Linguistic Fuzzy Rules in Data Mining: Follow-Up Mamdani Fuzzy Modeling Principle

A. Fernández and F. Herrera

Abstract. From the definition of fuzzy sets by Zadeh in 1965, fuzzy logic has become a significant area of interest for researchers on artificial intelligence. In particular, Professor Mamdani was the pioneer who investigated the use of fuzzy logic for interpreting the human derived control rules, and therefore his work was considered a milestone application of this theory.

In this work, we aim to carry out an overview of the principles of fuzzy modeling given by Mamdani and its application to different areas of data mining that can be exploited such as classification, association rule mining or subgroup discovery, among others. Specifically, we present a case of study on classification with highly imbalanced data-sets in which linguistic fuzzy rule based systems have shown to achieve a good behaviour among other techniques such as decision trees.

Keywords: Mamdani Fuzzy Rule Based Systems, Fuzzy Logic, Linguistic fuzzy partitions, Data Mining, Classification, Association rule mining, Subgroup discovery, Imbalanced data-sets.

1 Introduction

Fuzzy systems are one of the most important areas for the application of the Fuzzy Set Theory [Zadeh, 1965]. Usually it is considered a model structure in the form

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, CITIC-UGR
(Research Center on Information and Communications Technology), University of Granada, 18071 - Granada, Spain
e-mail: herrera@decsai.ugr.es

E. Trillas et al. (Eds.): Combining Experimentation and Theory, STUDFUZZ 271, pp. 103–122.
springerlink.com

of fuzzy rule based systems (FRBSs) [Yager and Filev, 1994]. FRBSs constitute an extension to classical rule-based systems, because they deal with "IF-THEN" rules, whose antecedents and consequents are composed of fuzzy logic statements, instead of classical ones.

The starting point of FRBSs is dated in 1973 when Professor Mamdani and the student S. Assilian were trying to stabilize the speed of a small steam engine. However, engine speed would either overshoot the target speed and arrive at the target speed after a series of oscillations, or the speed control was too sluggish, taking too long for the speed to arrive at the desired setting.

At that point, Dr. Mamdani decided to follow the theory proposed by Professor Zadeh and in this manner he could state that using a fuzzy logic controller for speed control of a steam engine was much superior to controlling the engine by conventional analytical control systems and logic control hardware [Mamdani and Assilian, 1975]. Dr. Mamdani found that, using the conventional approach, extensive trial and error work was necessary to arrive at successful control for a specific speed set-point. Further, due to the non-linearity of the steam engine operating characteristics, as soon as the speed set-point was changed, the trial and error effort had to be done all over again to arrive at effective control. This did not occur with the fuzzy logic controller, which adapted much better to changes, variations and non-linearity in the system.

Since then, linguistic FRBSs (also known as Mamdani FRBSs) have widely demonstrated their ability for control problems [Driankow et al, 1993] but have been also extended to numerous areas of data mining such as classification [Ishibuchi et al, 2004], association rule mining [Chan and Au, 1997], subgroup discovery [del Jesus et al, 2007] and so on.

Having this into account, the main aim of this chapter is to provide a brief overview of these applications of Mamdani FRBSs, showing how they are specifically adapted for each framework and providing a short description of their main features. With this objective, and trying to develop a self contained chapter, we will first introduce the concept of fuzzy set and linguistic variable and we will present the basic elements that compose an FRBS. Furthermore, we will enumerate the steps of the fuzzy inference system as proposed by Mamdani in order to obtain the output for an FRBS.

In order to show the significance and goodness of the use of linguistic FRBSs, we present a case of study on classification with imbalanced data-sets [He and Garcia, 2009; Sun et al, 2009], which refers to the context where the number of examples that represents one or more classes of the problem is much higher than that of the other classes. We will focus on those problems with a high degree of imbalance, showing that linguistic fuzzy models can obtain a higher precision than decision trees [Quinlan, 1993] in this domain.

The remainder of this chapter is organized as follows. In Section 2, we provide an overview of FRBSs. In Section 3, we describe the different applications of Mamdani FRBS over several topics of data mining. Next, Section 4 presents a case of study for linguistic fuzzy systems in the framework of classification with highly imbalanced data-sets. Finally, in Section 5, we provide some concluding remarks of this work.

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 Section 2.1, we introduce the important concepts of fuzzy set and linguistic variable. In Section 2.2, we deal with the basic elements of FRBSs. Finally, in Section 2.3 we describe the fuzzy inference system proposed by Mamdani for the output of an FRBS.

2.1 Preliminaries: Fuzzy Set and Linguistic Variables

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 [Zadeh, 1965], 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 takes on only two values $\{0,1\}$.

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°C 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°C in the class of high temperatures is zero. Likewise, 30°C 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°C to 30°C, the degree of membership in the fuzzy set high temperature gradually increases, as exemplified in Figure 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.

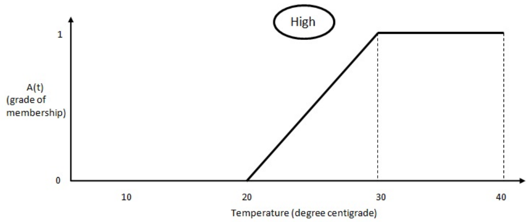

Fig. 1 Membership function

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 [Zadeh, 1975a,b,c]: $L = (x, A, X, g, m)$, 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 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.

