18	63.24	55.22	32.29	42.73	768.7	264.0
	FURIA	66.59	35.90	69.07	53.90	39.28	47.08	31.8	11.6
	C4.5	88.14	10.16	86.43	54.06	32.32	40.45	508.3	255.4
	C4.5-CS	87.37	21.72	96.46	50.95	43.29	45.20	412.4	207.2
A R B	IT2-PAES-RCS	94.73	6.88	96.33	94.25	7.04	95.55	51.4	21.1
	PAES-RCS	97.21	3.01	97.43	97.02	3.17	97.21	49.0	20.3
	PAES-RCS-SD	95.23	6.12	96.52	94.69	6.49	95.88	47.9	20.0
	FARC-HD	86.15	15.49	87.80	87.28	15.91	90.48	35.5	16.6
	FURIA	98.23	2.28	98.73	98.14	3.23	99.50	60.0	26.0
	C4.5	98.64	2.16	99.44	98.18	3.39	99.75	78.4	40.2
	C4.5-CS	98.73	2.42	99.88	98.37	2.50	99.25	32.4	17.2
C O M M	IT2-PAES-RCS	67.19	18.76	53.13	66.44	19.08	51.96	115.4	27.5
	PAES-RCS	67.42	17.59	52.44	66.05	18.17	50.26	145.1	30.9
	PAES-RCS-SD	67.33	17.38	52.04	66.33	17.79	50.44	167.6	35.4
	FARC-HD	76.77	10.03	63.57	65.86	10.18	41.91	313.3	118.8
	FURIA	80.47	9.33	70.27	65.14	9.76	40.05	73.6	20.8
	C4.5	93.91	4.36	92.18	61.04	22.18	44.26	2470.0	1236.0
	C4.5-CS	93.36	13.09	99.82	61.62	28.66	51.90	2139.6	1070.8
S L	IT2-PAES-RCS	60.03	34.50	54.56	59.64	34.61	53.90	249.4	21.9
	PAES-RCS	60.35	38.23	58.93	59.58	38.20	57.37	306.1	24.1
	PAES-RCS-SD	61.20	40.04	62.43	59.99	40.51	60.48	423.4	32.2
	FARC-HD	65.24	33.08	63.56	58.80	33.22	50.82	2154.6	720.6
	FURIA	69.20	36.42	74.82	58.13	37.22	53.48	70.2	19.2
	C4.5	93.72	4.53	91.97	54.15	22.36	30.67	7962.8	3982.4
	C4.5-CS	94.94	9.81	99.69	54.83	26.36	36.01	7836.8	3301.8
L E N	IT2-PAES-RCS	63.92	36.11	63.95	63.10	36.39	62.58	191.8	42.8
	PAES-RCS	64.14	35.95	64.24	62.96	36.42	62.33	236.6	50.6
	PAES-RCS-SD	64.42	36.61	65.46	63.29	37.07	63.64	300.4	59.2
	FARC-HD	70.59	23.77	64.96	59.34	24.08	42.75	1855.9	644.4

> PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

	FURIA	83.86	5.04	72.76	52.98	5.55	11.51	202.4	49.2
	C4.5	92.79	3.46	89.03	55.09	18.46	28.64	3987.2	1994.6
	C4.5-CS	94.91	10.15	99.97	56.35	25.57	38.28	3776.4	1889.2
D P K G	IT2-PAES-RCS	67.48	22.53	57.50	67.17	22.58	56.91	94.2	23.4
	PAES-RCS	67.89	23.46	59.24	67.50	23.62	58.62	102.4	24.4
	PAES-RCS-SD	67.88	23.72	59.48	67.53	23.74	58.80	138.2	29.5
	FARC-HD	70.66	27.48	68.79	64.97	27.54	57.48	1959.7	664.3
B A N	FURIA	81.67	14.11	77.44	63.58	14.55	41.71	560.8	100.4
	C4.5	96.50	1.88	94.88	75.23	8.10	58.56	9379.2	4690.6
	C4.5-CS	97.83	4.28	99.94	78.03	9.80	65.87	7111.2	3556.6
	IT2-PAES-RCS	80.85	16.85	78.56	80.59	17.24	78.44	111.5	28.3
G I V	PAES-RCS	80.26	22.44	82.96	79.76	23.40	82.93	98.6	24.6
	PAES-RCS-SD	81.02	16.73	78.77	80.76	17.20	78.74	114.0	30.2
	FARC-HD	85.41	13.82	84.64	77.24	30.07	84.55	810.9	331.2
	FURIA	92.96	9.12	95.04	77.61	39.09	94.31	233.4	41.8
C O I	C4.5	95.34	5.77	96.46	70.24	53.11	93.59	1932.8	966.6
	C4.5-CS	98.73	2.42	99.88	98.37	2.50	99.25	273.5	17.2
	IT2-PAES-RCS	72.59	17.68	62.87	72.54	17.69	62.77	40.3	19.8
	PAES-RCS	68.14	13.64	49.92	68.07	13.63	49.77	27.8	14.9
G I V	PAES-RCS-SD	72.77	17.47	63.01	72.67	17.44	62.78	43.1	19.9
	FARC-HD	73.44	13.07	59.96	66.51	13.05	46.07	253.2	87.6
	FURIA	78.35	14.46	71.15	73.26	14.53	61.04	129.6	19.8
	C4.5	86.38	10.03	82.79	70.53	13.40	54.46	7045.6	3523.8
C O I	C4.5-CS	95.99	8.03	100.00	71.51	11.82	54.84	9162.4	4582.2
	IT2-PAES-RCS	66.36	34.84	67.57	63.93	39.94	67.80	55.87	22.1
	PAES-RCS	67.88	31.45	67.21	62.62	41.36	66.68	92.63	29.9
	PAES-RCS-SD	66.87	32.94	66.69	63.74	39.25	66.74	85.5	27.3
C O I	FARC-HD	65.51	32.40	66.20	61.74	38.90	65.45	1304.1	457.6
	FURIA	95.50	7.10	98.10	62.13	73.29	97.54	510.5	91.4
	C4.5	97.60	3.68	98.89	61.34	73.48	96.17	596.4	299.2
	C4.5-CS	95.83	3.50	91.65	60.43	66.91	87.78	772.4	387.2

TABLE IV

RESULTS OF THE NON-PARAMETRIC STATISTICAL TESTS ON THE AUC COMPUTED ON THE TEST SET AMONG THE CLASSIFIERS WITH THE HIGHEST AUC ON THE TRAINING SET GENERATED BY IT2-PAES-RCS, PAES-RCS AND PAES-RCS-SD, AND THE CLASSIFIERS GENERATED BY FARC-HD, FURIA, C4.5 AND C4.5-CS

Algorithm	Friedman Rank	Iman and Davenport p-value	Hypothesis	
PAES-RCS-SD	2.3182			
IT2-PAES-RCS	2.7727			
PAES-RCS	3.6364			
FARC-HD	4.6364	0.003	Rejected	
FURIA	4.6364			
C4.5	4.6364			
C4.5-CS	5.3636			
<i>Holm post-hoc procedure</i>				
Algorithm	z-value	p-value	alpha/i	Hypothesis
C4.5-CS	3.30	9.46E-2	0.0083	Rejected
FURIA	2.51	1.18E-3	0.0100	Rejected
FARC-HD	2.51	1.18E-3	0.0125	Rejected
C4.5-CS	2.51	1.18E-3	0.0166	Rejected
PAES-RCS	1.43	1.57E-1	0.0250	Not Rejected
IT2-PAES-RCS	0.49	6.21E-1	0.0500	Not Rejected

Fig. 8 shows an example of partitions generated by IT2-PAES-RCS for one of the classifiers with the highest AUC on the training set for the ARB. Here, only 6 out of 7 attributes are

shown since one of the attributes was not used in the final rule base. We can observe that, although the evolutionary process has tuned the IT2 fuzzy sets on the specific dataset, the partitions of the different attributes result to be easily interpretable.

TABLE V

RESULTS OF THE NON-PARAMETRIC STATISTICAL TESTS ON THE NUMBER OF RULES AMONG THE CLASSIFIERS WITH THE HIGHEST AUC ON THE TRAINING SET GENERATED BY IT2-PAES-RCS, PAES-RCS AND PAES-RCS-SD, AND THE CLASSIFIERS GENERATED BY FARC-HD, FURIA, C4.5 AND C4.5-CS

Algorithm	Friedman Rank	Iman and Davenport p-value	Hypothesis	
IT2-PAES-RCS	2.1364			
FURIA	2.5909			
PAES-RCS	2.8182			
PAES-RCS-SD	3.5448	0.3E-9	Rejected	
C4.5-CS	5.0909			
FARC-HD	5.1818			
C4.5	6.6364			
<i>Holm post-hoc Procedure</i>				
Algorithm	z-value	p-value	alpha/i	Hypothesis
C4.5	4.88	1.00E-6	0.0015	Rejected
FARC-HD	3.30	9.46E-4	0.003	Rejected
C4.5-CS	3.20	1.13E-3	0.006	Rejected
PAES-RCS-SD	1.52	1.26E-1	0.0125	Not Rejected
PAES-RCS	0.74	4.59E-1	0.025	Not Rejected
FURIA	0.49	6.21E-1	0.05	Not Rejected

> PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <

As regards the interpretability of the rules, Fig. 9 shows the rule base of the classifier whose data base is shown in Fig. 8. Here, we do not show the “don’t care” conditions since they do not contribute to the inference process and penalize the interpretability of the rule base. The expert can deduce interesting knowledge from the rules of the classifier. Indeed, he/she can, for instance, discover that intermediate values of $C-P$ ($C-P$ is M) lead to conclude that the class is Arbitrage Opportunity. On the other hand, very high values of $Futures$ ($Futures$ is VH) allow inferring that the class is non Arbitrage Opportunity.

TABLE VI

RESULTS OF THE NON-PARAMETRIC STATISTICAL TESTS ON THE TRL AMONG THE CLASSIFIERS WITH THE HIGHEST AUC ON THE TRAINING SET GENERATED BY IT2-PAES-RCS, PAES-RCS AND PAES-RCS-SD, AND THE CLASSIFIERS GENERATED BY FARC-HD, FURIA, C4.5 AND C4.5-CS

Algorithm	Friedman Rank	Iman and Davenport p-value	Hypothesis
IT2-PAES-RCS	2.0000		
FURIA	2.6364		
PAES-RCS	2.7273		
PAES-RCS-SD	3.6364	0.17E-9	Rejected
C4.5-CS	5.1818		
FARC-HD	5.3636		
C4.5	6.6364		

Holm post-hoc Procedure				
Algorithm	z-value	p-value	alpha/i	Hypothesis
C4.5	4.83	0.1E-5	0.0015	Rejected
FARC-HD	3.65	2.1E-4	0.003	Rejected
C4.5-CS	3.45	5.1E-4	0.006	Rejected
PAES-RCS-SD	1.77	7.6E-3	0.0125	Rejected
PAES-RCS	0.78	4.29E-1	0.025	Not Rejected
FURIA	0.69	4.89E-1	0.05	Not Rejected

TABLE VII

MEANING OF THE ATTRIBUTES OF THE ARB DATASET

Name	Description
MoneyNess	Strike Price/Underlying Index Level
Basis % (x10000)	Futures price minus spot index level, divided by futures price, multiplied by 10,000
Und (x10)	Spot index level divided by futures price, multiplied by 10
Interest Ask %	The LIBOR ask rate for the maturity closest to the maturity of futures contract, multiplied by 100
Futures (T-t)	The nave trigger, profit after transaction costs, divided by futures price, multiplied by 1,000,000
C-P % (x100)	The difference between the call and the put prices, divided by futures price
Profit after TC (x 1,000,000)	The nave trigger, profit after transaction costs, divided by futures price, multiplied by 1,000,000

Table VIII shows the results of the test. The null hypothesis is not rejected for the AUC computed on the test set, but is rejected for the average number of rules and average TRL. We can conclude that IT2-PAES-RCS generated classifiers that achieve the same accuracy in terms of AUC as PAES-RCS and PAES-RCS-SD, but with a lower number of rules and a lower TRL. Thus, the classifiers generated by IT2-PAES-RCS result to be less complex and therefore more interpretable.

