Chapter 12
Evolutionary Fuzzy Systems: A Case Study in Imbalanced Classification

A. Fernández and F. Herrera

Abstract The use of evolutionary algorithms for designing fuzzy systems provides them with learning and adaptation capabilities, resulting on what is known as Evolutionary Fuzzy Systems. These types of systems have been successfully applied in several areas of Data Mining, including standard classification, regression problems and frequent pattern mining. This is due to their ability to adapt their working procedure independently of the context we are addressing. Specifically, Evolutionary Fuzzy Systems have been lately applied to a new classification problem showing good and accurate results. We are referring to the problem of classification with imbalanced datasets, which is basically defined by an uneven distribution between the instances of the classes. In this work, we will first introduce some basic concepts on linguistic fuzzy rule based systems. Then, we will present a complete taxonomy for Evolutionary Fuzzy Systems. Then, we will review several significant proposals made in this research area that have been developed for addressing classification with imbalanced datasets. Finally, we will show a case study from which we will highlight the good behavior of Evolutionary Fuzzy Systems in this particular context.

12.1 Introduction

Among all available strategies to be used in real application areas of engineering, those related to Computational Intelligence or Soft Computing have typically shown a good behavior [1]. In addition, we must stress that the collaboration among the components of Computational Intelligence can further improve the results than applying them on isolation. For this reason, hybrid approaches have attracted considerable attention in this community. Among them, the most popular is maybe the synergy

A. Fernández
Department of Computer Science, University of Jaén, 23071 Jaén, Spain
e-mail: alberto.fernandez@ujaen.es

F. Herrera (✉)
Department of Computer Science and Artificial Intelligence,
University of Granada, 18071 Granada, Spain
e-mail: herrera@decsai.ugr.es

T. Calvo Sánchez and J. Torrens Sastre (eds.), *Fuzzy Logic and Information Fusion*,
Studies in Fuzziness and Soft Computing 339, DOI 10.1007/978-3-319-30421-2_12

between Fuzzy Rule Based Systems (FRBSs) [2] and Evolutionary Computation [3, 4] leading to Evolutionary Fuzzy Systems (EFSs) [5, 6].

The automatic definition of an FRBS can be seen as an optimization or search problem. Regarding the former, the capabilities of Evolutionary Algorithms (EAs) [7] makes them an appropriate global search technique. They aim to explore a large search space for suitable solutions, only requiring a performance measure. In addition to their ability to find near optimal solutions in complex search spaces, the generic code structure and independent performance features of EAs allow them to incorporate a priori knowledge. In the case of FRBSs, this a priori knowledge may be in the form of linguistic variables, fuzzy membership function parameters, fuzzy rules, number of rules and so on. Furthermore, this approach has been recently extended by using Multi-Objective Evolutionary Algorithms (MOEAs) [8, 9], which can consider multiple conflicting objectives, instead of a single one. The hybridization between MOEAs and FRBSs is currently known as Multi-Objective Evolutionary Fuzzy Systems (MOEFSs) [10].

As stated previously, the adapting capabilities and goodness of EFSs has made their use to be spread successfully into different Data Mining areas [11]. Among these, possibly the most common application is related to classification problems. When working in this framework, we may find that they frequently present a very different distribution of examples inside their classes, which is known as the problem of imbalanced datasets [12, 13]. In the context of binary problems, the positive or minority class is represented by few examples, since it could represent a "rare case" or because the acquisition of this data is costly. For these reasons, the minority class is usually the main objective from the learning point of view. Therefore, the cost related to a poor classification of one example of this class is usually be greater than on the majority class.

Linguistic FRBSs have shown the achievement of a good performance in the context of classification with imbalanced datasets [14]. Specifically, linguistic fuzzy sets allow the smoothing of the borderline areas in the inference process, which is also a desirable behavior in the scenario of overlapping, which is known to highly degrade the performance in this context [15, 16]. In accordance with the former, and with aims at improving the behaviour and performance of these systems, a wide number of approaches have been proposed in the field of EFS for addressing classification with imbalanced datasets [15, 17, 18].

In this chapter, we will first introduce the existent taxonomy for the different types of EFSs, together with their main properties. Then, we focus on the main aim in this contribution, which is to present the use of EFSs in imbalanced classification, and to provide a list of the most relevant contributions in this area of work. Finally, we will show the goodness of this type of approaches presenting a case study on the topic over highly imbalanced datasets using the GP-COACH-H algorithm [19], a fuzzy rule-based technique based on genetic programming specifically designed to address the imbalance in data.

For achieving these objectives, the remainder of this chapter is organized as follows. In Sect. 12.2, we provide an overview of FRBSs. In Sect. 12.3, we focus our attention to EFSs. Section 12.4 is devoted to the application of EFSs in classification with imbalanced datasets, describing the features of this problem, presenting those EFSs approaches that have been designed for addressing this task, and introducing a brief case study to excel the good behaviour of EFSs in this work area. Finally, in Sect. 12.5, we provide some concluding remarks of this work.

12.2 Fuzzy Rule Based Systems

The basic concepts which underlie fuzzy systems are those of linguistic variable and fuzzy IF-THEN rule. A linguistic variable, as its name suggests, is a variable whose values are words rather than numbers, e.g., small, young, very hot and quite slow. Fuzzy IF-THEN rules are of the general form: IF antecedent(s) THEN consequent(s), where antecedent and consequent are fuzzy propositions that contain linguistic variables. A fuzzy IF-THEN rule is exemplified by “IF the temperature is high THEN the fan-speed should be high”. With the objective of modeling complex and dynamic systems, FRBSs handle fuzzy rules by mimicking human reasoning (much of which is approximate rather than exact), reaching a high level of robustness with respect to variations in the system’s parameters, disturbances, etc. The set of fuzzy rules of an FRBS can be derived from subject matter experts or extracted from data through a rule induction process.

In this section, we present a brief overview of the foundations of FRBSs, with the aim of illustrate the way they behave. In particular, in Sect. 12.2.1, we introduce the important concepts of fuzzy set and linguistic variable. In Sect. 12.2.2, we deal with the basic elements of FRBSs. Finally, in Sect. 12.2.3 we describe the fuzzy inference system proposed by Mamdani for the output of an FRBS, as in this work we focus on linguistic systems.

12.2.1 Preliminaries: Fuzzy Set and Linguistic Variable

A *fuzzy set* is distinct from a crisp set in that it allows its elements to have a degree of membership. The core of a fuzzy set is its membership function: a surface or line that defines the relationship between a value in the set’s domain and its degree of membership. In particular, according to the original ideal of Zadeh [20], membership of an element x to a fuzzy set A, denoted as $\mu_A(x)$ or simply $A(x)$, can vary from 0 (full non-membership) to 1 (full membership), i.e., it can assume all values in the interval [0, 1]. Clearly, a fuzzy set is a generalization of the concept of a set whose membership function can take only two values $\{0, 1\}$.

We must point out that this is clearly a generalization and extension of multi-valued logic, in which degrees of truth are introduced in terms of the aforementioned membership functions. These functions can be seen as mapping predicates into FSs (or more formally, into an ordered set of fuzzy pairs, called a fuzzy relation). Fuzzy logic can thus be defined as a logic of approximate reasoning that allows us to work

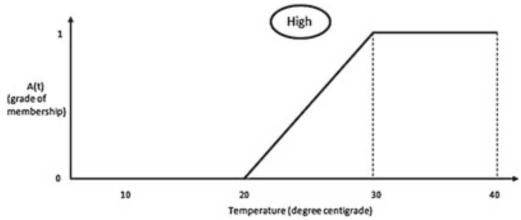

Fig. 12.1 Membership function

with FSs [21, 22]. In this manner, it allows a simplicity and flexibility which makes them superior with respect to classical logic for some complex problems. This can be achieved as they are able to cope with vague, imprecise or uncertain concepts that human beings use in their usual reasoning [23].

The value of $A(x)$ describes a degree of membership of x in A. For example, consider the concept of *high temperature* in an environmental context with temperatures distributed in the interval [0, 40] defined in degree centigrade. Clearly 0 ºC is not understood as a high temperature value, and we may assign a null value to express its degree of compatibility with the high temperature concept. In other words, the membership degree of 0 ºC in the class of high temperatures is zero. Likewise, 30 ºC and over are certainly high temperatures, and we may assign a value of 1 to express a full degree of compatibility with the concept. Therefore, temperature values in the range [30, 40] have a membership value of 1 in the class of high temperatures. From 20 to 30 ºC, the degree of membership in the fuzzy set high temperature gradually increases, as exemplified in Fig. 12.1, which actually is a membership function $A : T \rightarrow [0, 1]$ characterizing the fuzzy set of high temperatures in the universe $T = [0, 40]$. In this case, as temperature values increase they become more and more compatible with the idea of high temperature.

Linguistic variables are variables whose values are not numbers but words or sentences in a natural or artificial language. This concept has clearly been developed as a counterpart to the concept of a numerical variable. In concrete, a linguistic variable L is defined as a quintuple [24–26]: $L = (x, A, X, g, m)$, where x is the base variable, $A = \{A_1, A_2, \ldots, A_N\}$ is the set of *linguistic terms* of L (called *term-set*), X is the domain (universe of discourse) of the base variable, g is a syntactic rule for generating linguistic terms and m is a semantic rule that assigns to each linguistic term its *meaning* (a fuzzy set in X). Figure 12.2 shows an example of a linguistic variable *Temperature* with three linguistic terms "Low, Medium and High". The base variable is the temperature given in appropriate physical units.

Each underlying fuzzy set defines a portion of the variable's domain. But this portion is not uniquely defined. Fuzzy sets overlap as a natural consequence of their elastic boundaries. Such an overlap not only implements a realistic and functional semantic mechanism for defining the nature of a variable when it assumes various data values but also provides a smooth and coherent transition from one state to another.

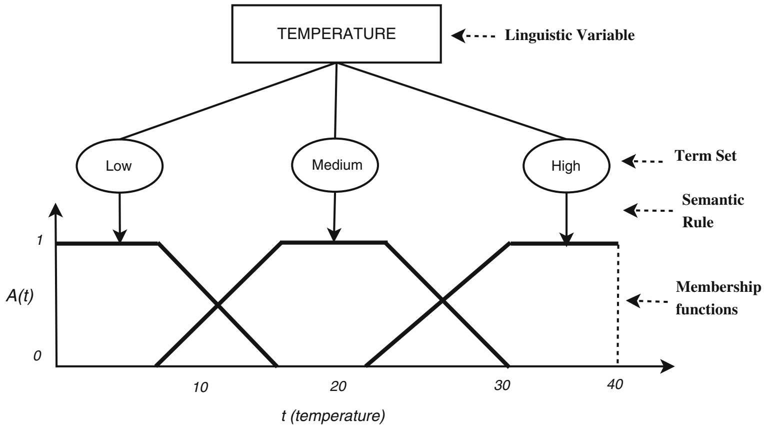

Fig. 12.2 Example of linguistic variable *Temperature* with three linguistic terms

12.2.2 Basic Elements of FRBSs

The essential part of FRBSs is a set of IF-THEN linguistic rules, whose antecedents and consequents are composed of fuzzy statements, related by the dual concepts of fuzzy implication and the compositional rule of inference.