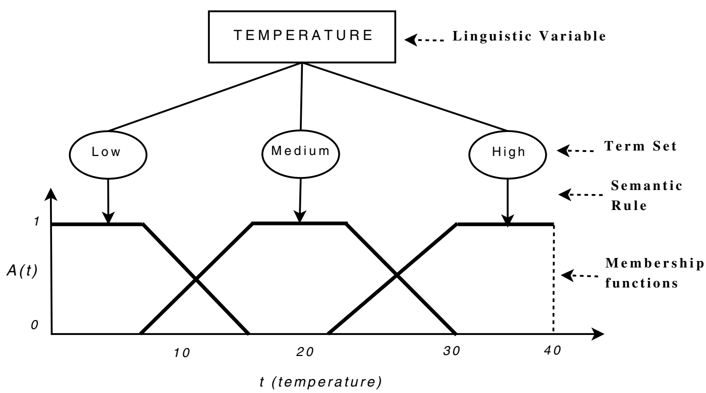

Fig. 2 Example of linguistic variable *Temperature* with three linguistic terms

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 provides a smooth and coherent transition from one state to another.

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.

As we state in the introduction of the paper, we will study linguistic models, which are based on collections of IF-THEN rules, whose antecedents are linguistic values, and the system behaviour 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 [Mamdani, 1974].

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 Figure 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 [Au et al, 2006], and the granularity of the fuzzy partition plays an important role for the FRBS behaviour [Cordón et al, 2000].

- An RB, comprised of a collection of linguistic rules that are joined by a rule connective (“also” operator). In other words, multiple rules can fire simultaneously for the same input.

The generic structure of an FRBS is shown in Figure 3.

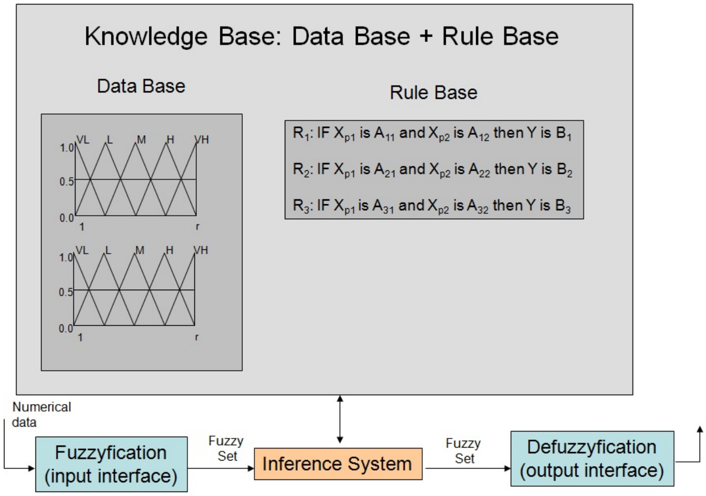

Fig. 3 Structure of an FRBS

For more information about fuzzy systems the following books may be consulted [Yager and Filev, 1994; Kuncheva, 2000; Cordón et al, 2001; Ishibuchi et al, 2004]. For different issues associated to the trade-off between interpretability and accuracy of FRBSs, the two following edited books present a collection of contributions in the topic [Casillas et al, 2003a,b].

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 output (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 [Mamdani, 1977], the fuzzy inference process comprises of five parts, which are 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 appropriate fuzzy sets via 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.

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

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.

$$\mu_{B'_j}(y) = I(\mu_{A_j}(x_p), \mu_{B_j}(y)) \qquad j = 1, \ldots, L. \tag{2}$$

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 via the *centre of gravity* of the individual fuzzy sets aggregated with the maximum connective *also*.

$$\mu_{B(y)} = \bigcup_j \mu_{B'_j}(y) \tag{3}$$

$$y_0 = \frac{\int_y y \cdot \mu_B(y)dy}{\int_y \mu_B(y)} \tag{4}$$

3 Extending Mamdani Fuzzy Rule Based Systems to Data Mining

As we have stressed in the introduction of this work, the first applications of FRBSs were focused in the field of control processes and directly to regression problems. Nevertheless, the properties of fuzzy logic make them an appropriate tool for many other fields of study, mainly because of their capability to built linguistic models interpretable to the users and the possibility of mixing different information as the one coming from expert knowledge and information coming from mathematical models or empiric measures.

For this and other reasons, the use of linguistic FRBSs has been successfully extended to the framework of data mining such as classification tasks, mining of association rules and subgroup discovery among others. In the following of this section we will briefly introduce the features of these problems and we will describe how FRBSs are adapted for each one of them.

3.1 Fuzzy Rule Based Systems for Classification

Classification is one of the most studied problems in machine learning and data mining [Duda et al, 2001; Han and Kamber, 2006]. It is a technique that, from a supervised learning point of view, consists on inducing a mapping which allow to determine the class of a new pattern from a set of attributes. A search algorithm is used to generate a classifier from a set of correctly classified patterns called training set.

Fuzzy sets have been widely employed in the field of pattern recognition and classification [Pedrycz, 1990, 1997], fundamentally because of, from the methodological point of view, the theory of fuzzy sets is an adequate theory to develop tools for modeling cognitive human processes related to the aspects of recognition. In this framework, if we join the use of fuzzy logic to the design of rule based systems, we will obtain which is known as Fuzzy Rule Based Classification Systems (FRBCSs).