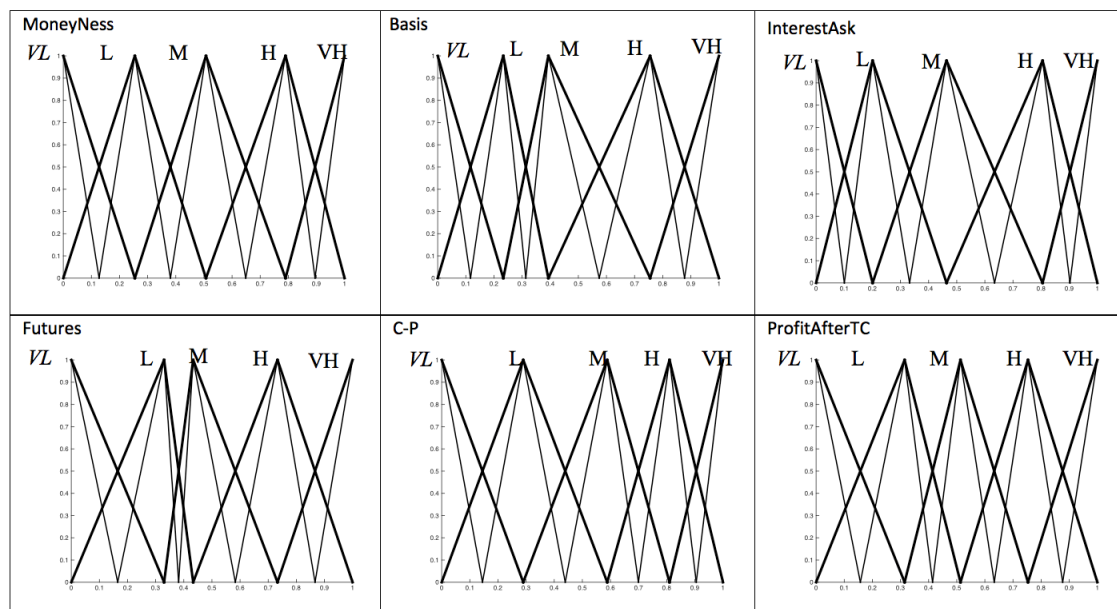


Figure 8. An example of partitions generated by IT2-PAES-RCS for one of the classifiers with the highest AUC on the training set for the dataset ARB.

<p>R_1: IF C-P is VL and ProfitAfterTC is VL THEN Class Y is Non Arbitrage Opportunity R_2: IF Basis is M and InterestAsk is L and Futures is L and ProfitAfterTC is L THEN Class Y is Non Arbitrage Opportunity R_3: IF Basis is H and C-P is L THEN Class Y is Non Arbitrage Opportunity R_4: IF InterestAsk is M and Futures is L THEN Class Y is Non Arbitrage Opportunity R_5: IF Futures is VH THEN Class Y is Non Arbitrage Opportunity R_6: IF ProfitAfterTC is L THEN Class Y is Non Arbitrage Opportunity R_7: IF Basis is M and Futures is VL THEN Class Y is Arbitrage Opportunity R_8: IF InterestAsk is M and Futures is L and C-P is L and ProfitAfterTC is VL THEN Class Y is Arbitrage Opportunity R_9: IF Futures is L and C-P is M THEN Class Y is Arbitrage Opportunity R_{10}: IF MoneyNess is L and C-P is M THEN Class Y is Arbitrage Opportunity R_{11}: IF MoneyNess is M and Futures is L and C-P is M and ProfitAfterTC is VL THEN Class Y is Arbitrage Opportunity</p>
--

Figure 9. The rule base of the classifier whose data base is shown in Fig. 8.

TABLE VIII

RESULTS OF THE WILCOXON SIGNED-RANK TEST ON AUC, TRL AND NUMBER OF RULES AMONG THE CLASSIFIERS WITH THE HIGHEST AUC ON THE TRAINING SET GENERATED BY IT2-PAES-RCS, PAES-RCS AND PAES-RCS-SD

AUC_{T_s}	R+	R-	p-value	Hypothesis ($\alpha=0.05$)
IT2-PAES-RCS vs. PAES-RCS	46.0	20.0	0.230	Not Rejected
IT2-PAES-RCS vs. PAES-RCS-SD	15.5	39.5	1	Not Rejected
TRL	R+	R-	p-value	Hypothesis ($\alpha=0.05$)
IT2-PAES-RCS vs. PAES-RCS	58.0	8.0	0.023	Rejected
IT2-PAES-RCS vs. PAES-RCS-SD	63.0	3.0	0.006	Rejected
#Rules	R+	R-	p-value	Hypothesis ($\alpha=0.05$)
IT2-PAES-RCS vs. PAES-RCS	58.0	8.0	0.023	Rejected
IT2-PAES-RCS vs. PAES-RCS-SD	63.0	3.0	0.006	Rejected

V. CONCLUSIONS

Financial data are often strongly unbalanced and characterized by a high level of uncertainty. In this paper, we have proposed to deal with financial data classification by adopting rule-based classifiers generated by a multi-objective evolutionary algorithm (MOEA). These classifiers have proved to be very effective in terms of accuracy. Further, they are generally characterized by a low number of rules and total rule length, and a good integrity of the partitions, thus making them very interpretable. Interpretability is considered essential in the financial context since the comprehension of how inputs and output are related to each other is crucial to take both operative and strategic decisions.

We have extended PAES-RCS, an MOEA-based approach to learn concurrently the rule and data bases of fuzzy rule-based classifiers. In order to cope with unbalanced datasets, we have split the accuracy into two objectives, namely True Positive Rate and False Positive Rate, and we have used an approach denoted as scaled dominance to give minority classes a fair chance when competing with a majority class. Further, we have coped with uncertainty by adopting IT2 fuzzy sets rather than type-1 fuzzy sets. This has required using a different inference mechanism. We have tested the three improvements on eleven financial datasets and compared the results with the ones obtained by the fuzzy rule-based classifiers FARC-HD and FURIA, the classical C4.5 decision tree algorithm and its version cost-sensitive. Using non-parametric statistical tests, we have shown that the three

improvements allow generating classifiers, which outperform the comparison approaches both in terms of accuracy, computed as area under the curve, and complexity, computed as number of rules. Finally, the extension of PAES-RCS, which integrates the three improvements, has proved to achieve high accuracy with, on average, the lowest number of rules and total rule length.