An FRBS is composed of a *knowledge base* (KB), that includes the information in the form of IF-THEN fuzzy rules;

IF *a set of conditions are satisfied*

THEN *a set of consequents can be inferred*

and an inference engine module that includes:

- A *fuzzification interface*, which has the effect of transforming crisp data into fuzzy sets.
- An *inference system*, that uses them together with the KB to make inference by means of a reasoning method.
- A *defuzzification interface*, that translates the fuzzy rule action thus obtained to a real action using a defuzzification method.

Linguistic models are based on collections of IF-THEN rules, whose antecedents are linguistic values, and the system behavior can be described in natural terms. The consequent is an output action or class to be applied. For example, we can denote them as:

$$R_j:\ \text{IF } x_{p1} \text{ IS } A_{j1} \text{ AND } \cdots \text{ AND } x_{pn} \text{ IS } A_{jn} \text{ THEN } y \text{ IS } B_j$$

with $j = 1$ to L, and with x_{p1} to x_{pn} and y being the input and output variables, with A_{j1} to A_{jn} and B_j being the involved antecedents and consequent labels, respectively. They are usually called *linguistic* FRBSs or *Mamdani* FRBSs [27].

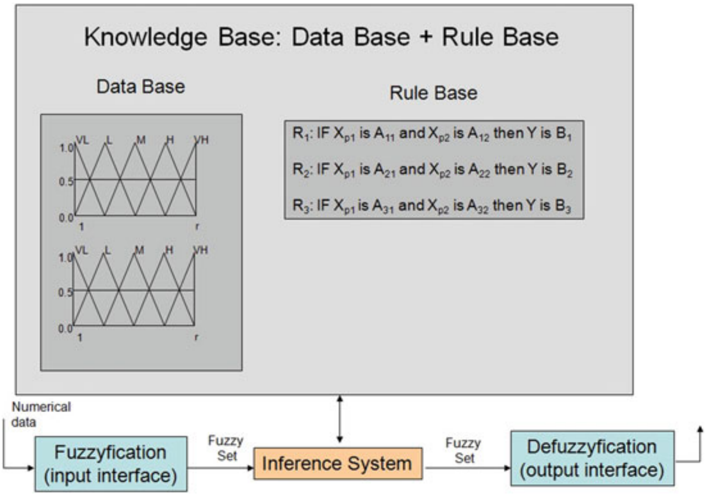

Fig. 12.3 Structure of an FRBS

In linguistic FRBSs, the KB is comprised by two components, a *data base* (DB) and a *rule base* (RB).

- A DB, containing the linguistic term sets considered in the linguistic rules and the membership functions defining the semantics of the linguistic labels.
  Each linguistic variable involved in the problem will have associated a fuzzy partition of its domain representing the fuzzy set associated with each of its linguistic terms. Reader is referred to recall Fig. 12.2 where we showed an example of fuzzy partition with three labels. This can be considered as a discretization approach for continuous domains where we establish a membership degree to the items (labels), we have an overlapping between them, and the inference engine manages the matching between the patterns and the rules providing an output according to the rule consequents with a positive matching. The determination of the fuzzy partitions is crucial in fuzzy modeling [28], and the granularity of the fuzzy partition plays an important role for the FRBS behavior [29].
- A RB, comprised of a collection of linguistic rules that are joined by a rule connective ("also" operator). In other words, multiple rules can be triggered simultaneously for the same input.

The generic structure of an FRBS is shown in Fig. 12.3.

12.2.3 Mamdani Fuzzy Inference Process

The inference engine of FRBSs acts in a different way depending of the kind of problem (classification or regression) and the kind of fuzzy rules. It always includes a fuzzification interface that serves as the input to the fuzzy reasoning process, an inference system that infers from the input to several resulting outputs (fuzzy set, class, etc.) and the defuzzification interface or output interface that converts the fuzzy sets obtained from the inference process into a crisp action that constitutes the global output of the FRBS, in the case of regression problems, or provide the final class associated to the input pattern according to the inference model.

According to Mamdani principles [30], the fuzzy inference process includes five parts, which contain a very simple structure of "max-min" operators, specifically fuzzification of the input variables, application of the fuzzy operator (AND or OR) in the antecedent, implication from the antecedent to the consequent, aggregation of the consequents across the rules and defuzzification. These five operations can be compressed into three basic steps, which are described below:

Step 1. **Computation of the Matching Degree**. The first step is to take the inputs and determine the degree to which they belong to each of the fuzzy sets considering the membership functions. In order to compute the matching degree to which each part of the antecedent is satisfied for each rule, a conjunction operator C is applied. Specifically, Mamdani recommended the use of the minimum t-norm.

$$\mu_{A_j}(x_p) = C(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})), \qquad j = 1, \ldots, L. \tag{12.1}$$

Step 2. **Apply an Implication Operator**. In this step, the consequent is reshaped using a function associated with the antecedent (a single number). The input for the implication process is a single number given by the antecedent, and the output is a fuzzy set. Implication is implemented for each rule. Usually, two approaches for the implication operator I are employed, i.e. minimum t-norm, which truncates the output fuzzy set, and product t-norm, which scales the output fuzzy set. Mamdani also recommended the use of the minimum t-norm in this case.

$$\mu_{B'_j}(y) = I(\mu_{A_j}(x_p), \mu_{B_j}(y)) \qquad j = 1, \ldots, L. \tag{12.2}$$

Step 3. **Defuzzification process**. Decisions are based on the testing of all of the rules in a fuzzy inference system, so rules must be combined in order to make a decision. There are two modes of obtaining the output value of a fuzzy system, namely "*aggregation first, defuzzification after*" and "*defuzzification first, aggregation after*". The defuzzification method suggested by Mamdani considers the first method using the *centre of gravity* of the individual fuzzy sets aggregated with the maximum connective *also*.

$$\mu_{B(y)} = \bigcup_j \mu_{B'_j}(y) \tag{12.3}$$

$$y_0 = \frac{\int_y y \cdot \mu_B(y)dy}{\int_y \mu_B(y)} \tag{12.4}$$

12.3 Evolutionary Fuzzy Systems: Taxonomy and Analysis

EFSs are a family of approaches that are built on top of FRBSs, whose components are improved by means of an evolutionary learning/optimization process as depicted in Fig. 12.4. This process is designed for acting or tuning the elements of a fuzzy system in order to improve its behavior in a particular context. Traditionally, this was carried out by means of Genetic Algorithms, leading to the classical term of Genetic Fuzzy Systems [5, 31–33]. In this chapter, we consider a generalization of the former by the use of EAs [7].

The central aspect on the use of EAs for automatic learning of FRBSs is that the design process can be analyzed as a search problem in the space of models, such as the space of rule sets, membership functions, and so on. This is carried out by means of the coding of the model in a chromosome. Therefore, the first step in designing an EFS is to decide which parts of the fuzzy system are subjected to optimization by the EA coding scheme. Hence, EFS approaches can be mainly divided into two types of processes: tuning and learning. Additionally, we must make a decision whether to just improve the accuracy/precision of the FRBS or to achieve a tradeoff between accuracy and interpretability (and/or other possible objectives) by means of a MOEA. Finally, we must stress that new fuzzy set representations have been designed, which implies a new aspect to be evolved in order to take the highest advantage of this approach.

This high potential of EFSs implies the development of many different types of approaches. In accordance with the above, and considering the FRBSs' components involved in the evolutionary learning process, a taxonomy for EFS was proposed by Herrera in [33] (please refer to its thematic Website at http://sci2s.ugr.es/gfs/). More recently, in [6] we extended the former by distinguishing among the learning of the

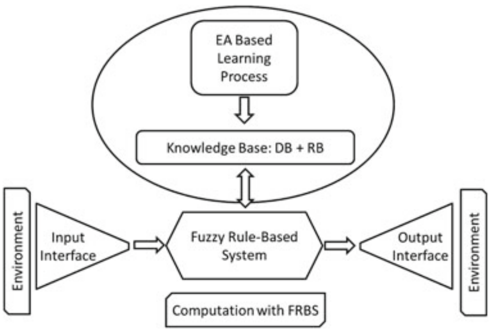

Fig. 12.4 Integration of an EFS on top of an FRBS

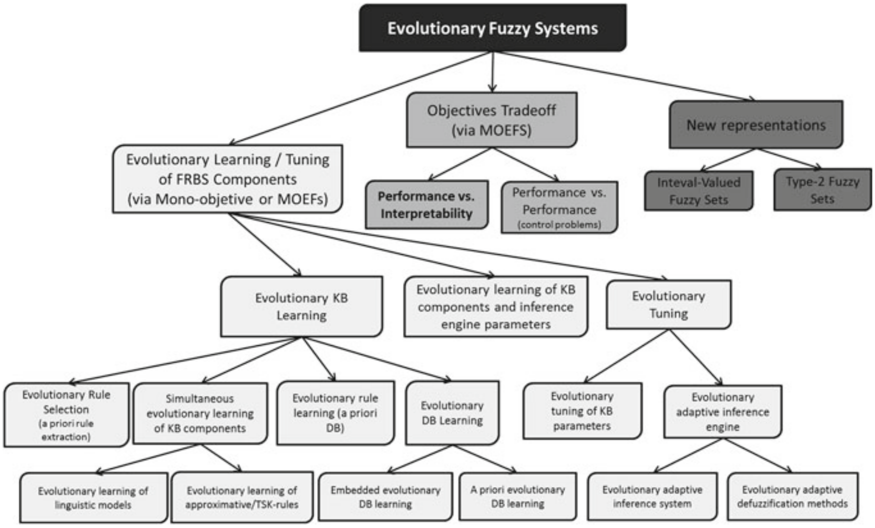

Fig. 12.5 Evolutionary fuzzy systems taxonomy

FRBSs' elements, the EA components and tuning, and the management of the new fuzzy sets representation. This novel EFS taxonomy is depicted in Fig. 12.5.

In order to describe this taxonomy tree of EFSs, this section is arranged as follows. First, we present those models according to the FRBS components involved in the evolutionary learning process (Sect. 12.3.1). Afterwards, we focus on the multi-objective optimization (Sect. 12.3.2). Finally, we provide some brief remarks regarding the parametrized construction for new fuzzy representations (Sect. 12.3.3).

12.3.1 Evolutionary Learning and Tuning of FRBSs' Components

When addressing a given Data Mining problem, the use of any fuzzy sets approach is usually considered when an interpretable system is sought, when the uncertainty involved in the data must be properly managed, or even when a dynamic model is under consideration. Then, we must make the decision on whether a simple FRBS is enough for the given requirements, or if a more sophisticated solution is needed, thus exchanging computational time for accuracy.