We can find three different types of fuzzy classification rules in the specialised literature, which are enumerated below:

1. Rules with a single class in the consequent [Kuncheva, 1996; Nauck and Kruse, 1997]:

$$\text{Rule } R_j: \text{ If } x_{p1} \text{ is } A_{j1} \text{ and } \dots \text{ and } x_{pn} \text{ is } A_{jn} \text{ then Class} = C_j$$

2. Rules with a single class and a rule weight associated to this class in the consequent [Ishibuchi et al, 1992; Nozaki et al, 1996]:

$$\text{Rule } R_j: \text{ If } x_{p1} \text{ is } A_{j1} \text{ and } \dots \text{ and } x_{pn} \text{ is } A_{jn} \text{ then Class} = C_j \text{ with } RW_j$$

3. Rules with rule weights associated to each one of the class of the consequent [Mandal et al, 1992; Pal and Mandal, 1992]:

$$\text{Rule } R_j: \text{ If } x_{p1} \text{ is } A_{j1} \text{ and } \dots \text{ and } x_{pn} \text{ is } A_{jn} \text{ then } (RW_j^1, \dots, RW_M^1)$$

In all cases, R_j is the label of the jth rule, $x_p = (x_{p1}, \dots, x_{pn})$ is an n-dimensional pattern vector, A_{ji} is an antecedent fuzzy set representing a linguistic term, C_j is a class label, and RW_j is the rule weight [Ishibuchi and Nakashima, 2001].

For classification tasks, the fuzzy inference needs some modifications in order to adapt it to this specific problem, since the output value is no longer a fuzzy set but a class label. Therefore, if $x_p = (x_{p1}, \ldots, x_{pn})$ is a new pattern and if L denotes the number of rules in the RB and M the number of classes of the problem, the steps of the fuzzy reasoning method for classification [Cordón et al, 1999] are the following:

1. *Matching degree*, that is, *the strength of activation of the if-part for all rules in the RB with the pattern* x_p. To compute it a T-norm is used as conjunctive connector.

$$\mu_{A_j}(x_p) = C(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})), \qquad j = 1, \ldots, L. \tag{5}$$

2. *Association degree*. To compute the *association degree of the pattern* x_p *with the M classes according to each rule in the RB.*

$$\mathrm{b}_j^k = I(\mu_{A_j}(x_p), RW_j^k) \qquad k = 1, \ldots, M, \quad j = 1, \ldots, L. \tag{6}$$

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

$$Y_k = f(b_j^k,\ j = 1, \ldots, L \text{ and } b_j^k > 0), \qquad k = 1, \ldots, M. \tag{7}$$

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

$$F(Y_1, \ldots, Y_M) = \arg\max(Y_k), \qquad k = 1, \ldots, M \tag{8}$$

We must point out that the FRBCSs we have described follow the same philosophy that Mamdani's fuzzy models for regression/control problems, sharing most of their features and extending the inference mechanism as it has been explained.

3.2 Fuzzy Rule Based Systems for Association Mining

Association rules are used to represent and identify dependencies between items in a database [Zhang and Zhang, 2002]. These are an expression of the type $X \rightarrow Y$, where X and Y are sets of items and $X \cap Y = \emptyset$. This means that if all the items in X exist in a transaction then all the items in Y with a high probability are also in the transaction, and X and Y should not have any common items [Agrawal and Srikant, 1994]. Knowledge of this type of relationship can enable proactive decision making to proceed from the inferred data. Many problem domains have a need for this type of analysis, including risk management, medical diagnostic, fire management in national forests and so on.

The first studies on the topic focused on databases with binary values, however the data in real-world applications usually consist of quantitative values. In this context, different studies have proposed methods for mining fuzzy association rules from quantitative data. Specifically, Chan and Au proposed an F-APACS algorithm to mine fuzzy association rules [Chan and Au, 1997]. They first transformed quantitative attribute values into linguistic terms and then used the adjusted difference analysis to find interesting associations among attributes. In addition, both positive and negative associations could be found.

Kuok et al. proposed a fuzzy mining approach to handle numerical data in databases with attributes and derived fuzzy association rules [Kuok et al, 1998]. At nearly the same time, Hong et al. proposed a fuzzy mining algorithm to mine rules from quantitative transaction data [Hong et al, 1999] by means of linguistic terms, where the membership functions were assumed to be known in advance. Whereas these classical algorithms use a predefined DB, as we have mentioned, recent approaches on fuzzy association rule mining are focused to learn both the fuzzy rules and the membership functions of the fuzzy labels [Hong et al, 2008; Alcalá-Fdez et al, 2009].

In contrast to the standard inference modeling of a rule set, in this case rules are individually analyzed according to specifical measures of quality for evaluating their interestingness such as support and confidence. Let $R_j:$ IF X is A Then Y is B be a fuzzy association rule where $X = x_1, \ldots, x_n$ and $Y = y_1, \ldots, y_n$ are itemsets which must be disjoint between them. In order to enable the evaluation of a fuzzy association rule, we use the standard approach for calculating support and confidence, replacing the set-theoretic operations by the corresponding fuzzy set-theoretic operations:

$$Supp(A \rightarrow B) = \frac{\sum_{x_p \in B} \mu_A(x_p)}{m} \tag{9}$$

$$Conf(A \rightarrow B) = \frac{\sum_{x_p \in B} \mu_A(x_p)}{\sum_{p=1}^{m} \mu_A(x_p)} \tag{10}$$

The support measure is especially important to determine frequent itemsets with respect to the user-defined minimum support, just as in binary association rules. The confidence is particularly used for investigating the interestingness of the discovered rules. A rule will only be interesting if its confidence is above the specified minimum, and it becomes more interesting the bigger the support is. Additionally, there exists some other measures of interest that can be used to complement confidence in order to measure the goodness of the rules [Noda et al, 1999].

