A Proposal to Solve Rule Conflicts in the Wang-Mendel Algorithm for Fuzzy Classification Using Evidential Theory

Diego Alvarez-Estevez[1*][0000-0001-5790-0577] and Vicente Moret-Bonillo[2][0000-0002-9435-3151]

[1] Haaglanden Medisch Centrum, The Hague, The Netherlands
diego.alvarez@udc.es
[2] University of A Coruña, A Coruña, Spain
vicente.moret@udc.es

Abstract. This paper addresses the problem of solving rule conflicts in a modified version of the Wang-Mendel algorithm for the induction of fuzzy classification rules. At this respect we propose a solution based on the reinterpretation of the conflict resolution mechanism as an evidence reassignment problem. The Evidential-theory framework developed by Dempster and Shafer is used for this purpose, and different alternative conflict handling strategies are explored. Experiments are carried out using a benchmark of well-known classification problems. Our preliminary results are encouraging toward supporting the usefulness of the proposed approach.

Keywords: Learning from examples, Wang-Mendel algorithm, Fuzzy classification, Evidential Theory.

1 Introduction

Several methods are available in the literature for the induction of fuzzy rules in the context of supervised learning, also often referred as learning from examples. Examples can be found in fields such as function approximation [1-3], clustering [4-5], and of course, classification [6-9]. In this paper we focus on the domain of solving classification problems. As such, given a set of N input patterns of the form $\overrightarrow{x_i} = (x_{i1}, x_{i2}, \dots, x_{im}), i = 1..N$ and a set of output classes $C = \{C_1, C_2, \dots, C_K\}$, then the classification problem consists in finding the appropriate mapping f such that $f(\overrightarrow{x_i})$ returns the class $C_i \in C$ to which the pattern $\overrightarrow{x_i}$ belongs, $\forall i = 1..N$. The goal is therefore to find the best possible approximation $\widehat{f(\vec{x})}$ of $f(\vec{x})$. Hence, ideally one would like that $f(\vec{x}) = \widehat{f(\vec{x})}$. Notice that we are implicitly assuming that each input pattern has one dominant class membership in *C*, that is, it should be always possible to find *j* such that $\mu_j(\vec{x}) > \mu_i(\vec{x}), \forall i \neq j, i = 1..K$. We are, on purpose, not considering the case in which the same pattern could be assigned to two or more classes with exactly the same degree of membership. Further, in the context of this work, we will interpret this situation as a *conflict*.

In particular in this paper we analyze the case of the Wang-Mendel (WM) algorithm for the induction of fuzzy IF-THEN rules [10]. Specifically we focus on the version adapted by Chi et al. [11] for pattern classification (hereafter referred as the WM/Chi method). As it will be described later on, the WM/Chi method for the generation of fuzzy classification rules includes, as part of the algorithm, a rule conflict resolution step by which, among the generated rules that share the same antecedent while pointing to different classes, one winner rule has to be chosen to solve the conflict, discarding the rest. The original algorithm (both on its regression and classification forms) proposes a method to choose the winner rule based on the calculation of ad-hoc scorings over the involved conflicting rules. Different methods to obtain these scorings have been proposed, for example, in Chi et al. [11] and in Cordón et al. [12]. In this work, we contribute by exploring an alternative approach by reinterpreting the conflict resolution mechanism as an evidence reassignment problem, and for that purpose we explore the use of the Evidential Theory (ET) framework developed in the past by Dempster and Shafer [13]. The idea is explained throughout the subsequent sections, and experiments are carried out to test this approach by using several well-known machine learning classification benchmarks.