REFERENCES

- 1) <http://butleranalytics.com/regulatory-opportunity-financial-services/>
- 2) <http://butleranalytics.com/predictive-models-risks-benefits/>
- 3) <http://butleranalytics.com/financial-regulation-predictive-models/>
- 4) A. Garcia-Almanza and E. Tsang, "Forecasting stock prices using genetic programming and chance discovery," Proceedings of the 12th International Conference on Computing in Economics and Finance, Limassol, Cyprus, June 22-24, 2006.
- 5) A. Garcia-Almanza and E. Tsang, "Evolving decision rules to predict investment opportunities," *The International Journal of Automation and Computers*, vol. 5, no. 1, pp. 22-31, 2008.
- 6) A. Garcia-Almanza, "New classification methods for gathering patterns in the context of genetic programming," PhD Thesis, Department of Computing and Electronic Systems, University of Essex, 2008.
- 7) L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- 8) J. Casillas, O. Cordon, F. Herrera, L. Magdalena (Eds.), "Interpretability issues in fuzzy modelling," Springer, 2003.
- 9) M. Setnes, H. Roubos, "GA-fuzzy modelling and classification: complexity and performance," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 5, pp. 509-522, 2000.
- 10) H. Ishibuchi, T. Yamamoto, "Interpretability Issues in Fuzzy Genetic-Based Machine Learning for Linguistic Modelling," in *Modelling with Words*, Lecture notes in Computer Science, vol. 2873, pp. 209-228, 2003.
- 11) H. Ishibuchi, T. Nakashima and M. Nii, *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining*, Berlin, Germany: Springer-Verlag, 2004.
- 12) P. Ducange and F. Marcelloni, "Multi-objective evolutionary fuzzy systems," 9th International Workshop on Fuzzy Logic and Applications, Trani, Italia, 29-31 Agosto, Lecture Notes in Computer Science, vol. 6857, pp. 83-90, 2011.
- 13) M. Fazzolari, R. Alcalá, Y. Nojima, H. Ishibuchi and F. Herrera, "A review of the application of multiobjective evolutionary fuzzy systems: Current status and further directions," *IEEE Transactions on Fuzzy Systems*, vol. 21, no.1, pp.45-65, 2013.
- 14) P. Melin and O. Castillo, "A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition," *Applied Soft Computing*, vol. 21, pp. 568-577, 2014.
- 15) C. Glackin, L. Maguire, R. McIvor, P. Humphreys and P. Herman, "A comparison of fuzzy strategies for corporate acquisition analysis," *Fuzzy Sets and Systems*, vol. 159, no. 18, pp. 2039-2056, 2007.
- 16) D. Bernardo, H. Hagrais and E. Tsang, "A Genetic Type-2 fuzzy logic based system for the generation of summarised linguistic predictive