The reader must have realized that conceptually, fuzzy association rules follow the same scheme proposed by Mamdani in regression/control. What it is used in this case is simply their meaning as descriptive rules of information, hence having a descriptive rule set that are a particular chunk of information.

3.3 Fuzzy Rule Based Systems for Subgroup Discovery

Subgroup discovery is a data mining technique aimed at discovering interesting relationships between different objects in a set with respect to a specific property which is of interest to the user. In this way, it is somewhere halfway between supervised and unsupervised learning [Kralj-Novak et al, 2009]. Indeed, the final aim of subgroup discovery is not to perform a good classification of new examples but to cover them within the correct subgroup with a high confidence.

Due to the fact that subgroup discovery is focused on the extraction of relations with interesting characteristics, it is not necessary to obtain complete but partial relations. These relations are described in the form of individual rules. Then, a rule (R), which consists of an induced subgroup description, can be formally defined as [Lavrac et al, 2004]:

$$R : Cond \rightarrow Class$$

where *Class* is not considered as an actual class of the problem, but rather a target value or property of interest for the subgroup which appears in the consequent part of the rule; and the antecedent part *Cond* is a conjunction of features (attribute-value pairs) selected from the features describing the training instances. In this way, for the representation of the rule it is only necessary to codify the antecedent part.

Currently, some approaches make use of fuzzy logic for representing the continuous variables that form the antecedent of these rules, by means of linguistic variables, such as SDIGA [del Jesus et al, 2007], MESDIF [Berlanga et al, 2006] and NMEEF-SD [Carmona et al, 2010]. Specifically, a fuzzy rule describing a subgroup is represented in the same way as for classification tasks, where the antecedent describes the subgroup in canonical form or disjunctive normal form and the classes are treated as the target values.

One of the most important aspects in subgroup discovery is the choice of the quality measures employed to extract and evaluate the rules. There is no current consensus in the field about which are the most suitable for both processes, and there are a wide number of measures presented throughout the bibliography. For example, we can find measures of complexity, generality and precision which, in the case of FRBSs, must be computed in concordance to the properties of fuzzy logic. Specifically, most of these measures compute the number of examples covered by the rules which, in this case, is given by obtaining a positive compatibility degree of the example from a given rule, such as confidence and support which are obtained in the same way as in fuzzy association rule mining:

$$Supp(Cond^i \rightarrow Class_j) = \frac{\sum_{x_p \in Class_j} \mu_{A_i}(x_p)}{m} \tag{11}$$

$$Conf(Cond^i \rightarrow Class_j) = \frac{\sum_{x_p \in Class_j} \mu_{A_i}(x_p)}{\sum_{p=1}^{m} \mu_{A_i}(x_p)} \tag{12}$$

Despite the lack of consensus, the most commonly used metric of performance in the field of subgroup discovery is known as *Unusualness* and it is defined as the weighted relative accuracy of a rule [Lavrac et al, 1999]:

$$WRAcc(Cond^i \rightarrow Class_j) = \frac{n(Cond_i)}{m} \left(\frac{n(Class_j.Cond^i)}{n(Cond^i)} - \frac{n(Class_j)}{m} \right) \tag{13}$$

where $n(Cond^i)$ is the number of examples which verifies the condition $Cond^i$ described in the antecedent (independently of the class to which belongs), $n(Class_j.Cond^i)$ is the number of correctly covered examples of class j and $n(Class_j)$ the number of examples of the former. Therefore, the weighted relative accuracy of a rule can be described as the balance among the coverage, interest and accuracy gain of the rule. It must be noted that the higher a rule's unusualness, the more relevant it is.

We must emphasize that fuzzy rules for subgroup discovery are treated at descriptive level in a similar way as fuzzy association rules, being linked in the same manner with Mamdani's work.

4 Case of Study: Addressing Highly Imbalanced Classification Problems with Linguistic Fuzzy Rule Based Systems

In this section we present a case of study aiming to show the goodness of the application of linguistic fuzzy systems in a relevant problem such as the classification of imbalanced data-sets [He and Garcia, 2009; Sun et al, 2009], which has been identified as one of the current challenges in data mining [Yang and Wu, 2006].

In the remaining of this section we will first develop a brief introduction to the problem of imbalanced data-sets in classification and the evaluation measures employed in this topic. Next, we will describe the methodology we proposed in our former work on the topic [Fernández et al, 2009] in order to deal with the imbalance problem using linguistic hierarchical FRBCSs. Finally, we will present the experimental framework for this work, together with the tables of results and the statistical study carried out.

4.1 Imbalanced Data-Sets in Classification

We refer to imbalanced data-sets when the distribution between the classes is not uniform, being the number of examples that represents one of the classes much lower than the other, adding that the characterization of this class often has a higher practical interest [Chawla et al, 2004]. The significance of this problem relies on its presence in numerous real classification problems including, but not limited to, telecommunications, finances, biology or medicine.

Standard classification algorithms from examples are often biased towards the negative class (majority class), since the rules that correctly classify a higher number of examples are selected in the learning process while increasing the considered metric (that it is often based in the percentage of well-classified examples). Hence, the instances of the positive class (minority class) are misclassified with a higher frequency than those that belong to the negative class [Weiss, 2004]. Another important feature of this type of problems are the "small disjuncts", that is, a data concentration of one class in a small area of the problem being surrounded by examples of the contrary class [Orriols-Puig and Bernadó-Mansilla, 2009]; this type of regions are hard to detect for most of the learning algorithms. Furthermore, another main problem of imbalanced data-sets is the higher probability of overlapping between the positive and negative examples [García et al, 2008].