2 Methods

2.1 Overview of the WM Algorithm

The WM algorithm for the induction of fuzzy rules was one of the first approaches to design fuzzy rule based systems by learning from examples [10]. The interest in this method resides in its simplicity and its straightforward approach, which at the same time has proven capabilities to provide reasonably good performance. These properties have made of this algorithm one of reference benchmark methods in the field [9, 14-17]. Originally the WM algorithm was developed to be used in Mamdani-type fuzzy reasoning systems on the context of regression and function approximation [10]. A first extension of this algorithm for its use in classification problems was proposed by Chi et al. [11]. Later on, Cordón et al. used the Chi's et al. approach to investigate the consequences of using different rules types, including the possibility to assign weights to the generated rules. They also explored the use of different aggregation strategies [12]. In Wang [14] the WM method is reviewed presenting it as a flexible fuzzy system approach to data mining. In this review, the original method for function approximation is extended, but also classification problems are considered using a piecewise constant fuzzy predictive model. In this work we will focus on the Chi's et al. [11] formulation by considering the mapping function of the system as a direct application from $\mathbb{R}^m$ to C. This approximation has the advantage of avoiding the necessity to define fuzzy partitions over the output space. In our view, this qualifies for a simpler and more natural classification approach.

2.2 The WM/Chi Method for Classification

Let us assume that a set of N input-output patterns of the form

$$(\vec{x_i}, C_i), \vec{x_i} \epsilon R^m, C_i \epsilon C, i = 1,2, \ldots, N \tag{1}$$

are available, which are split into two disjoint subsets. We refer to them as the training (TR) and the testing (TS) sets, respectively, containing N_{TR} and N_{TS} input-output patterns each, therefore $N = N_{TR}+N_{TS}$. Our learning algorithm is allowed to use only the information contained in TR in order to construct $\widehat{f(\vec{x})}$, i.e. to infer a set of R rules that would integrate the so-called Knowledge Base (KB) of the system.

In the following we describe the WM/Chi method for classification problems (for detailed information see [11]). The algorithm is composed of three steps:

-Step 1: Divide the Input Space into Fuzzy Regions

The original method does not impose any specific partition for the input variables, but it proposes an equally spaced partition based on triangular fuzzy sets [10]. In this kind of partition each domain interval k, k=1…m, is divided into $Q_k = 2n + 1$ regions, n being a natural number, where the center of each corresponding membership function (MF) lies in the center of the region, and the extrema lie at the center of the neighboring regions. Several other methods for fuzzy partition design can be used as well which can be found in the literature [4, 14].

-Step 2: Generate Fuzzy Rules from the Given Data Pairs

For each input-output pair $(\vec{x_p}, C_p) \in TR$, $\vec{x_p} \epsilon R^m$, p = 1.. N_{TR}, a rule Rp is generated by computing the membership values $\mu_{jk}(x_{pk})$ for k=1…m, j=1...Q_k. For each k then select the fuzzy set $A_{j_{max}k}$ in which the maximum membership takes place, j_{max} = $\{j \mid \mu_{j_{max}k}(x_{pk}) \geq \mu_{jk}(x_{pk})$ for all j=1...$Q_k\}$.

The following rule is then generated:

$$Rule\ R_p\!: IF\ x_1\ is\ A_{j_{max}1}\ \wedge \ldots \wedge\ x_m\ is\ A_{j_{max}m}, THEN\ C_p \tag{2}$$

The process is repeated for all the training patterns and an initial set of N_{TR} “and” rules is therefore obtained as the output.

-Step 3: Remove conflicting rules

Each training pattern generates a rule, and therefore there might be rules that “conflict” with each other as result. The definition of “conflict” is of fundamental importance here. In the original paper for function approximation a conflict is considered when two or more rules share the same antecedent part while having different consequents [10]. The natural extension for classification followed by Chi et al. on their adaptation for classification [11] is to consider a conflict when two or more rules have the same antecedent and output to different classes. Indeed, one should notice that allowing rules to share the same antecedent while pointing to different classes is likely (depending on the aggregation strategy) to cause the conflicting situation of the same input being assigned to different classes with the same activation strength. This is a problem when one has to unambiguously assign a unique final class label. Indeed, this is especially the case, for example, when the fuzzy inference engine does only use the rule with the highest activation to perform the classification (the so-called one-winner aggregation approach, as it was the case in the original WM method [10]).