As introduced previously, this can be achieved either by designing approaches to learn the KB components, including an adaptive inference engine, or by starting from a given FRBS, developing approaches to tune the aforementioned components. Therefore, we may distinguish among the evolutionary KB learning, the evolutionary learning of KB components and inference engine parameters, and the evolutionary tuning. These approaches are described below, which can be performed via a standard mono-objective approach or a MOEA.

12.3.1.1 Evolutionary KB Learning

The following four KB learning possibilities can be considered:

1. *Evolutionary rule selection.* In order to get rid of irrelevant, redundant, erroneous and/or conflictive rules in the RB, which perturb the FRBS performance, an optimized subset of fuzzy rules can be obtained [34].
2. *Simultaneous evolutionary learning of KB components.* Working in this way, there is possibility of generating better definitions of these components [35]. However, a larger search space is associated with this case, which makes the learning process more difficult and slow.
3. *Evolutionary rule learning.* Most of the approaches proposed to automatically learn the KB from numerical information have focused on the RB learning, using a predefined DB [36].
4. *Evolutionary DB learning.* A DB generation process allows the shape or the membership functions to be learnt, as well as other DB components such as the scaling functions, the granularity of the fuzzy partitions, and so on. Two possibilities can be used: "a priori evolutionary DB learning" and "embedded evolutionary DB learning [37]."

12.3.1.2 Evolutionary Learning of KB Components and Inference Engine Parameters

This area belongs to a hybrid model between adaptive inference engine and KB components learning. These type of approaches try to find high cooperation between the inference engine via parameters adaptation and the learning of KB components, including both in a simultaneous learning process [38].

12.3.1.3 Evolutionary Tuning

With the aim of making the FRBS perform better, some approaches try to improve the preliminary DB definition or the inference engine parameters once the RB has been derived. The following three tuning possibilities can be considered (see the sub-tree under "evolutionary tuning").

1. *Evolutionary tuning of KB parameters.* A tuning process considering the whole KB obtained is used a posteriori to adjust the membership function parameters, i.e. the shapes of the linguistic terms [39].
2. *Evolutionary adaptive inference systems.* This approach uses parameterized expressions in the inference system, sometimes called adaptive inference sys-

tems, for getting higher cooperation among the fuzzy rules without losing the linguistic rule interpretability [40].

3. *Evolutionary adaptive defuzzification methods.* When the defuzzification function is applied by means of a weighted average operator, i.e. parameter based average functions, the use of EAs can allow us to adapt these defuzzification methods [41].

12.3.2 Approaches for Optimizing Several Objectives

Traditionally, the efforts in developing EFSs were aimed at improving the accuracy/precision of the FRBS in a mono-objective way. However, in current applications the interest of researchers in obtaining more interpretable linguistic models has significantly grown [42]. The hitch is that accuracy and interpretability represent contradictory objectives. A compromise solution is to address this problem using MOEAs [8] leading to a set of fuzzy models with different tradeoffs between both objectives instead of a biased one. These hybrid approaches are known as MOEFSs [10] that, in addition to the two aforementioned goals, may include any other kind of objective, such as the complexity of the system, the cost, the computational time, additional performance metrics, and so on.

In this case, the division of this type of techniques is first based on the multi-objective nature of the problem faced and second on the type of FRBS components optimized. Regarding the previous fact, those of the second level present a clear correspondence with the types previously described for EFSs in the previous section.

Here, we will only present a brief description for each category under consideration. For more detailed descriptions or an exhaustive list of contributions see [10] or its associated Webpage (http://sci2s.ugr.es/moefs-review/).

12.3.2.1 Accuracy-Interpretability Trade-Offs

The comprehensibility of fuzzy models began to be integrated into the optimization process in the mid 1990s [43], thanks to the application of MOEAs to fuzzy systems. Nowadays, researchers agree on the need to consider two groups of interpretability measures, complexity-based and semantic-based ones. While the first group is related to the dimensionality of the system (simpler is better) the second one is related to the comprehensibility of the system (improving the semantics of the FRBS components). For a complete survey on interpretability measures for linguistic FRBSs see [42].

The differences between both accuracy and interpretability influence the optimization process, so that researchers usually include particular developments in the proposed MOEA making it able to handle this particular trade-off. An example can be seen in [44] where authors specifically force the search to focus on the most accurate solutions.

12.3.2.2 Performance Versus Performance (Control Problems)

In control system design, there are often multiple objectives to be considered, i.e. time constraints, robustness and stability requirements, comprehensibility, and the compactness of the obtained controller. This fact has led to the application of MOEAs in the design of Fuzzy Logic Controllers.

The design of these systems is defined as the obtaining of a structure for the controller and the corresponding numerical parameters. In a general sense, they fit with the tuning and learning presented for EFSs in the previous section. In most cases, the proposal deals with the postprocessing of Fuzzy Logic Controller parameters, since it is the simplest approach and requires a reduced search space.

12.3.3 Novel Fuzzy Representations

Classical approaches on FRBSs make use of standard fuzzy sets [20], but in the specialized literature we found extensions to this approach with aim to better represent the uncertainty inherent to fuzzy logic. Among them, we stress Type-2 fuzzy sets [45] and Interval-Valued Fuzzy Sets (IVFSs) [46] as two of the main exponents of new fuzzy representations.

Type-2 fuzzy sets reduce the amount of uncertainty in a system because this logic offers better capabilities to handle linguistic uncertainties by modeling vagueness and unreliability of information. In order to obtain a type-2 membership function, we start from the type-1 standard definition, and then we blur it to the left and to the right. In this case, for a specific value, the membership function, takes on different values, which are not all weighted the same. Therefore, we can assign membership grades to all of those points.

For IVFS [46], the membership degree of each element to the set is given by a closed sub-interval of the interval [0,1]. In such a way, this amplitude will represent the lack of knowledge of the expert for giving an exact numerical value for the membership. We must point out that IVFSs are a particular case of type-2 fuzzy sets, having a zero membership out of the ranges of the interval.

In neither case, there is a general design strategy for finding the optimal fuzzy models. In accordance with the former, EAs have been used to find the appropriate parameter values and structure of these fuzzy systems.

In the case of type-2 fuzzy models, EFSs can be classified into two categories [47]: (1) the first category assumes that an "optimal" type-1 fuzzy model has already been designed, and afterwards a type-2 fuzzy model is constructed through some sound augmentation of the existing model [48]; (2) the second class of design methods is concerned with the construction of the type-2 fuzzy model directly from experimental data [49].

Regarding IVFS, current works initialize type-1 fuzzy sets as those defined homogeneously over the input space. Then, the upper and lower bounds of the interval for each fuzzy set are learnt by means of a weak-ignorance function (amplitude tuning) [50], which may also involve a lateral adjustment for the better contextualization of the fuzzy variables [51]. Finally, in [52] IVFS are built ad-hoc, using an interval-valued restricted equivalence functions within a new interval-valued fuzzy reasoning method. The parameters of these equivalence functions per variable are learnt by means of an EA, which is also combined with rule selection in order to decrease the complexity of the system.

12.4 The Use of Evolutionary Fuzzy Systems for Classification with Imbalanced Datasets

As EFSs have improved their performance from its initial models, they have been applied to novel challenges like the problem of classification with imbalanced datasets [12, 13]. This classification scenario has gained recognition in the last years as its importance comes from its presence in numerous real-world problems and the necessity of using specific approaches to address them.

In Sect. 12.4.1, we will briefly introduce the problem of classification with imbalanced datasets, outlining the approaches that are usually applied in the area and the evaluation metrics that are specifically used in this case. Then, in Sect. 12.4.2, we will provide an analysis over the EFS approaches that have been proposed to handle datasets with imbalanced distributions. Finally, one of the EFS approaches that we have developed in the topic, GP-COACH-H [19], is further described in Sect. 12.4.3 together with an experimental analysis to prove its usefulness in the imbalanced scenario.

12.4.1 Introduction to Classification with Imbalanced Datasets

The classification problem with imbalanced datasets arises when the number of examples belonging to one class is negligible with respect to the number of examples that represent the other classes [53–55]. In this problem, it is precisely the underrepresented class, also known as the minority or positive class, the one which needs to be properly identified, as its incorrect identification usually entails high costs [56, 57]. This fact contrasts with the more represented classes, also known as majority or negative classes, which are typically well identified.

The importance of this problem comes from its presence in a high number of real-world problems, becoming one of the top challenges in data mining research [58]. We can find imbalanced distributions in areas like risk management [59], bioactivity of chemical substances [60], fraud detection [61], system failure detection [62] or medical applications [63, 64], just mentioning some of them.

Standard classification algorithms are not usually able to provide a good identification of samples belonging to the minority class as they are guided by global search measures, like a percentage of well-classified examples. Thus, following the standards models are created trying to cover as many samples as possible while maintaining their simplicity. In these cases, the developed models properly identify many examples of the majority class, as they represent the higher number of examples of the whole dataset. However, minority class examples are not usually covered because their representation is small and therefore, it could have no influence in the learning stage and hence no classification rule is created for them.

In this context, the imbalance ratio (IR) [65], is traditionally used to determine how difficult a classification problem with imbalanced datasets is. Specifically, it is defined as the quotient between the number of examples belonging to the majority class and the number of samples belonging to the minority class, $IR = \frac{\#numMaj}{\#numMin}$. Although the imbalanced distribution poses a major challenge to classifiers, there are also some data intrinsic characteristics that difficult the classification with imbalanced datasets, further degrading the performance of methods than when these issues arise separately [12]. These data intrinsic characteristics include the small sample size or lack of density problem [66], the overlapping of the samples belonging to each class [16], the presence of small disjuncts in the data [67], the existence of borderline [68] or noisy samples [69] and the dataset shift [70].

Numerous approaches have been proposed to address the problem of classification with imbalanced datasets. They are commonly divided into approaches at the data-level, at the algorithm-level, and cost-sensitive learning, all of which can be embedded into an ensemble learning scheme:

- Approaches at the data-level [71, 72] modify the original training set to obtain a more or less balanced dataset that can be addressed using standard classification algorithms. These modifications to the dataset can be performed generating additional examples associated to the minority class (oversampling) or removing examples from the minority class (undersampling).
- Algorithm-level approaches [73] modify existing standard classification methods in order to enhance the identification of the minority class examples. These modifications may include the use of imbalanced measures to guide the search, a limitation of procedures designed to generalize the models or even new operations specifically designed to focus on the minority class.
- Cost-sensitive learning solutions combine approaches at the algorithm-level and the data-level for imbalanced classification considering the variable costs of misclassifying an instance as belonging to the other class [74, 75]. In imbalanced datasets, the misclassification costs associated to a minority class instance are higher than the costs associated to the misclassification of a majority class instance $C(min, maj) > C(maj, min)$, as the minority class is the main interest in the learning process.
- Ensembles have also been adapted to imbalance learning [76] obtaining a high performance when applied. In general, these new ensemble approaches introduce

in their way of running some cost-sensitive learning [77] or data preprocessing features [78, 79].