- models for financial applications,” *Soft Computing*, vol. 17, no. 12, pp. 2185-2201, 2013.
- [17] J.A Sanz, D. Bernardo, F. Herrera, H. Bustince and H. Hagrás, “A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data,” *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 4, pp.973-990, 2015.
- [18] D. Wu, W. Tan, “Genetic learning and performance evaluation of interval type-2 fuzzy logic controllers,” *Engineering Applications of Artificial Intelligence*, vol. 19, no. 8, pp. 819-841, 2006.
- [19] [19] R. Martínez-Soto, O. Castillo, L.T. Aguilar and A.R. Díaz, “A hybrid optimization method with PSO and GA to automatically design Type-1 and Type-2 fuzzy logic controllers,” *Int. J. Machine Learning & Cybernetics*, vol. 6, no. 2, pp. 175-196, 2015.
- [20] [20] O. Linda and M. Manic, “Uncertainty-robust design of interval type-2 fuzzy logic controller for delta parallel robot,” *IEEE Transactions on Industrial Informatics*, vol. 7, no. 4, pp. 661-670, 2011.
- [21] [21] M. Almarahi, R. John, S. Ahmadi, “Learning of type-2 fuzzy logic systems by simulated annealing with adaptive step size,” *Lecture Notes in Electrical Engineering*, vol. 130, pp. 53-64, 2013.
- [22] [22] T. Kumbasar and H. Hagrás, “Big bang-big crunch optimization based Interval type-2 fuzzy PID cascade controller design strategy,” *Information Sciences*, vol. 282, pp. 277-295, 2014.
- [23] J. Mendel, H. Hagrás, W-W. Tan, W. Melek and H. Ying, “Introduction to type-2 fuzzy logic control: Theory and applications,” Wiley-IEEE Press, 2014.
- [24] G.G. Sundarkumar and V. Ravi, “A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance,” *Engineering Applications of Artificial Intelligence*, vol. 36, pp. 368-377, 2015.
- [25] M. Antonelli, P. Ducange, F. Marcelloni, “Multi-objective evolutionary rule and condition selection for designing fuzzy rule-based classifiers,” 2012 IEEE International Conference on Fuzzy Systems, Brisbane, Australia, June 10-15, 2012, pp. 794-800.
- [26] M. Antonelli, P. Ducange, F. Marcelloni, “A fast and efficient multi-objective evolutionary learning scheme for fuzzy rule-based classifiers,” *Information Sciences*, vol. 283, pp. 36-54, 2014.
- [27] M. Antonelli, P. Ducange, F. Marcelloni, “An experimental study on evolutionary fuzzy classifiers designed for managing imbalanced datasets,” *Neurocomputing*, vol. 146, pp. 125-136, 2014.
- [28] 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 Transactions on Fuzzy Systems*, vol. 19, no. 5, pp. 857-872, 2011.
- [29] J. Huhn, E. Hullermeier, “FURIA: an algorithm for unordered fuzzy rule induction,” *Data mining and knowledge discovery*, vol. 19, no. 3, pp. 293-319, 2009.
- [30] J. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufman, 1993.
- [31] K.M. Ting, “An instance-weighting method to induce cost-sensitive trees,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 3, pp. 659-665, 2002.
- [32] Q. Liang, J. M. Mendel, “Interval Type-2 Fuzzy Logic Systems: Theory and Design,” *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 5, pp. 535-550, 2000.
- [33] D. Bernardo, H. Hagrás, E. Tsang “An interval type-2 fuzzy logic system for the modelling and prediction of financial applications”, Proceedings of the International Conference on Autonomous and Intelligent Systems (AIS 2012), Aveiro, Portugal, June 25-27, 2012.
- [34] D. Bernardo, H. Hagrás, E. Tsang “An interval type-2 fuzzy logic based system for model generation and summarization of arbitrage opportunities in stock markets,” Proceedings of the UK Workshop on Computational Intelligence (UKCI 2012), Edinburgh, Scotland, September 5-7 2012.
- [35] D. Bernardo, H. Hagrás, E. Tsang “A genetic type-2 fuzzy logic based system for financial applications modelling and prediction,” Proceeding of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2013), Hyderabad India, 7-10 July 2013.
- [36] A. Zhou, B-Y. Qu, H. Li, S-Z. Zhao, P.N. Suganthan, Q. Zhang, “Multiobjective evolutionary algorithms: A survey of the state of the art,” *Swarm and Evolutionary Computation*, vol. 1, No 1, pp 32-49, 2011.
- [37] P. Ducange, B. Lazzerini, F. Marcelloni, “Multi-objective genetic fuzzy classifiers for imbalanced and cost-sensitive datasets,” *Soft Computing*, vol. 14, no. 10, pp. 713-728, 2010.
- [38] F. Herrera, “Genetic fuzzy systems: taxonomy, current research trends and prospects,” *Evolutionary Intelligence*, vol. 1, no. 1, pp. 27-46, 2008.
- [39] J. M. Alonso and L. Magdalena, “Editorial: Special issue on interpretable fuzzy systems,” *Information Sciences*, vol. 181, no. 20, pp. 4331-4339, 2011.
- [40] M.J. Gacto, R. Alcalá, F. Herrera, “Interpretability of linguistic fuzzy rule-based systems: an overview of interpretability measures,” *Information Sciences*, vol.128, no. 20, pp. 4340-4360, 2011.
- [41] M. Cococcioni, P. Ducange, B. Lazzerini, and F. Marcelloni, “A Pareto-based multi-objective evolutionary approach to the identification of mamdani fuzzy systems,” *Soft Computing*, vol. 11, no. 11, pp. 1013-1031, 2007.
- [42] A. Botta, B. Lazzerini, F. Marcelloni, D. Stefanescu, “Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index,” *Soft Computing*, vol. 13, no. 5, pp. 437-449, 2009.
- [43] R. Alcalá, P. Ducange, F. Herrera, B. Lazzerini, F. Marcelloni, “A multi-objective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy rule-based systems,” *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 5, pp 1106-1122, 2009.
- [44] M. Antonelli, P. Ducange, B. Lazzerini, F. Marcelloni, “Multi-objective evolutionary learning of granularity, membership function parameters and rules of Mamdani fuzzy systems,” *Evolutionary Intelligence*, vol. 2, no. 1-2, pp. 21-37, 2009.
- [45] J. Casillas, P. Martinez, A. D. Benitez, “Learning consistent, complete and compact sets of fuzzy rules in conjunctive normal form for regression problems,” *Soft Computing*, vol. 13, no. 5, pp. 451-465, 2009.
- [46] P. Pulkkinen and H. Koivisto, “A dynamically constrained multiobjective genetic fuzzy system for regression problems,” *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 1, pp. 161-177, 2010.
- [47] M. Antonelli, P. Ducange, B. Lazzerini and F. Marcelloni, “Learning concurrently data and rule bases of Mamdani fuzzy rule-based systems by exploiting a novel interpretability index,” *Soft Computing*, vol. 15, no. 10, pp. 1981-1998, 2011.
- [48] M. Antonelli, P. Ducange, B. Lazzerini and F. Marcelloni, “Learning knowledge bases of multi-objective evolutionary fuzzy systems by simultaneously optimizing accuracy, complexity and partition integrity,” *Soft Computing*, vol. 15, no. 12, pp. 2335-2354, 2011.
- [49] M. J. Gacto, R. Alcalá and F. Herrera, “Integration of an index to preserve the semantic interpretability in the multi-objective evolutionary rule selection and tuning of linguistic fuzzy systems,” *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 3, pp. 515-531, 2010.
- [50] R. Alcalá, M.J. Gacto and F. Herrera, “A fast and scalable multiobjective genetic fuzzy system for linguistic fuzzy modeling in high-dimensional regression problems,” *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 4, pp. 666-681, 2011.
- [51] M. Antonelli, P. Ducange and F. Marcelloni, “Genetic training instance selection in multi-objective evolutionary fuzzy systems: a co-evolutionary approach,” *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 2, pp. 276-290, 2012.
- [52] H. Ishibuchi, T. Murata and I. B. Turksen, “Single-objective and two objective genetic algorithms for selecting linguistic rules for pattern classification problems,” *Fuzzy Sets and Systems*, vol. 89, no. 2, pp. 135-150, 1997.
- [53] O. Cordon, M. J. del Jesus, J. Casillas, F. Herrera, L. Magdalena and P. Villar, “A multiobjective genetic learning process for joint feature selection and granularity and context learning in fuzzy rule-based classification systems,” in Interpretability Issues in Fuzzy Modeling, J. Casillas, F. Herrera, O. Cordoon, and L. Magdalena, Eds. Secaucus, NJ, USA: Springer-Verlag, pp. 79-99, 2003.
- [54] H. Ishibuchi and T. Yamamoto, “Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining,” *Fuzzy Sets and Systems*, vol. 141, pp. 59-88, 2004.