In order to deal with this problem, we may find external approaches that preprocess the training data in order to rebalance the class distribution prior to the learning stage [Batista et al, 2004]. According to our previous empirical results on the topic [Fernández et al, 2008], we selected the "Synthetic Minority Over-sampling Technique" (SMOTE) [Chawla et al, 2002] as preprocessing mechanism for our current study. This technique is an over-sampling method whose main idea is to form new minority class examples by interpolating between several minority class examples that lie together. Thus, the overfitting problem is avoided and causes the decision boundaries for the minority class to spread further into the majority class space. We considered only the 1-nearest neighbor to generate the synthetic samples, and we balanced both classes to the 50% distribution.

Regarding performance metric, standard quality measures for classification can lead to erroneous conclusions over imbalanced data-sets since they do not take into account the proportion of examples for each class, as we stated before. For this reason, in this work we will use the Area Under the ROC Curve (AUC) [Huang and Ling, 2005], which is defined as:

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2} \tag{14}$$

where TP_{rate} is the ratio of examples of the positive class that are well-classified and FP_{rate} is the ratio of examples of the negative class misclassified.

Finally, we must point out that there exist different imbalance degrees between the data. In this work we will use the *imbalance ratio* (IR) [Orriols-Puig and Bernadó-Mansilla, 2009] to distinguish among different categories. This metric is defined as the ratio between the number of examples of the negative class and the positive class. We consider that a data-set present a high degree of imbalance when its IR is higher than 9 (less than a 10% of instances of the positive class).

4.2 A Methodology for Dealing with Imbalanced Data-Sets with Hierarchical FRBCSs

In our previous work on the topic, we propose the use of a hierarchical environment in order to improve the behaviour of linguistic FRBCSs in the framework of imbalanced data-sets [Fernández et al, 2009]. A Hierarchical Fuzzy Rule Based Classification System (HFRBCS) [Cordón et al, 2002], is based on the refinement of a simple linguistic fuzzy model by means of the extension of the structure of the KB in a hierarchical way by using the concept of "layers", i.e. fuzzy partitions with different granularity. The final aim is the application of a thick granularity for inferring the initial RB, and a fine granularity in those areas of the problem where these low granularity rules have a bad performance.

In this manner, this approach preserves the original descriptive power and increases its accuracy by reinforcing those problem subspaces that are specially difficult. Therefore, we focus our efforts on enhancing the classification performance in the boundary areas of the problem, obtaining a good separability among the classes in an imbalanced environment.

The algorithm to obtain an HFRBCS is based on a two-stage methodology, which includes the following processes:

1. *Hierarchical KB Generation Process*: A hierarchical RB is created from a simple RB obtained by a linguistic rule generation method. In our work we employed a simple inductive rule generation method, we named as Chi et al.'s method [Chi et al, 1996]. It is worth to mention that this process is divided into two main steps: the first one identifies bad performance rules and the second one expands these rules into a higher granularity space. Finally, both "good rules" and expanded rules are joint together.
2. *Hierarchical RB Genetic Selection Process*: The best cooperative rules from the previous stage are selected by means of an evolutionary algorithm. We considered the CHC evolutionary model [Eshelman, 1991] in order to make the rule selection.

This approach allow us to get a compact set of fuzzy rules with different granularity in the fuzzy partition, adapted to each region of the data.

4.3 Experimental Study

We will study the performance of linguistic FRBCSs employing a large collection of imbalanced data-sets with high IR. Specifically, we have considered twenty-two data-sets from KEEL data-set repository [Alcalá-Fdez et al, 2010], as shown in Table 1, where we denote the number of examples (#Ex.), number of attributes (#Atts.), class name of each class (minority and majority), class attribute distribution and IR. This table is in ascendant order according to the IR.

To develop the different experiments we consider a *5-folder cross-validation model*, i.e., 5 random partitions of data with a 20%, and the combination of 4 of them (80%) as training and the remaining one as test. We must point out that the data-set partitions employed in this paper are available for download at the KEEL data-set repository (`http://www.keel.es/dataset.php`) both for the original partitions and those preprocessed data with the SMOTE method. Therefore, any interested researcher can use the same data for comparison.

We will use the following configuration for the FRBCS approach: product T-norm as conjunction operator, together with the Penalized Certainty Factor heuristic Ishibuchi and Yamamoto [2005] for the rule weight and the winning rule approach

Table 1 Summary Description for Imbalanced Data-Sets

Data-set	#Ex.	#Atts.	Class (min., maj.)	%Class(min.; maj.)	IR
Yeast2vs4	514	8	(cyt; me2)	(9.92, 90.08)	9.08
Yeast05679vs4	528	8	(me2; mit,me3,exc,vac,erl)	(9.66, 90.34)	9.35
Vowel0	988	13	(hid; remainder)	(9.01, 90.99)	10.10
Glass016vs2	192	9	(ve-win-float-proc; build-win-float-proc, build-win-non_float-proc,headlamps)	(8.89, 91.11)	10.29
Glass2	214	9	(Ve-win-float-proc; remainder)	(8.78, 91.22)	10.39
Ecoli4	336	7	(om; remainder)	(6.74, 93.26)	13.84
Yeast1vs7	459	8	(vac; nuc)	(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, build-win-non_float-proc,headlamps)	(4.89, 95.11)	19.44
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

for the fuzzy reasoning method. Furthermore, we selected the following values for the parameters in the learning method for building HFRBCSs:

- Rule Generation:
  - δ,*n(t+1)-linguistic partition* terms selector: 0.1
  - α, used to decide the expansion of the rule: 0.2
- Evolutionary Algorithm Selection:
  - Number of evaluations: 10,000
  - Population length: 61

As algorithm of comparison we have selected the well-known C4.5 decision tree [Quinlan, 1993], which is a method of reference in the field of classification with imbalanced data-sets [Batista et al, 2004; Orriols-Puig and Bernadó-Mansilla, 2009; Su and Hsiao, 2007]. We have set a confidence level of 0.25, the minimum number of item-sets per leaf was set to 2 and the application of pruning was used to obtain the final tree.

Finally, we have made use of statistical tests for performance comparison. Specifically, we apply the Wilcoxon signed-rank test [Sheskin, 2006] as non-parametric statistical procedure for performing pairwise comparisons between two algorithms. We will also compute the p-value associated to each comparison, which represents the lowest level of significance of a hypothesis that results in a rejection. In this manner, we can know whether two algorithms are significantly different and how different they are.

Table 2 shows the results in performance using the AUC metrics, for the HFRBCS methodology and C4.5 being the values for the precision grouped for the training and test partitions in the first and second columns respectively.

Table 2 Detailed table of results for the linguistic HFRBCS and C4.5 in both training and test

Data-set	HFRBCS		C4.5	
yeast2vs4	.9527 ± .0104	.8952 ± .0373	.9814 ± .0088	.8588 ± .0878
yeast05679vs4	.9296 ± .0107	.7475 ± .0608	.9526 ± .0094	.7602 ± .0936
vowel0	.9999 ± .0003	.9883 ± .0160	.9967 ± .0048	.9494 ± .0495
glass016vs2	.8766 ± .0233	.6262 ± .1824	.9716 ± .0186	.6062 ± .1266
glass2	.8390 ± .0146	.5695 ± .1929	.9571 ± .0151	.5424 ± .1401
ecoli4	.9870 ± .0140	.9325 ± .0788	.9769 ± .0196	.8310 ± .0990
shuttle0vs4	1.000 ± .0000	.9912 ± .0114	.9999 ± .0002	.9997 ± .0007
yeast1vs7	.9181 ± .0221	.7234 ± .1016	.9351 ± .0220	.7003 ± .0146
glass4	.9981 ± .0017	.8059 ± .1995	.9844 ± .0229	.8508 ± .0935
page-blocks13vs4	.9989 ±.0012	.9865 ± .0064	.9975 ± .0021	.9955 ± .0047
abalone9-18	.8367 ± .0290	.7108 ± .0790	.9531 ± .0444	.6215 ± .0496
glass016vs5	.9971 ± .0030	.8743 ± .2257	.9921 ± .0047	.8129 ± .2444
shuttle2vs4	.9990 ± .0023	.9755 ± .0263	.9990 ± .0023	.9917 ± .0186
yeast1458vs7	.9076 ± .0136	.6474 ± .0454	.9158 ± .0278	.5367 ± .0209
glass5	.9768 ± .0215	.7988 ± .1842	.9976 ± .0040	.8829 ± .1331
yeast2vs8	.8462 ± .0139	.7685 ± .1066	.9125 ± .0184	.8066 ± .1122
yeast4	.9002 ± .0194	.8293 ± .0205	.9101 ± .0264	.7004 ± .0565
yeast1289vs7	.8713 ± .0229	.7040 ± .0343	.9465 ± .0113	.6832 ± .0616
yeast5	.9785 ± .0032	.9427 ± .0257	.9777 ± .0145	.9233 ± .0472
yeast6	.9344 ± .0174	.8619 ± .1077	.9242 ± .0354	.8280 ± .1277
ecoli0137vs26	.9868 ± .0078	.8226 ± .2103	.9678 ± .0328	.8136 ± .2168
abalone19	.8405 ± .0307	.7001 ± .1070	.8544 ± .0249	.5202 ± .0441
average	.9352 ± .0129	**.8137** ± **.0936**	.9593 ± .0168	.7825 ± .0838

As it can be observed, the prediction ability obtained by the linguistic HFRBCS is higher than that of C4.5, showing the goodness of the use of fuzzy systems achieving a high classification accuracy in this context of highly imbalanced data-sets and therefore emphasizing the robustness of this approach. We must also stress the significance of these results since the obtained fuzzy model has an implicit high interpretability because of the inclusion of linguistic fuzzy terms in the antecedents.

In order to validate these results, we perform a Wilcoxon test for detecting significant differences between the results of HFRBCS and C4.5. The result of this test is shown in Table 3, where we observe that the fuzzy approach clearly outperforms C4.5 considering a degree of confidence over the 95%.

Table 3 Wilcoxon Test to compare the HFRBCS method with C4.5 regarding the AUC metric. R^+ stands for the sum of the ranks for the first method and R^- for the second.

Comparison	R^+	R^-	Hypothesis $\alpha = 0.05$	p-value
HFRBCS vs. C4.5	195	58	Rejected for HFRBCS	0.026

In brief, it is possible to improve the behaviour of the linguistic FRBCS by a simple and effective methodology, that is, applying a higher granularity in the areas where the RB has a bad performance in order to obtain a better coverage of that area of the space of solutions.