Finally, when considering the evaluation of the performance of classifiers in this context, we must proceed carefully. The most common measure of performance, the overall accuracy in classification, is not appropriate in a dataset with an uneven class distribution, as a high value in the measure can be obtained correctly classifying the instances associated to the majority class, even when all the minority class samples are not properly identified. This situation is completely undesirable as the minority class is the most interesting from the learning point of view. For this reason, more appropriate performance metrics are used in the imbalanced classification scenario.

The geometric mean (GM) [80] of the true rates is one measure that is able to avoid the problems related to the traditional accuracy metrics and is defined as:

$$GM = \sqrt{sensitivity \cdot specificity} \tag{12.5}$$

where $sensitivity = \frac{TP}{TP+FN}$ and $specificity = \frac{TN}{FP+TN}$. The sensitivity and specificity values represent the true rate for each class, computed from TP and TN which are the true rate for the minority and majority instances and the FP and FN which are the rate for the false minority samples and majority examples respectively. This metric tries to balance the performance over the two classes, combining both objectives into one.

12.4.2 EFS Approaches for Imbalanced Classification

EFSs have evolved and addressed new challenges and problems since they were first proposed. There are several proposals of EFS for imbalanced datasets; some of them study the impact and improvement of these systems modifying some of the fuzzy components while others introduce new operations in the methods without changing the basic fuzzy inference process. As the EFS methods applied to imbalanced classification are quite varied, we have organized them in four groups considering how they approach the imbalanced classification problem, namely, with data-level approaches, algorithm-level approaches, cost-sensitive learning and ensembles.

12.4.2.1 EFS and Data-Level Approaches

Data preprocessing techniques have been used together with EFSs because of their versatility as they are independent of the classifier used. FRBSs have demonstrated a good performance [14] in the imbalanced classification scenario, especially for oversampling techniques. This has encouraged the development of different fuzzy based classifiers together with preprocessing approaches. One of the most popular oversampling techniques, the "Synthetic Minority Oversampling TEchnique" (SMOTE) algorithm [72], has been extensively used in combination with EFSs.

One of the EFSs approximations developed for imbalanced datasets is the one described in [81]. In it, an adaptive inference system with parametric conjunction operators is presented. To deal with the imbalance, the SMOTE algorithm is used to balance the datasets. The idea presented in the paper is based on the suitability of adaptive t-norms, like the Dubois t-norm where a parameter α modifies how the t-norm behaves. Using the CHC evolutionary algorithm [82] with the GM performance measure as evaluation function, the α parameter can be learned for the whole RB or for each specific rule improving the overall performance of the system.

An analysis about the evolutionary tuning of the KB for classification with imbalanced datasets is performed in [83]. In order to avoid the imbalanced problem, the SMOTE algorithm is again used to obtain a balanced distribution of the train set. In this work, a genetic process based on the CHC evolutionary algorithm is used to learn the lateral displacement of the DB using the 2-tuples linguistic representation [84]. This lateral translation can be learned over the full DB for the complete RB or adapting each set of fuzzy labels according to each specific rule in the RB. In this way, the performance of the Chi et al. method [85] and the FH-GBML classifier [86] is improved.

The imbalanced problem can emerge in conjunction with other problems like the availability of low quality data. Therefore, the uncertainty that needs to be managed does not refer only to the difficult identification of samples for each class but also to the values associated to the input values of the samples. In [18], several preprocessing techniques are adapted to the low quality data scenario to obtain a more or less balanced distribution that can be managed more easily. Specifically, low quality data versions of the ENN [87], NCL [88], CNN [89], SMOTE [72] and SMOTE+ENN [71] algorithms are designed to classify low quality imbalanced data using a genetic cooperative-competitive learning algorithm. The performance of these versions is similar to the one obtained with the preprocessing methods for the standard imbalanced problems.

In [90], a genetic procedure for learning the KB in imbalanced datasets, GA-FS+GL, is proposed. In this case, the SMOTE algorithm is again used to balance the training set. The idea presented in this work is the use of a GA to perform a feature selection and a selection of the granularity of the data base. To perform the feature selection, a binary part of the chromosome is used to determine if an attribute is used or not. To select the granularity of the DB, the algorithm searches for the best performing set of labels considering different number of labels between two and seven using equally distributed triangular membership functions. The approach is tested over the Chi et al. method [85] obtaining competitive results.

The data intrinsic characteristics can degrade the performance of imbalanced classifiers in a further extent than when they appear in isolation [12]. In [91], the impact of dataset shift over imbalanced classifiers is studied. Specifically, two partitioning techniques for the validation of classifiers are compared: the stratified cross-validation and a novel cross-validation approach named DOB-SCV [92], an approach that tries to limit the covariate shift that is induced when partitions are created. The FARC-HD classifier [93] is selected to compare how the dataset shift changes the behavior of EFS classifiers. The results obtained show that it is advisable to limit the dataset shift introduced in these processes even when we are using fuzzy systems that can cope with uncertainty and imprecision.

12.4.2.2 EFS and Algorithm-Level Approaches

Modifying operations within the algorithm design to further enhance the correct identification of examples belonging to the minority class is a popular solution to adapt EFS for imbalanced classification. In some cases, these alterations are enough to address the imbalanced distributions, however, in others it is necessary to combine them with preprocessing techniques to further improve the performance of these methods.

One of the approaches that follows a design with specific operations for the imbalance is the one described in [94], renamed in [95] as FLAGID, Fuzzy Logic And Genetic algorithms for Imbalanced Datasets. This approach follow several stages, starting with a first step that is a modified version of the RecBF/DDA algorithm [96]. This first step, creates the membership functions that are going to be used afterwards, creating a smaller number of membership functions for the minority class. The second stage is called ReRecBF and its aim is to simplify the previously created functions so that they cover important regions being able to at least represent a 10 % of the class. Finally, the third stage learns the RB considering the previously generated membership functions using a genetic algorithm procedure that uses the GM as fitness function.

The use of a hierarchical fuzzy rule-based classification system (HFRBCS) for imbalanced classifiers has been considered in [17], using Chi et al.'s method as baseline classifier. Additionally, in [19] authors propose GP-COACH-H, which is also based on a hierarchical system. In both HFRBCSs, the KB is structured following different levels of learning, being the lower levels more general and the higher levels more specific. This type of approaches aim to improve the performance of methods in difficult data areas like the data intrinsic characteristics that further difficult the classification with imbalanced datasets. In a first stage, the SMOTE algorithm is used to balance the data that will be later processed by the hierarchical methods that try to identify the samples in difficult areas. For HFRBCS(Chi) [17], the generation process of the hierarchical rules is based on the Chi et al. method [85]. When the hierarchical rule base has been obtained, a rule selection process to select a subset of rules for the final classifier is used. For GP-COACH-H [19], the hierarchical rule base is obtained modifying the GP-COACH algorithm [97]. In this method, a subsequent step is applied with a genetic tuning of the knowledge base where a combined selection of the rules and a lateral tuning based on the 2-tuples representation is developed. GP-COACH-H will be later described as our selected method for the case of study.

Another algorithm that has been modified for imbalanced classification is the FARC-HD algorithm [93]. Specifically, this method is applied in the case study of intrusion detection systems [98, 99], following two different schemes. In the first case [100], the alterations to the method follow two aspects: the first one is the use of the SMOTE algorithm to preprocess the data for the subsequent operations; whereas the second one is related to the changes introduced in the genetic tuning of the knowledge base phase that is performed in the FARC-HD method. This genetic procedure changes its evaluation function to the GM performance measure. On the other hand, in [101] the FARC-HD EFS baseline algorithm is embedded in a pairwise learning scheme [102]. This is done for being able to improve the recognition of the minority class instances by simplifying the borderline areas in each binary classifier. Finally, Sanz et al. applied an interval-valued fuzzy sets version of FARC-HD in the context of the modeling and prediction of imbalanced financial datasets [103].

Following diverse evolutionary approaches, we can find another proposal for imbalanced classification in [104]. The method is divided in two steps, a first step that applies a feature selection process to reduce the dimensionality of the training set and a second step to generate the fuzzy rules using evolutionary techniques. This second step is divided in several steps as well: first, a differential evolution optimization process is performed to estimate a *radii* value; then, this *radii* value will be used in a subtractive clustering method to generate fuzzy TSK rules; finally, a genetic programming step is used to improve the fuzzy TSK rules and convert them to Mamdani classification rules.

12.4.2.3 EFS and Cost-Sensitive Learning

As in the previous cases, EFSs have also been adapted to classification with imbalanced datasets following cost-sensitive learning, that is, considering the costs within the algorithm to favor the classification of the minority class examples. In [105], the FH-GBML-CS method is proposed, which is a cost-sensitive version of the FH-GBML algorithm [86]. The costs are introduced in the evaluation function of the genetic procedure and in the computation of the rule weight, using the what is called as the cost-sensitive penalized certainty factor, a modified version including costs of the penalized certainty factor [106]. Moreover, the inference process for the fuzzy reasoning method considers the compatibility degree of the example and the fuzzy label together with the cost associated to that example.

In [107], a cost-sensitive MOEFS for imbalanced classification is presented. In this method, the NSGA-II algorithm is used to create an FRBS using the specificity performance measure as an objective, the sensitivity performance measure as another objective and the complexity as a third objective. This complexity measure is computed adding the number of antecedents in all the rules in the RB. Finally, a ROC convex hull technique is used together with the misclassification costs to select the best solution of the pareto of solutions obtained by the multi-objective method.

12.4.2.4 EFS and Ensemble Learning

The existing learning approaches for imbalanced classification have been typically used with weak learners like decision trees instead of using other approaches. In this way, ensembles are rarely combined with EFS in the imbalanced scenario. However, even when this type of systems are not specifically designed for imbalance, we can find proposals that are tested over imbalanced datasets.

For example, in [108], a boosting method based on fuzzy rules is proposed. This method is divided in two stages. In the first one, a fuzzy rule learning method is applied. This method is based on the AdaBoost algorithm [109], which has been modified following an iterative rule learning approach. Rules are created with an evolutionary rule generation method that adapts the rules to the dataset according to the learned weights in each iteration until the performance does not improve or even decreases. When this first stage has finished, a genetic tuning step is performed. There are no specific operations for the imbalanced distributions in this method as it was not specifically designed for imbalanced data, however, the weights associated to each sample in the boosting can favor the correct identification of minority class examples. Furthermore, the method is tested in the land cover classification of remote sensing imagery problem, which can feature imbalanced distributions.