To avoid this problem a conflict resolution strategy was originally proposed in Wang and Mendel [10] and in Chi et al. [11], based on the scoring of the conflicting rules. Using this mechanism, only the rule with the maximum score is selected to be included in the final KB, discarding the rest. To compute the score, in Chi et al. [11] the following custom product strategy based on the respective associated input activation was proposed:

$$s_p = \prod_{k=1}^{m} \mu_{j_{max}k}(x_{pk}) \tag{3}$$

This is an adaptation for classification of the original scoring rule proposed in Wang and Mendel [10]. Alternative approaches have been proposed too, for example, a scoring approach based on the relative proportion of patterns supporting each of the conflicting rules was also suggested in Chi et al. [11]. On the other hand, in the unlikely situation of a tie in the scorings, then the conflict resolution mechanism should provide of a complementary selection criterion. This situation however, is not explicitly mentioned in the original method. One simple possibility is just to pick one of the conflicting rules at random, or else use any other alternative (tie breaking) criterion.

As a result of this step, we obtain a subset of R rules, $R \leq N_{TR}$, where conflicting rules have been removed. This set of rules defines the final KB of the fuzzy inference system.

2.3 Conflict Resolution Using Evidential Theory

While the original proposal of choosing the rule with the maximum score seems in principle straightforward, this criterion explicitly discards all the information carried out by the non-selected rules. However, it seems logical to consider that this discarded information might relevantly contribute to the final output. After all, the existence of these (preliminary) rules is a consequence of the information carried out by the (training) data, as described in step 1) of the WM/Chi algorithm. Depending on the amount data available, the number of output classes and the schema of the fuzzy rules (i.e. for example we could allow a simple rule to point to more than one class at the same time, see for example [12]), the complexity of the relationships can rapidly increase, and the amount of discarded information might indeed be relevant and have potential contribution to the final optimal decision. In this scenario the application of a more formal and powerful evidence combinational model seems adequate.

The Evidential Theory (ET) framework proposed by Dempster and Shafer is, at this respect, attractive, among others for which: (*i*) it easily allows modelling of uncertainty associated to evidences and hypotheses, (*ii*) it allows considering sets of hypotheses without the need that the confidence placed on a particular hypothesis spreads in any particular mode on each of the remaining ones, (*iii*) it elegantly reflects the lack of knowledge, so commonly associated to reasoning processes, (*iv*) it is a formal model which contains probability theory as a particular case, and also some of the evidence combination functions of other effective (but ad-hoc) models such as the model of certainty factors, developed by Shortliffe and Buchanan, and used in the well-known MYCIN system [18].

Our proposal for the reinterpretation of the conflict resolution problem in the context of the ET framework is rather simple. As such, when two or more rules conflict (according to the above mentioned notion of conflict) then we might just consider that each rule's output class accounts for a possible hypothesis, and that each individual rule's associated score (regardless of the specific method for computing this score) is an independent item of evidence pointing to the consequent hypothesis.

Let us consider the following simple example in the context of a two-class problem, in which we have three conflicting rules R_1, R_2, and R_3, pointing respectively to classes C_1, C_1, and C_2. Let us also assume the rule scoring approach provides the following respective rule scores s = {0.7, 0.1, 0.6}. According to the above described reinterpretation, and put in terms of the ET framework, then the universe set representing all possible hypotheses of the domain U, results $U = \{H_1, H_2\}$, or equivalently $U = \{C_1, C_2\}$. Consequently the power set $\Omega = 2^U = \{\emptyset, \{C_1\}, \{C_2\}, \{C_3\}, \{C_1, C_2\}=U\}$, contains all the possible subsets of U, including the empty set $\emptyset$. The resulting Basic Belief Assignment (BBA) function, $m: \Omega \rightarrow [0,1]$, assigns the following belief masses due to the conflicting rules:

$$R_1: m_1(C_1)=0.70 \Rightarrow m_1(\Omega_1)=1-0.70=0.30$$

$$R_2: m_2(C_1)=0.10 \Rightarrow m_2(\Omega_2)=1-0.10=0.90$$

$$R_3: m_3(C_2)=0.60 \Rightarrow m_3(\Omega_3)=1-0.60=0.40$$

Taking into account that

$$\sum_{A \subseteq U} m(A) = 1 \qquad m(\emptyset) = 0 \tag{4}$$

and applying the Dempster's rule of combination, such that

$$m_{1,2}(Z) = \frac{\sum_{Z=X_i \cap Y_j} m_1(X_i) m_2(Y_j)}{\sum_{X_i \cap Y_j \neq \emptyset} m_1(X_i) m_2(Y_j)} \tag{5}$$