- [55] H. Ishibuchi and Y. Nojima, "Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning," *International Journal of Approximate Reasoning*, vol. 44, no. 1, pp. 4–31, 2007.
- [56] R. Alcalá, Y. Nojima, F. Herrera and H. Ishibuchi, "Multiobjective genetic fuzzy rule selection of single granularity-based fuzzy classification rules and its interaction with the lateral tuning of membership functions," *Soft Computing*, vol. 15, no. 12, pp. 2303–2318, 2011.
- [57] P. Pulkkinen and H. Koivisto, "Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms," *International Journal of Approximate Reasoning*, vol. 48, pp. 526–543, 2008.
- [58] K. Trawinski, O. Cordon and A. Quirin, "A Study on the use of multiobjective genetic algorithms for classifier selection in FURIA-based fuzzy multiclassifiers," *International Journal of Computational Intelligence Systems*, vol. 5, no. 2, pp. 231-253, 2012.
- [59] O. Cordon, M. J. del Jesus and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems," *International Journal of Approximate Reasoning*, vol. 20, no. 1, pp. 21-45, 1999.
- [60] W. W. Cohen, "Fast effective rule induction," in: Proceedings of the Twelfth International Conference on Machine Learning, Morgan Kaufmann, 1995, pp. 115-123.
- [61] J.R. Quinlan, "Boosting, Bagging, and C4.5," Proc. of the 13th national conference on Artificial Intelligence, vol. 1, pp. 725-730, 1996.
- [62] N.V. Chawla, K.W. Bowyer, L.O. Hall and W.P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [63] J. Alcalá-Fdez, A. Fernandez, J. Luengo, J. Derrac and S. Garcia, "KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," *Multiple-Valued Logic and Soft Computing*, vol. 2-3, pp. 255-287, 2011.
- [64] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874, 2006.
- [65] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, vol. 32, pp. 675–701, 1937.
- [66] R. L. Iman, J. H. Davenport, "Approximations of the critical region of the Friedman statistic," *Comm. Statist. Part A Theory Methods*, vol. 9, pp. 571–595, 1980.
- [67] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Statist.*, vol. 6, pp. 65–70, 1979.
- [68] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 4th Edition, Chapman & Hall/CRC, 2007.