5 Concluding Remarks

In this work, we have discussed the extension of the use of linguistic fuzzy rules in order to represent the information in numerous areas of data mining such as classification, association rule mining or subgroup discovery among others. We have shown the specific features for linguistic FRBSs for adapting them to each case, also providing a brief description of their use and most significant characteristics.

Finally, we have proved the usefulness of linguistic FRBCSs within an emerging and significant problem in data mining such as the classification of imbalanced datasets and specifically for those with a high imbalance degree. Specifically, we have shown the good behaviour of a linguistic hierarchical FRBCS, enhancing the classification performance in the overlapping areas between the minority and majority classes and outperforming the well-known C4.5 decision tree.

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

References

Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Bocca, J.B., Jarke, M., Zaniolo, C. (eds.) Proceedings of 20th International Conference on Very Large Data Bases, VLDB, pp. 487–499. Morgan Kaufmann (1994)

Alcalá-Fdez, J., Alcalá, R., Gacto, M.J., Herrera, F.: Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms. Fuzzy Sets and Systems 160(7), 905–921 (2009)

Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.: KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. Journal of Multi-Valued Logic and Soft Computing (2010) (in press)

Au, W.H., Chan, K.C.C., Wong, A.K.C.: A fuzzy approach to partitioning continuous attributes for classification. IEEE Transactions on Knowledge Data Engeneering 18(5), 715–719 (2006)

Batista, G., Prati, R.C., Monard, M.C.: A study of the behaviour of several methods for balancing machine learning training data. SIGKDD Explorations 6(1), 20–29 (2004)

Berlanga, F.J., del Jesus, M.J., González, P., Herrera, F., Mesonero, M.: Multiobjective Evolutionary Induction of Subgroup Discovery Fuzzy Rules: A Case Study in Marketing. In: Perner, P. (ed.) ICDM 2006. LNCS (LNAI), vol. 4065, pp. 337–349. Springer, Heidelberg (2006)

Carmona, C.J., González, P., del Jesus, M.J., Herrera, F.: NMEEF-SD: Non-dominated multiobjective evolutionary algorithm for extracting fuzzy rules in subgroup discovery. IEEE Transactions on Fuzzy Systems (2010) (in press); doi:10.1109/TFUZZ.2010.2060200

Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (eds.): Accuracy improvements in linguistic fuzzy modeling. Springer, Berlin (2003a)

Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (eds.): Interpretability issues in fuzzy modeling. Springer, Berlin (2003b)

Chan, C.C., Au, W.H.: Mining fuzzy association rules. In: Conference on Information and Knowledge Management, vol. 215, pp. 209–215 (1997)

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: Synthetic minority over–sampling technique. Journal of Artificial Intelligent Research 16, 321–357 (2002)

Chawla, N.V., Japkowicz, N., Kolcz, A.: Editorial: special issue on learning from imbalanced data sets. SIGKDD Explorations 6(1), 1–6 (2004)

Chi, Z., Yan, H., Pham, T.: Fuzzy algorithms with applications to image processing and pattern recognition. World Scientific (1996)

Cordón, O., del Jesus, M.J., Herrera, F.: A proposal on reasoning methods in fuzzy rule-based classification systems. International Journal of Approximate Reasoning 20(1), 21–45 (1999)

Cordón, O., Herrera, F., Villar, P.: Analysis and guidelines to obtain a good fuzzy partition granularity for fuzzy rule-based systems using simulated annealing. International Journal of Approximate Reasoning 25(3), 187–215 (2000)

Cordón, O., Herrera, F., Hoffmann, F., Magdalena, L.: Genetic fuzzy systems. In: Evolutionary tuning and learning of fuzzy knowledge bases. World Scientific, Singapore (2001)

Cordón, O., Herrera, F., Zwir, I.: Linguistic modeling by hierarchical systems of linguistic rules. IEEE Transactions on Fuzzy Systems 10(1), 2–20 (2002)

Driankow, D., Hellendoorn, H., Reinfrank, M.: An introduction to fuzzy control. Springer, Berlin (1993)

Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley, New York (2001)

Eshelman, L.J.: The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. In: Foundations of Genetic Algorithms, pp. 265–283. Morgan Kaufman (1991)

Fernández, A., García, S., del Jesus, M.J., Herrera, F.: A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data–sets. Fuzzy Sets and Systems 159(18), 2378–2398 (2008)

Fernández, A., del Jesus, M.J., Herrera, F.: Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data–sets. International Journal of Approximate Reasoning 50, 561–577 (2009)

García, V., Mollineda, R., Sánchez, J.S.: On the k-NN performance in a challenging scenario of imbalance and overlapping. Pattern Analysis Applications 11(3-4), 269–280 (2008)

Han, J., Kamber, M.: Data Mining - Concepts and Techniques, 2nd edn. Morgan Kaufmann, San Francisco (2006)

He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Transactions on Knowledge And Data Engineering 21(9), 1263–1284 (2009)

Hong, T.P., Kuo, C.S., Chi, S.C.: Mining association rules from quantitative data. Intelligent Data Analysis 3(5), 363–376 (1999)

Hong, T.P., Chen, C.H., Lee, Y.C., Wu, Y.L.: Genetic-fuzzy data mining with divide-and-conquer strategy. IEEE Transactions on Evolutionary Computation 12(2), 252–265 (2008)