Finally, in [110], three EFS systems for imbalanced classification are compared, which are the GA-FS+GL method described in [90], the GP-COACH-H algorithm presented in [19] and the MOEFS developed in [107]. The authors use 22 datasets to perform this comparison being the MOEFS approach the one with the best performance supported by a Holm test. However, the results extracted from this comparison need to be treated with care, as the AUC performance measure [111] is not computed in the same way for the three methods: for the MOEFS approach, the AUC measure is computed considering all the classifiers obtained in the multi-objective process without applying the ROC convex hull technique that selects one of them, while for the other two methods, the AUC measure is calculated considering the one point formula.

12.4.3 Case Study: Addressing Highly Imbalanced Datasets with GP-COACH-H

Having analyzed how EFS adopt the different strategies for imbalanced datasets, we have selected one of the EFS approaches to demonstrate its effectiveness for classification with imbalanced datasets. In a first step, we will further describe the GP-COACH-H algorithm, a fuzzy rule-based classification system for imbalanced data. Then, we will present the experimental framework associated to the study performed, the result tables and its associated statistical tests.

12.4.3.1 GP-COACH-H: A Hierarchical Genetic Programming Fuzzy Rule-Based Classification System with Rule Selection and Tuning

The GP-COACH-H algorithm [19] is a fuzzy rule-based classification system that has been developed to effectively address imbalanced datasets in arduous imbalanced scenarios, such as highly imbalanced datasets and borderline imbalanced datasets. To do so, the GP-COACH-H algorithm follows an algorithm-level approach as its behavior is modified in order to favor the correct identification of samples belonging to the minority class.

The proposal combines several strategies that are able to obtain a good synergy between them, namely, data preprocessing, a hierarchical linguistic classifier and genetic tuning of KB parameters. As previously mentioned, the data preprocessing modifies the input dataset to ease the learning process of the subsequent classifier.

An HFRBCS [112] extends the standard definition of the KB so it can better model complex search spaces such as imbalanced datasets entangled with some data intrinsic characteristics like borderline examples, overlapping between the classes or even small disjuncts. In these systems, the KB is called hierarchical knowledge base (HKB), as the changes introduced by the hierarchical model affect both the DB and the RB.

An HKB is composed by a set of layers, which represent different granularity levels. Each layer has its own DB and RB which determine the specificity level of description that can be achieved by the model. The RB of a layer can only use the fuzzy linguistic labels defined in the associated DB. The layers are organized in a hierarchical way: a new layer level has a higher number of fuzzy labels than the previous level and the fuzzy labels built in the new level are created preserving the membership function modal points, adding a new linguistic term between each two consecutive terms of the linguistic partition belonging to the previous model. The idea behind the usage of this system is to use a low hierarchical level to describe general areas of data, while using a larger hierarchical level to illustrate more complex areas.

Another strategy used in this method is the genetic tuning of KB parameters. As it was explained in the previous sections, the objective of this component is to enhance the previously learned KB using genetic algorithms so that the final model is able to better characterize the classes of the dataset.

The GP-COACH-H algorithm follows a three stages approach in order to integrate the different strategies proposed. A flowchart of the building of this model can be found in Fig. 12.6. The three stages of the algorithm involve the following operations:

1. *Data preprocessing*: As a first step, the GP-COACH-H algorithm needs to modify the original training set so it displays a more or less balanced distribution. In order to do so, an oversampling scheme is used, following the SMOTE algorithm [72] to generate synthetic samples associated to the minority class.
2. *HKB generation process with an EFS*: Considering the dataset obtained in the previous step, a genetic programming approach is used to generate the HKB. GP-

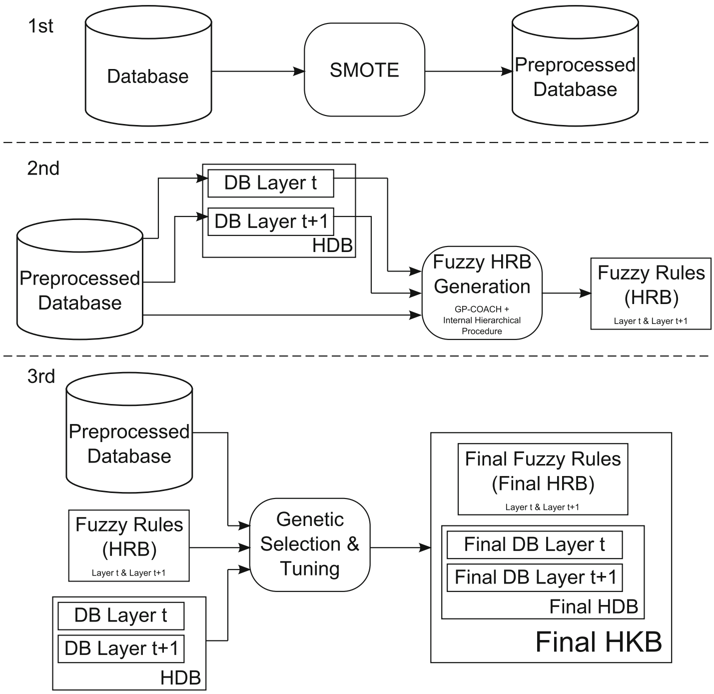

Fig. 12.6 Flowchart of the GP-COACH-H algorithm

COACH-H is based on the GP-COACH algorithm [97], and therefore, it follows the genetic programming procedure adopted in that method. To obtain a HKB, some of the GP-COACH steps need to be modified to enable the usage of rules of different hierarchies in the same population. The first alteration, is the addition of a new step in each generation of the genetic approach: the identification of *good* and *bad* rules in the current population. *Bad* rules are discarded and are replaced by new high granularity rules. Another modification to the genetic process is the use of a new evaluation function that is able to consider different granularity levels in the rules. New constraints over the crossover operator need to be performed to ensure that it is only applied to rules of the same hierarchy.

3. *Genetic tuning of the HKB parameters*: When the building of the HKB has ended, a genetic tuning process of the HKB parameters is started to further adapt the classifier to the available data. In this step, we try to perform a selection of rules that demonstrate a good cooperation [34] while also tuning the existing hierarchical DBs following a 2-tuples linguistic representation [84]. These optimizations are done together with an unique genetic procedure using the CHC evolutionary algorithm to profit from the synergy that these optimizations can achieve. To do so, each part of the chromosome codifies the rule selection process or the tuning adjustment where genetic operators are modified to consider this situation.

12.4.3.2 Experimental Study

To evaluate the performance of the GP-COACH-H algorithm, we have selected 44 highly imbalanced datasets with an IR higher than 9 from the KEEL dataset repository[1] [113]. Table 12.1 summarizes the datasets used in this study, showing for each one the number of examples (#Ex.), the number of attributes (#Atts.), the class name associated to each class (minority and majority), the attribute class distribution and the IR. This table is sorted by increasing order of IR.

To carry out this study, we follow a five-fold cross-validation approach, that is, performing five random partitions of the data where the aggregation of four of them constitutes the training set, and the other partition is considered as the test set. The results presented in this work display the average results obtained in the five partitions.

For the GP-COACH-H algorithm, we have set the parameters of its different components according to what is usual in these domains. For the SMOTE preprocessing part, we consider only the 1-nearest neighbor (using the euclidean distance) to generate the synthetic examples to obtain a balanced dataset. The parameters associated to the fuzzy rule-based classification system are the use of the minimum t-norm as conjunction operator, the maximum t-norm as disjunction operator, the certainty factor is used to compute the rule weight, the normalized sum is used as fuzzy reasoning method and 5 fuzzy labels are used for low granularity rules while 9 fuzzy labels are used for high granularity rules. Considering the genetic part of the building of the model, the number of evaluations used are 20000, the initial population has a size of 200, the α value for the raw fitness is 0.7, the crossover probability is 0.5, the mutation probability is 0.2, the dropping condition probability is 0.15, the insertion probability is 0.15 as well, the tournament size is 2, while the weights associated to the fitness function are $w_1 = 0.8$, $w_2 = w_3 = 0.05$, $w_4 = 0.1$. Finally, the parameters associated to the hierarchical procedure and the last genetic tuning phase are an α of 0.2 to detect *good* and *bad* rules, the number of evaluations is 10000, the population size is 61 and the bits per gene are 30.

To demonstrate the good performance of the GP-COACH-H algorithm for imbalanced datasets, we have selected the C4.5 algorithm [114], a well-known decision tree that has displayed a good behavior for imbalanced datasets [71]. The parameters associated to this method are the ones recommended by the author, namely, the use of a pruned tree, a confidence of 0.25 and 2 as the minimum number of item-sets per leaf. To deal with the imbalance, we have combined C4.5 with the SMOTE+ENN algorithm [71], where the ENN cleaning method [87] is directly applied after the SMOTE algorithm to generalize the borders between the classes. For generating synthetic samples, the 1-nearest neighbor is used to obtain a balanced dataset. For the ENN part, 3-nearest neighbors are considered. In both cases, the euclidean distance is applied.

Furthermore, we have used statistical tests to detect whether there are significant differences among the results achieved by the different tested methods [115]. We will use non-parametric tests as the conditions that guarantee the reliability of parametric

[1] http://www.keel.es/imbalanced.php.

Table 12.1 Summary of highly imbalanced datasets

Datasets	#Ex.	#Atts.	Class (maj;min)	%Class(maj; min)	IR
ecoli034vs5	200	7	(p,imL,imU; om)	(10.00, 90.00)	9.00
yeast2vs4	514	8	(cyt; me2)	(9.92, 90.08)	9.08
ecoli067vs35	222	7	(cp,omL,pp; imL,om)	(9.91, 90.09)	9.09
ecoli0234vs5	202	7	(cp,imS,imL,imU; om)	(9.90, 90.10)	9.10
glass015vs2	172	9	(build-win-non_float-proc,tableware,	(9.88, 90.12)	9.12
			build-win-float-proc; ve-win-float-proc)		
yeast0359vs78	506	8	(mit,me1,me3,erl; vac,pox)	(9.88, 90.12)	9.12
yeast02579vs368	1004	8	(mit,cyt,me3,vac,erl; me1,exc,pox)	(9.86, 90.14)	9.14
yeast0256vs3789	1004	8	(mit,cyt,me3,exc; me1,vac,pox,erl)	(9.86, 90.14)	9.14
ecoli046vs5	203	6	(cp,imU,omL; om)	(9.85, 90.15)	9.15
ecoli01vs235	244	7	(cp,im; imS,imL,om)	(9.83, 90.17)	9.17
ecoli0267vs35	224	7	(cp,imS,omL,pp; imL,om)	(9.82, 90.18)	9.18
glass04vs5	92	9	(build-win-float-proc,containers; tableware)	(9.78, 90.22)	9.22
ecoli0346vs5	205	7	(cp,imL,imU,omL; om)	(9.76, 90.24)	9.25
ecoli0347vs56	257	7	(cp,imL,imU,pp; om,omL)	(9.73, 90.27)	9.28
yeast05679vs4	528	8	(me2; mit,me3,exc,vac,erl)	(9.66, 90.34)	9.35
ecoli067vs5	220	6	(cp,omL,pp; om)	(9.09, 90.91)	10.00
vowel0	988	13	(hid; remainder)	(9.01, 90.99)	10.10
glass016vs2	192	9	(ve-win-float-proc; build-win-float-proc,	(8.89, 91.11)	10.29
			build-win-non_float-proc,headlamps)		
glass2	214	9	(Ve-win-float-proc; remainder)	(8.78, 91.22)	10.39
ecoli0147vs2356	336	7	(cp,im,imU,pp; imS,imL,om,omL)	(8.63, 91.37)	10.59
led7digit02456789vs1	443	7	(0,2,4,5,6,7,8,9; 1)	(8.35, 91.65)	10.97
glass06vs5	108	9	(build-win-float-proc,headlamps; tableware)	(8.33, 91.67)	11.00
ecoli01vs5	240	6	(cp,im; om)	(8.33, 91.67)	11.00
glass0146vs2	205	9	(build-win-float-proc,containers,headlamps,	(8.29, 91.71)	11.06
			build-win-non_float-proc;ve-win-float-proc)		
ecoli0147vs56	332	6	(cp,im,imU,pp; om,omL)	(7.53, 92.47)	12.28