Michela Antonelli received the M.Sc. degree in Computer Engineering and the Ph.D. degree in Information Engineering from the University of Pisa (Italy), in 2003 and 2007, respectively. She has been a research fellow of the Computational Intelligence Group at the Department of Information Engineering, University of Pisa from 2008 to 2014. Currently, she is

a research associate in the Centre for Medical Image Computing at University College London.

Her main research interests are in the area of Computational Intelligence, with particular emphasis to fuzzy systems and multi-objective evolutionary algorithms.



Dario Bernardo received the B.Sc. and M.Sc. degrees in computer engineering for enterprise management from the University of Pisa, Pisa, Italy, and another M.Sc. degree in computational and software techniques in engineering from

Cranfield University, Cranfield, U.K. He is currently pursuing the Ph.D. degree in computer science with the University of Essex, Colchester, U.K., during which he has been a Knowledge Partnership Associate. He is currently a Data Scientist working in London, U.K., in the field of predictive analytics in finance. His major research interests include computational intelligence, machine learning, type-2 fuzzy systems, fuzzy logic, genetic algorithms, and evolutionary computation. His research interests also include big data analytics, algorithm scalability, and optimization.



Hani Hagra (M'03–SM'05, F'13) received the B.Sc. and M.Sc. degrees in electrical engineering from Alexandria University, Alexandria, Egypt, and the Ph.D. degree in computer science from the University of Essex, Colchester, U.K. He is a Professor in the School of Computer Science and Electronic Engineering, Director of the

Computational Intelligence Centre and the Head of the Fuzzy Systems Research Group in the University of Essex, UK. His major research interests are in computational intelligence, notably type-2 fuzzy systems, fuzzy logic, neural networks, genetic algorithms, and evolutionary computation. His research interests also include ambient intelligence, pervasive computing and intelligent buildings. He is also interested in embedded agents, robotics and intelligent control. He has authored more than 300 papers in international journals, conferences and books. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) and he is also a Fellow of the Institution of Engineering and Technology (IET) (IIE). He was the Chair of IEEE Computational Intelligence Society (CIS) Senior Members Sub-Committee. His research has won numerous prestigious international awards where most recently he was awarded by the IEEE Computational Intelligence Society (CIS), the 2013 Outstanding Paper Award in the IEEE Transactions on Fuzzy Systems and he was also awarded the 2006 Outstanding Paper Award in the IEEE Transactions on Fuzzy Systems. He is an Associate Editor of the IEEE Transactions on Fuzzy Systems. He is also an Associate Editor of the International Journal of Robotics and Automation, the Journal of Cognitive Computation and the Journal of Ambient Computing and Intelligence. He is a member of the IEEE Computational Intelligence Society (CIS) Fuzzy Systems Technical Committee and IEEE CIS conference committee. Prof. Hagra chaired several international conferences where he will act as the Programme Chair of the 2017 IEEE International Conference on Fuzzy Systems, Naples, Italy, July 2017 and he served as the General Co-Chair of the 2007 IEEE International Conference on Fuzzy systems London.



Francesco Marcelloni (M'06) received the Laurea degree in Electronics Engineering and the Ph.D. degree in Computer Engineering from the University of Pisa in 1991 and 1996, respectively. He is currently a full

professor at the Department of Information Engineering of the University of Pisa. He has co-founded the Computational Intelligence Group in 2002 and is the founder and head of the Competence Centre on MOBILE Value Added Services (MOVAS), both at the Department of Information Engineering of the University of Pisa.

His main research interests include fuzzy classifiers for big data, multi-objective evolutionary algorithms, genetic fuzzy systems, fuzzy clustering algorithms, pattern recognition, signal analysis, neural networks, mobile information systems, and data compression and aggregation in wireless sensor

networks. He has co-edited three volumes, four journal special issues, and is (co-)author of a book and of more than 200 papers in international journals, books and conference proceedings. He has been TPC co-chair, general co-chair and tutorial chair of some international conferences and has held invited talks in a number of events. Currently, he serves as associate editor of Information Sciences (Elsevier) and Soft Computing (Springer), and is on the editorial board of a number of other international journals.