Huang, J., Ling, C.X.: Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on Knowledge and Data Engineering 17(3), 299–310 (2005)

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

Ishibuchi, H., Yamamoto, T.: Rule weight specification in fuzzy rule–based classification systems. IEEE Transactions on Fuzzy Systems 13, 428–435 (2005)

Ishibuchi, H., Nozaki, K., Tanaka, H.: Distributed representation of fuzzy rules and its application to pattern classification. Fuzzy Sets and Systems 52(1), 21–32 (1992)

Ishibuchi, H., Nakashima, T., Nii, M.: Classification and modeling with linguistic information granules: Advanced approaches to linguistic data mining. Springer, Berlin (2004)

del Jesus, M.J., González, P., Herrera, F., Mesonero, M.: Evolutionary Fuzzy Rule Induction Process for Subgroup Discovery: A case study in marketing. IEEE Transactions on Fuzzy Systems 15(4), 578–592 (2007)

Kralj-Novak, P., Lavrac, N., Webb, G.I.: Supervised descriptive rule discovery: A unifying survey of constrast set, emerging pattern and subgroup mining. Journal of Machine Learning Research 10, 377–403 (2009)

Kuncheva, L.I.: On the equivalence between fuzzy and statistical classifiers. International Journal of Uncertainty, Fuzzyness and Knowledge Based Systems 4(3), 245–253 (1996)

Kuncheva, L.I.: Fuzzy classifier design. Springer, Berlin (2000)

Kuok, C.M., Fu, A.W.C., Wong, M.H.: Mining fuzzy association rules in databases. SIGMOD Record 27(1), 41–46 (1998)

Lavrač, N., Flach, P.A., Zupan, B.: Rule Evaluation Measures: A Unifying View. In: Džeroski, S., Flach, P.A. (eds.) ILP 1999. LNCS (LNAI), vol. 1634, p. 174. Springer, Heidelberg (1999)

Lavrac, N., Cestnik, B., Gamberger, D., Flach, P.A.: Decision Support Through Subgroup Discovery: Three Case Studies and the Lessons Learned. Machine Learning 57(1-2), 115–143 (2004)

Mamdani, E.H.: Applications of fuzzy algorithm for control a simple dynamic plant. Proceedings of the IEEE 121(12), 1585–1588 (1974)

Mamdani, E.H.: Application of fuzzy logic to approximate reasoning using linguistic synthesis. IEEE Transactions on Computers 26(12), 1182–1191 (1977)

Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies 7(1), 1–13 (1975)

Mandal, D.P., Murphy, C.A., Pal, S.K.: Formulation a multivalued recognition system. IEEE Transations on System man and cibernetics 22(4), 607–620 (1992)

Nauck, D., Kruse, R.: A neuro-fuzzy method to learn fuzzy classification rules from data. Fuzzy Sets and Systems 89(3), 277–288 (1997)

Noda, E., Freitas, A.A., Lopes, H.S.: Discovering interesting prediction rules with a genetic algorithm. In: Proceedings of the Congress on Evolutionary Computation, pp. 1322–1329 (1999)

Nozaki, K., Ishibuchi, H., Tanaka, H.: Adaptive fuzzy rule-based classification systems. IEEE Transactions on Fuzzy Systems 4(3), 238–250 (1996)

Orriols-Puig, A., Bernadó-Mansilla, E.: Evolutionary rule–based systems for imbalanced datasets. Soft Computing 13(3), 213–225 (2009)

Pal, S.K., Mandal, D.P.: Linguistic recognition system based on approximate reasoning. Information Sciences 61(1-2), 135–161 (1992)

Pedrycz, W.: Fuzzy sets in pattern recognition: Methodology and methods. Pattern Recognition 23(1-2), 121–146 (1990)

Pedrycz, W.: Fuzzy sets in pattern recognition: Accomplishments and challenges. Fuzzy sets and systems 90, 171–176 (1997)

Quinlan, J.R.: C4. Machine Learning. Morgan Kaufmann Publishers, San Mateo–California (1993)

Sheskin, D.: Handbook of parametric and nonparametric statistical procedures, 2nd edn. Chapman and Hall, CRC (2006)

Su, C.T., Hsiao, Y.H.: An evaluation of the robustness of MTS for imbalanced data. IEEE Transactions on Knowledge Data Engineering 19(10), 1321–1332 (2007)

Sun, Y., Wong, A.K.C., Kamel, M.S.: Classification of imbalanced data: A review. International Journal of Pattern Recognition and Artificial Intelligence 23(4), 687–719 (2009)

Weiss, G.M.: Mining with rarity: a unifying framework. SIGKDD Explorations 6(1), 7–19 (2004)

Yager, R.R., Filev, D.P.: Essentials of fuzzy modeling and control. John Wiley & Sons (1994)

Yang, Q., Wu, X.: 10 challenging problems in data mining research. International Journal of Information Technology and Decision Making 5(4), 597–604 (2006)

Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)

Zadeh, L.A.: The concept of a linguistic variable and its applications to approximate reasoning. part I. Information Sciences 8, 199–249 (1975a)

Zadeh, L.A.: The concept of a linguistic variable and its applications to approximate reasoning. part II. Information Sciences 8, 301–357 (1975b)

Zadeh, L.A.: The concept of a linguistic variable and its applications to approximate reasoning. part III. Information Sciences 9, 43–80 (1975c)

Zhang, C., Zhang, S.: Association Rule Mining. LNCS (LNAI), vol. 2307. Springer, Heidelberg (2002)