Table 12.1 (continued)

Datasets	#Ex.	#Atts.	Class (maj;min)	%Class(maj; min)	IR
cleveland0vs4	177	13	(0; 4)	(7.34, 92.66)	12.62
ecoli0146vs5	280	6	(cp,im,imU,omL; om)	(7.14, 92.86)	13.00
ecoli4	336	7	(om; remainder)	(6.74, 93.26)	13.84
yeast1vs7	459	8	(nuc; vac)	(6.72, 93.28)	13.87
shuttle0vs4	1829	9	(Rad Flow; Bypass)	(6.72, 93.28)	13.87
glass4	214	9	(containers; remainder)	(6.07, 93.93)	15.47
page-blocks13vs2	472	10	(graphic; horiz.line,picture)	(5.93, 94.07)	15.85
abalone9vs18	731	8	(18; 9)	(5.65, 94.25)	16.68
glass016vs5	184	9	(tableware; build-win-float-proc,	(4.89, 95.11)	19.44
			build-win-non_float-proc,headlamps)		
shuttle2vs4	129	9	(Fpv Open; Bypass)	(4.65, 95.35)	20.5
yeast1458vs7	693	8	(vac; nuc,me2,me3,pox)	(4.33, 95.67)	22.10
glass5	214	9	(tableware; remainder)	(4.20, 95.80)	22.81
yeast2vs8	482	8	(pox; cyt)	(4.15, 95.85)	23.10
yeast4	1484	8	(me2; remainder)	(3.43, 96.57)	28.41
yeast1289vs7	947	8	(vac; nuc,cyt,pox,erl)	(3.17, 96.83)	30.56
yeast5	1484	8	(me1; remainder)	(2.96, 97.04)	32.78
ecoli0137vs26	281	7	(pp,imL; cp,im,imU,imS)	(2.49, 97.51)	39.15
yeast6	1484	8	(exc; remainder)	(2.49, 97.51)	39.15
abalone19	4174	8	(19; remainder)	(0.77, 99.23)	128.87

tests may not be satisfied [116]. As we are performing a pairwise comparison, we use the Wilcoxon test to search for statistical differences between two methods. The objective in this comparison is to obtain an adjusted p-value, which represents the lowest level of significance of a hypothesis that results in a rejection, which means the detection of significant differences between the methods.

Table 12.2 shows the average GM values in training and test obtained by the algorithms included in the comparison for the 44 highly imbalanced datasets, namely, the GP-COACH-H algorithm and the C4.5 decision tree together with SMOTE+ENN. The best average values in test per dataset are highlighted in bold.

From these results we can observe that the best performing method is GP-COACH-H, showing the good synergy between the elements integrated into this approach. The GP-COACH-H algorithm is able to obtain better results than the C4.5 decision tree combined with SMOTE+ENN, showing that the goodness in its predictive behavior is obtained through the appropriate integration of the EFS with the hierarchical approach and not by the use of the data preprocessing techniques.

Table 12.2 Detailed table of results for GP-COACH-H and SMOTE+ENN+C4.5

Dataset	GP-COACH-H		SMOTE+ENN+C4.5	
	GM_{tr}	GM_{tst}	GM_{tr}	GM_{tst}
ecoli034vs5	0.9833	0.8660	0.9762	**0.8761**
yeast2vs4	0.9647	**0.9304**	0.9745	0.9029
ecoli067vs35	0.9707	**0.7286**	0.9771	0.7206
ecoli0234vs5	0.9966	0.8473	0.9827	**0.8861**
glass015vs2	0.9503	0.6301	0.9066	**0.7788**
yeast0359vs78	0.8919	**0.7189**	0.9213	0.6894
yeast02579vs368	0.9298	0.9107	0.9572	**0.9125**
yeast0256vs3789	0.8348	**0.7982**	0.9173	0.7707
ecoli046vs5	0.9952	0.8677	0.9834	**0.8776**
ecoli01vs235	0.9845	**0.8471**	0.9649	0.8277
ecoli0267vs35	0.9707	**0.9028**	0.9825	0.8061
glass04vs5	0.9909	0.9429	0.9909	**0.9748**
ecoli0346vs5	0.9993	0.8847	0.9884	**0.8946**
ecoli0347vs56	0.9881	**0.8767**	0.9566	0.8413
yeast05679vs4	0.8961	0.6988	0.9197	**0.7678**
ecoli067vs5	0.9849	**0.8671**	0.9740	0.8376
vowel0	0.9947	**0.9465**	0.9943	0.9417
glass016vs2	0.9415	**0.6467**	0.9365	0.6063
glass2	0.9663	0.5886	0.9261	**0.7377**
ecoli0147vs2356	0.9594	**0.8263**	0.9563	0.8119
led7digit02456789vs1	0.9142	**0.9000**	0.9217	0.8370
glass06vs5	0.9975	0.9120	0.9911	**0.9628**
ecoli01vs5	0.9977	**0.8946**	0.9828	0.8081
glass0146vs2	0.9313	**0.7300**	0.9010	0.6157
ecoli0147vs56	0.9852	**0.8372**	0.9608	0.8250
cleveland0vs4	0.9719	**0.8646**	0.9819	0.7307
ecoli0146vs5	0.9952	**0.9194**	0.9850	0.8880
ecoli4	0.9936	**0.9357**	0.9826	0.8947
yeast1vs7	0.8988	0.6900	0.9093	**0.7222**
shuttle0vs4	1.0000	**1.0000**	0.9999	0.9997
glass4	0.9906	0.7303	0.9665	**0.7639**
page-blocks13vs4	0.9994	0.9482	0.9975	**0.9909**
abalone9-18	0.8595	**0.7500**	0.9273	0.6884
glass016vs5	0.9921	**0.8550**	0.9863	0.7738
shuttle2vs4	1.0000	0.9918	1.0000	**1.0000**
yeast1458vs7	0.8952	**0.6304**	0.8717	0.3345
glass5	0.9957	**0.7877**	0.9698	0.5851
yeast2vs8	0.9937	0.7381	0.8923	**0.8033**

(continued)

Table 12.2 (continued)

Dataset	GP-COACH-H		SMOTE+ENN+C4.5	
	GM_{tr}	GM_{tst}	GM_{tr}	GM_{tst}
yeast4	0.9001	**0.8175**	0.8984	0.6897
yeast1289vs7	0.8843	**0.6939**	0.9408	0.5522
yeast5	0.9724	**0.9428**	0.9819	0.9390
ecoli0137vs26	0.9843	**0.7067**	0.9650	0.7062
yeast6	0.9319	**0.8170**	0.9301	0.8029
abalone19	0.8558	**0.5532**	0.8838	0.1550
Mean	0.9576	**0.8175**	0.9549	0.7848

Table 12.3 Wilcoxon test to compare GP-COACH-H against SMOTE+ENN+C4.5 in highly imbalanced datasets

Comparison	R^+	R^-	p-Value
GP-COACH-H versus SMOTE+ENN+C4.5	667.0	323.0	0.0446

R^+ corresponds to the sum of the ranks for GP-COACH-H and R^- to C4.5

To further support this evidence, we use the Wilcoxon test [116] to develop the statistical study that aims to find statistical differences. Table 12.3 shows the rankings associated to each method and the adjusted p-value that has been calculated. The p-value obtained by the Wilcoxon test, 0.0446, is low enough to reject the null hypothesis, which means that statistical differences are found with a degree of confidence near to the 95 %.

To sum up, we have presented an EFS based approach, GP-COACH-H, to deal with highly imbalanced datasets. The approach is based on a genetic programming approach to build the KB which combines several strategies, including the use of data preprocessing, HKB and genetic tuning of the KB to enhance the performance. The experimental results have demonstrated that GP-COACH-H outperforms the C4.5 decision tree combined with data preprocessing over 44 highly imbalanced datasets, becoming a competitive method in the imbalanced classification scenario.

12.5 Concluding Remarks

In this chapter, we have reviewed the topic of EFSs focusing on the application of this type of systems for classification with imbalanced datasets.

With this aim, we have introduced some preliminaries about linguistic fuzzy rule based systems in order to set the context of this work. Then, we have presented a complete taxonomy for the current types of associated methodologies. Specifically, we have distinguished between three approaches, namely the learning of the FRBS' elements, the different schemes regarding the evolutionary components, and finally the optimization of novel fuzzy representations.

Regarding imbalanced classification, we have paid special interest in providing the design principles for those algorithms that have been used in this work area. Among them, we have divided into those solutions applied in conjunction with data level techniques, algorithm-level approaches, cost-sensitive learning, and the ones embedded into ensemble learning.

Finally, we have analyzed and evaluated the good properties and features of EFSs in this context. In order to do so, we have presented a case study with a recent EFS approach, the GP-COACH-H algorithm. Experimental results stressed the goodness of these types of approaches for addressing the problem of classification with imbalanced data.

Acknowledgments This work have been partially supported by the Spanish Ministry of Science and Technology under projects TIN-2012-33856, TIN2014-57251-P; the Andalusian Research Plans P12-TIC-2958, P11-TIC-7765 and P10-TIC-6858; and both the University of Jaén and Caja Rural Provincial de Jaén under project UJA2014/06/15.

References

1. A. Konar, *Computational Intelligence: Principles, Techniques and Applications* (Springer, Berlin, 2005)
2. R.R. Yager, D.P. Filev, *Essentials of Fuzzy Modeling and Control* (Wiley, 1994)
3. D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley Professional, Upper Saddle River, 1989)
4. J.H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press, Ann Arbor, 1975)
5. O. Cordon, F. Herrera, F. Hoffmann, L. Magdalena, Genetic fuzzy systems. *Evolutionary Tuning and Learning of Fuzzy Knowledge Bases* (World Scientific, Singapore, Republic of Singapore, 2001)
6. A. Fernandez, V. Lopez, M.J. del Jesus, F. Herrera, Revisiting evolutionary fuzzy systems: taxonomy, applications, new trends and challenges. Knowl. Based Syst. **80**, 109–121 (2015)
7. A.E. Eiben, J.E. Smith, *Introduction to Evolutionary Computation* (Springer, Berlin, 2003)
8. C.A. Coello-Coello, G. Lamont, D. van Veldhuizen, *Evolutionary Algorithms for Solving Multi-objective Problems*, Genetic and Evolutionary Computation, 2nd edn. (Springer, Berlin, 2007)
9. K. Deb, *Multi-objective Optimization using Evolutionary Algorithms* (Wiley, New York, 2001)
10. M. Fazzolari, R. Alcala, Y. Nojima, H. Ishibuchi, F. Herrera, A review of the application of multi-objective evolutionary systems: current status and further directions. IEEE Trans. Fuzzy Syst. **21**(1), 45–65 (2013)
11. J. Han, M. Kamber, J. Pei, *Data Mining: Concepts and Techniques*, 3rd edn. (Morgan Kaufmann, San Mateo, 2011)
12. V. Lopez, A. Fernandez, S. Garcia, V. Palade, F. Herrera, An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics. Inf. Sci. **250**(20), 113–141 (2013)
13. R.C. Prati, G.E.A.P.A., Batista, D.F. Silva, Class imbalance revisited: a new experimental setup to assess the performance of treatment methods. Knowl. Inf. Syst. **45**(1), 247–270 (2015)

14. A. Fernandez, S. Garcia, M.J. del Jesus, F. Herrera, A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. Fuzzy Sets Syst. **159**(18), 2378–2398 (2008)
15. S. Alshomrani, A. Bawakid, S.O. Shim, A. Fernandez, F. Herrera, A proposal for evolutionary fuzzy systems using feature weighting: dealing with overlapping in imbalanced datasets. Knowl.-Based Syst. **73**, 1–17 (2015)
16. V. Garcia, R.A. Mollineda, J.S. Sanchez, On the k-NN performance in a challenging scenario of imbalance and overlapping. Pattern Anal. Appl. **11**(3–4), 269–280 (2008)
17. A. Fernandez, M.J. del Jesus, F. Herrera, Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data-sets. Int. J. Approximate Reasoning **50**, 561–577 (2009)
18. A. Palacios, L. Sanchez, I. Couso, Equalizing imbalanced imprecise datasets for genetic fuzzy classifiers. Int. J. Comput. Intell. Syst. **5**(2), 276–296 (2012)
19. V. Lopez, A. Fernandez, M.J. del Jesus, F. Herrera, A hierarchical genetic fuzzy system based on genetic programming for addressing classification with highly imbalanced and borderline data-sets. Knowl. Based Syst. **38**, 85–104 (2013)
20. L.A. Zadeh, Fuzzy sets. Inf. Control **8**, 338–353 (1965)
21. G. Klir, B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications* (Prentice-Hall, 1995)
22. H. Zimmermann, Fuzzy set theory. WIREs Comput. Stat. **2**(3), 317–332 (2010)
23. W. Pedrycz, F. Gomide, *An Introduction to Fuzzy Sets: Analysis and Design* (Prentice-Hall, 1998)
24. L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning. part I. Inf. Sci. **8**, 199–249 (1975)
25. L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning. part II. Inf. Sci. **8**, 301–357 (1975)
26. L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning. part III. Inf. Sci. **9**, 43–80 (1975)
27. E.H. Mamdani, Applications of fuzzy algorithm for control a simple dynamic plant. Proc. Inst. Electr. Eng. **121**(12), 1585–1588 (1974)
28. W.H. Au, K.C.C. Chan, A.K.C. Wong, A fuzzy approach to partitioning continuous attributes for classification. IEEE Trans. Knowl. Data Eng. **18**(5), 715–719 (2006)
29. O. Cordon, F. Herrera, P. Villar, Analysis and guidelines to obtain a good fuzzy partition granularity for fuzzy rule-based systems using simulated annealing. Int. J. Approximate Reasoning **25**(3), 187–215 (2000)
30. E.H. Mamdani, Application of fuzzy logic to approximate reasoning using linguistic synthesis. IEEE Trans. Comput. **26**(12), 1182–1191 (1977)
31. O. Cordon, F. Gomide, F. Herrera, F. Hoffmann, L. Magdalena, Ten years of genetic fuzzy systems: current framework and new trends. Fuzzy Sets Syst. **141**, 5–31 (2004)
32. O. Cordon, A historical review of evolutionary learning methods for mamdani-type fuzzy rule-based systems: designing interpretable genetic fuzzy systems. Int. J. Approximate Reasoning **52**(6), 894–913 (2011)
33. F. Herrera, Genetic fuzzy systems: taxonomy, current research trends and prospects. Evol. Intel. **1**(1), 27–46 (2008)
34. H. Ishibuchi, K. Nozaki, N. Yamamoto, H. Tanaka, Selection of fuzzy IF-THEN rules for classification problems using genetic algorithms. IEEE Trans. Fuzzy Syst. **3**(3), 260–270 (1995)
35. A. Homaifar, E. McCormick, Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms. IEEE Trans. Fuzzy Syst. **3**(2), 129–139 (1995)
36. P. Thrift, Fuzzy logic synthesis with genetic algorithms, in *Proceedings of the 4th International Conference on Genetic Algorithms (ICGA'91)*. pp. 509–513 (1991)
37. O. Cordon, F. Herrera, P. Villar, Generating the knowledge base of a fuzzy rule-based system by the genetic learning of data base. IEEE Trans. Fuzzy Syst. **9**(4), 667–674 (2001)
38. F. Marquez, A. Peregrín, F. Herrera, Cooperative evolutionary learning of linguistic fuzzy rules and parametric aggregation connectors for mamdani fuzzy systems. IEEE Trans. Fuzzy Syst. **15**(6), 1162–1178 (2008)

39. J. Casillas, O. Cordon, M.J. del Jesus, F. Herrera, Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. IEEE Trans. Fuzzy Syst. **13**(1), 13–29 (2005)
40. J. Alcala-Fdez, F. Herrera, F.A. Marquez, A. Peregrin, Increasing fuzzy rules cooperation based on evolutionary adaptive inference systems. Int. J. Intell. Syst. **22**(9), 1035–1064 (2007)
41. D. Kim, Y. Choi, S.Y. Lee, An accurate cog defuzzifier design using lamarckian co-adaptation of learning and evolution. Fuzzy Sets Syst. **130**(2), 207–225 (2002)
42. M.J. Gacto, R. Alcala, F. Herrera, Interpretability of linguistic fuzzy rule-based systems: an overview of interpretability measures. Inf. Sci. **181**(20), 4340–4360 (2011)
43. H. Ishibuchi, T. Murata, I. Turksen, Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems. Fuzzy Sets Syst. **8**(2), 135–150 (1997)
44. M.J. Gacto, R. Alcala, F. Herrera, Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems. Soft. Comput. **13**(5), 419–436 (2009)
45. N.N. Karnik, J.M. Mendel, Q. Liang, Type-2 fuzzy logic systems. IEEE Trans. Fuzzy Syst. **7**(6), 643–658 (1999)
46. R. Sambuc, Function $\Phi-$Flous, Application a l'aide au Diagnostic en Pathologie Thyroidienne. Ph.D. thesis, University of Marseille (1975)
47. O. Castillo, P. Melin, Optimization of type-2 fuzzy systems based on bio-inspired methods: a concise review. Inf. Sci. **205**, 1–19 (2012)
48. O. Castillo, P. Melin, A.A. Garza, O. Montiel, R. Sepulveda, Optimization of interval type-2 fuzzy logic controllers using evolutionary algorithms. Soft. Comput. **15**(6), 1145–1160 (2011)
49. C. Wagner, H. Hagras, A genetic algorithm based architecture for evolving type-2 fuzzy logic controllers for real world autonomous mobile robots, in *FUZZ-IEEE*. pp. 1–6. IEEE (2007)
50. J.A. Sanz, A. Fernandez, H. Bustince, F. Herrera, Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning. Inf. Sci. **180**(19), 3674–3685 (2010)
51. J. Sanz, A. Fernandez, H. Bustince, F. Herrera, A genetic tuning to improve the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets: degree of ignorance and lateral position. Int. J. Approximate Reasoning **52**(6), 751–766 (2011)
52. J.A. Sanz, A. Fernandez, H. Bustince, F. Herrera, IVTURS: a linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection. IEEE Trans. Fuzzy Syst. **21**(3), 399–411 (2013)
53. H. He, E.A. Garcia, Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**(9), 1263–1284 (2009)
54. V. Lopez, A. Fernandez, J.G. Moreno-Torres, F. Herrera, Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics. Expert Syst. Appl. **39**(7), 6585–6608 (2012)
55. Y. Sun, A.K.C. Wong, M.S. Kamel, Classification of imbalanced data: a review. Int. J. Pattern Recognit. Artif. Intell. **23**(4), 687–719 (2009)
56. C. Elkan, The foundations of cost-sensitive learning, in *Proceedings of the 17th IEEE International Joint Conference on Artificial Intelligence (IJCAI'01)*. pp. 973–978 (2001)
57. B. Zadrozny, J. Langford, N. Abe, Cost-sensitive learning by cost-proportionate example weighting, *in Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM'03)*. pp. 435–442 (2003)
58. Q. Yang, X. Wu, 10 challenging problems in data mining research. Int. J. Inform. Technol. Decis. Making **5**(4), 597–604 (2006)
59. S.J. Lin, M.F. Hsu, Enhanced risk management by an emerging multi-agent architecture. Connection Sci. **26**(3), 245–259 (2014)
60. M. Hao, Y. Wang, S.H. Bryant, An efficient algorithm coupled with synthetic minority over-sampling technique to classify imbalanced pubchem bioassay data. Anal. Chim. Acta **806**, 117–127 (2014)

61. R. Oentaryo, E.P. Lim, M. Finegold, D. Lo, F. Zhu, C. Phua, E.Y. Cheu, G.E. Yap, K. Sim, M.N. Nguyen, K. Perera, B. Neupane, M. Faisal, Z. Aung, W.L. Woon, W. Chen, D. Patel, D. Berrar, Detecting click fraud in online advertising: a data mining approach. J. Mach. Learn. Res. **15**, 99–140 (2014)
62. W.Y. Hwang, J.S. Lee, Shifting artificial data to detect system failures. Int. Trans. Oper. Res. **22**(2), 363–378 (2015)
63. B. Krawczyk, G. Schaefer, A hybrid classifier committee for analysing asymmetry features in breast thermograms. Appl. Soft Comput. J. **20**, 112–118 (2014)
64. C. Lu, M. Mandal, Toward automatic mitotic cell detection and segmentation in multispectral histopathological images. IEEE J. Biomed. Health Inform. **18**(2), 594–605 (2014)
65. A. Orriols-Puig, E.B. Mansilla, Evolutionary rule-based systems for imbalanced datasets. Soft. Comput. **13**(3), 213–225 (2009)
66. M. Wasikowski, X.W. Chen, Combating the small sample class imbalance problem using feature selection. IEEE Trans. Knowl. Data Eng. **22**(10), 1388–1400 (2010)
67. G.M. Weiss, in *The Impact of Small Disjuncts on Classifier Learning*, eds. by R. Stahlbock, S.F. Crone, S. Lessmann. Data Mining, Annals of Information Systems, vol. 8 (Springer, 2010), pp. 193–226
68. J. Stefanowski, Overlapping, rare examples and class decomposition in learning classifiers from imbalanced data. Smart Innovation Syst. Technol. **13**, 277–306 (2013)
69. C. Seiffert, T.M. Khoshgoftaar, J. Van Hulse, A. Folleco, An empirical study of the classification performance of learners on imbalanced and noisy software quality data. Inf. Sci. **259**, 571–595 (2014)
70. J.G. Moreno-Torres, T. Raeder, R. Alaiz-Rodriguez, N.V. Chawla, F. Herrera, A unifying view on dataset shift in classification. Pattern Recogn. **45**(1), 521–530 (2012)
71. G.E.A.P.A. Batista, R.C. Prati, M.C. Monard, A study of the behaviour of several methods for balancing machine learning training data. SIGKDD Explor. **6**(1), 20–29 (2004)
72. N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority oversampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
73. V. Lopez, I. Triguero, C.J. Carmona, S. Garcia, F. Herrera, Addressing imbalanced classification with instance generation techniques: IPADE-ID. Neurocomputing **126**, 15–28 (2014)
74. P. Domingos, MetaCost: A general method for making classifiers cost-sensitive, in *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining (KDD'99)*. pp. 155–164 (1999)
75. B. Zadrozny, C. Elkan, Learning and making decisions when costs and probabilities are both unknown, in *Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining (KDD'01)*. pp. 204–213 (2001)
76. M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, F. Herrera, A review on ensembles for class imbalance problem: bagging, boosting and hybrid based approaches. IEEE Trans. Syst. Man Cybern.—Part C: Appl. Rev. **42**(4), 463–484 (2012)
77. W. Fan, S.J. Stolfo, J. Zhang, P.K. Chan, AdaCost: misclassification cost-sensitive boosting, in *Proceedings of the 16th International Conference on Machine Learning (ICML'96)*. pp. 97–105 (1999)
78. C. Seiffert, T.M. Khoshgoftaar, J. Van Hulse, A. Napolitano, RUSBoost: a hybrid approach to alleviating class imbalance. IEEE Trans. Syst. Man Cybern.—Part A **40**(1), 185–197 (2010)
79. S. Wang, X. Yao, Diversity analysis on imbalanced data sets by using ensemble models, in *Proceedings of the 2009 IEEE Symposium on Computational Intelligence and Data Mining (CIDM'09)*. pp. 324–331 (2009)
80. R. Barandela, J.S. Sanchez, V. Garcia, E. Rangel, Strategies for learning in class imbalance problems. Pattern Recogn. **36**(3), 849–851 (2003)
81. A. Fernandez, M.J. del Jesus, F. Herrera, On the influence of an adaptive inference system in fuzzy rule based classification systems for imbalanced data-sets. Expert Syst. Appl. **36**(6), 9805–9812 (2009)
82. L.J. Eshelman, The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. In: Rawlin, G. (ed.) Foundations of Genetic Algorithms, pp. 265–283. Morgan Kaufman (1991)

83. A. Fernandez, M.J. del Jesus, F. Herrera, On the 2-tuples based genetic tuning performance for fuzzy rule based classification systems in imbalanced data-sets. Inf. Sci. **180**(8), 1268–1291 (2010)
84. R. Alcala, J. Alcala-Fdez, F. Herrera, A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection. IEEE Trans. Fuzzy Syst. **15**(4), 616–635 (2007)
85. Z. Chi, H. Yan, T. Pham, Fuzzy Algorithms with Applications to Image Processing and Pattern Recognition (World Scientific, 1996)
86. H. Ishibuchi, T. Yamamoto, T. Nakashima, Hybridization of fuzzy GBML approaches for pattern classification problems. IEEE Trans. Syst. Man Cybern.—Part B **35**(2), 359–365 (2005)
87. D.L. Wilson, Asymptotic properties of nearest neighbor rules using edited data. IEEE Trans. Syst. Man Cybern. **2**(3), 408–421 (1972)
88. J. Laurikkala, Improving identification of difficult small classes by balancing class distribution, in *Proceedings of the Artificial Intelligence in Medicine, 8th Conference on AI in Medicine in Europe (AIME 2001)*. pp. 63–66 (2001)
89. P.E. Hart, The condensed nearest neighbor rule. IEEE Trans. Inf. Theory **14**, 515–516 (1968)
90. P. Villar, A. Fernandez, R.A. Carrasco, F. Herrera, Feature selection and granularity learning in genetic fuzzy rule-based classification systems for highly imbalanced data-sets. Int. J. Uncertainty Fuzziness Knowl. Based Syst. **20**(3), 369–397 (2012)
91. V. Lopez, A. Fernandez, F. Herrera, Addressing covariate shift for genetic fuzzy systems classifiers: a case of study with FARC-HD for imbalanced datasets, in *Proceedings of the 2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2013)*. pp. 1–8 (2013)
92. J.G. Moreno-Torres, J.A. Saez, F. Herrera, Study on the impact of partition-induced dataset shift on k-fold cross-validation. IEEE Trans. Neural Netw. Learn. Syst. **23**(8), 1304–1313 (2012)
93. J. Alcala-Fdez, R. Alcala, F. Herrera, A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. IEEE Trans. Fuzzy Syst. **19**(5), 857–872 (2011)
94. V. Soler, J. Cerquides, J. Sabria, J. Roig, M. Prim, Imbalanced datasets classification by fuzzy rule extraction and genetic algorithms, in *Proceedings of the 2006 IEEE International Conference on Data Mining (ICDM 2006)*. pp. 330–334 (2006)
95. V. Soler, M. Prim, in *Extracting a Fuzzy System by Using Genetic Algorithms for Imbalanced Datasets Classification: Application on Down's Syndrome Detection*, eds. by D.A. Zighed, S. Tsumoto, Z.W. Ras, H. Hacid. Mining Complex Data, Studies in Computational Intelligence, vol. 165 (Springer, 2009), pp. 23–39
96. M.R. Berthold, K.P. Huber, Constructing fuzzy graphs from examples. Intell. Data Anal. **3**(1), 37–53 (1999)
97. F.J. Berlanga, A.J. Rivera, M.J. del Jesus, F. Herrera, GP-COACH: genetic Programming-based learning of COmpact and ACcurate fuzzy rule-based classification systems for High-dimensional problems. Inf. Sci. **180**(8), 1183–1200 (2010)
98. S. Axelsson, Research in intrusion-detection systems: a survey. Technical Report 98–17, Department of Computer Engineering, Chalmers University of Technology, Goteborg, Sweden (1998)
99. W. Lee, S. Stolfo, A framework for constructing features and models for intrusion detection systems. ACM Trans. Inform. Syst. Secur. **3**(4), 227–261 (2000)
100. S.M. Gaffer, M.E. Yahia, K. Ragab, Genetic fuzzy system for intrusion detection: analysis of improving of multiclass classification accuracy using KDDCup-99 imbalance dataset, in *Proceedings of the 2012 12th International Conference on Hybrid Intelligent Systems (HIS 2012)*. pp. 318–323 (2012)
101. S. Elhag, A. Fernandez, A. Bawakid, S. Alshomrani, F. Herrera, On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on intrusion detection systems. Expert Syst. Appl. **42**(1), 193–202 (2015)
102. T. Hastie, R. Tibshirani, Classification by pairwise coupling. Ann. Stat. **26**(2), 451–471 (1998)

103. J. Sanz, D. Bernardo, F. Herrera, H. Bustince, H. Hagras, A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data. IEEE Trans. Fuzzy Syst. **23**(4), 973–990 (2015)
104. M. Mahdizadeh, M. Eftekhari, Designing fuzzy imbalanced classifier based on the subtractive clustering and genetic programming, in *Proceedings of the 13th Iranian Conference on Fuzzy Systems (IFSC 2013)*. pp. 318–323 (2013)
105. V. Lopez, A. Fernandez, F. Herrera, A first approach for cost-sensitive classification with linguistic genetic fuzzy systems in imbalanced data-sets, in *Proceedings of the 10th International Conference on Intelligent Systems Design and Applications (ISDA'10)*. pp. 676–681 (2010)
106. H. Ishibuchi, T. Yamamoto, Rule weight specification in fuzzy rule-based classification systems. IEEE Trans. Fuzzy Syst. **13**, 428–435 (2005)
107. P. Ducange, B. Lazzerini, F. Marcelloni, Multi-objective genetic fuzzy classifiers for imbalanced and cost-sensitive datasets. Soft. Comput. **14**(7), 713–728 (2010)
108. D.G. Stavrakoudis, J.B. Theocharis, G.C. Zalidis, A boosted genetic fuzzy classifier for land cover classification of remote sensing imagery. ISPRS J. Photogrammetry Remote Sens. **66**(4), 529–544 (2011)
109. R.E. Schapire, A brief introduction to boosting. In: Proceedings of the 16th International Joint Conference on Artificial Intelligence, Vol. 2 (IJCAI'99). pp. 1401–1406 (1999)
110. M. Antonelli, P. Ducange, F. Marcelloni, A. Segatori, Evolutionary fuzzy classifiers for imbalanced datasets: an experimental comparison, in *Proceedings of the 2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS 2013)*. pp. 13–18 (2013)
111. J. Huang, C.X. Ling, Using AUC and accuracy in evaluating learning algorithms. IEEE Trans. Knowl. Data Eng. **17**(3), 299–310 (2005)
112. O. Cordon, F. Herrera, I. Zwir, Linguistic modeling by hierarchical systems of linguistic rules. IEEE Trans. Fuzzy Syst. **10**(1), 2–20 (2002)
113. J. Alcala-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. Garcia, L. Sanchez, F. Herrera, KEEL data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. J. Multi-Valued Logic Soft Comput. **17**(2–3), 255–287 (2011)
114. J.R. Quinlan, *C4.5: Programs for Machine Learning* (Morgan Kaufmann Publishers, San Mateo-California, 1993)
115. S. Garcia, A. Fernandez, J. Luengo, F. Herrera, A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability. Soft. Comput. **13**(10), 959–977 (2009)
116. D. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 2nd edn. (Chapman & Hall/CRC, 2006)