we obtain the following final mass assignments:

$$m(C_1)=0.52$$

$$m(C_2)=0.29$$

$$m(\Omega)=0.19$$

At this point the conflict resolution criterion might be modified to choose one of the conflicting rules, as the one whose class associated mass assignment is the maximum (*criterion A*). Using this criterion notice that the result matches the one of the classical approach in the sense that $max\{0.7, 0.1, 0.6\} = 0.7 \Rightarrow C_1$. In this case the classical conflict resolution approach and the proposal using ET agree. However, we may just consider exactly the same example, but adding one additional rule to the conflicting set: R_4 -> C_2 with associated score s_4 = 0.5. The interested reader might do the corresponding calculations to obtain in this case that:

$$m(C_1)=0.35$$

$$m(C_2)=0.52$$

$$m(\Omega)=0.13$$

and therefore, in this case the selected class would be C_2, which differs from the analogous result applying the original WM/Chi criterion: $max\{0.7, 0.1, 0.6, 0.5\}= 0.7$ => C_1.

Another possible rule conflict resolution criterion (*criterion B*) might be to allow the implementation of all the possible rules in the KB, by assigning a rule's weight in each case, equal to the resulting final mass assignments. Actually, using this alternative conflict resolution criterion, we are not selecting a unique output class for the antecedent to solve the conflict, but allowing several rules to have the same antecedent while pointing to different classes with different strength a priori. That is, our KB will be composed in fact of rules of the form:

$$Rule\ R_p: IF\ x_1\ is\ A_{j_{max}1}\ \wedge \dots \wedge\ x_m\ is\ A_{j_{max}m}, THEN\ C_p \text{ with } w_p = m(C_p) \quad (6)$$

Experiments will be carried out in the next sections to compare both approaches.

3 Experimental Setup

The main objective here is to evaluate the proposed approach as described in Section 2. For this purpose we have chosen to use a set of well-known classification benchmarks selected from the University of Carolina Irvine (UCI) machine learning repository [19]. For each dataset we schedule 10x10-fold cross-validation in which, for each fold, we apply the WM/Chi method to build the KB and evaluate the performance over the corresponding TS set. For comparison purposes, for each iteration, the resolution of conflicting rules in step 3 of the algorithm will be performed applying the classical approach as stated originally by Wang and Mendel [10] and later by Chi et al. [11] (baseline method), and then using the two alternative criteria (*A* and *B*) using the ET framework, as described before in Section 2.3.

Table 1 characterizes the selected datasets. In Table 1 the first column refers to the name of the dataset (for detailed information please check [19]), the second column shows the number of instances that form the dataset, column three indicates the input dimensionality, and the last column shows the number of output classes and their corresponding distribution.

Note that during the cross-validation process all datasets are partitioned using the original class distribution according to column four of Table 1, thus no class distribution normalization is attempted. In our experiments, for all the datasets, the input space is partitioned using three equally spaced triangular fuzzy sets.

Table 1. Summary characterization of the classification benchmarks selected for the tests. In the fourth column the notation "number of classes: distribution" is used

Dataset	Patterns	Attributes	Classes
IRIS	150	4	3: [50, 50, 50]
WINE	178	13	3: [59, 71, 48]
BREAST CANCER	569	32	2: [257, 212]
GLASS	214	9	6: [70, 76, 17, 13, 9, 29]
PIMA	768	8	2: [500, 268]

Two specific realizations of a general fuzzy inference engine are tested: (a) using the classical one-winner reasoning method (*max*) in which only the rule with the highest activation is used to classify the current input pattern; and (b) using a general reasoning model that combines the information provided by all the activated rules, in our case using a normalized addition aggregation operator (*sum*). The objective is to test the validity of the previous analyses of Cordon et al. using WM/Chi approach, who hypothesized that the one-winner approach would tend to misuse the information provided by the places of overlapping fuzzy subsets, usually offering worst results than combinative approaches [12]. In both approaches the product is used to implement the *T-norm* operator that calculates the matching degree between each rule R_j and the current input pattern $\vec{x}$. The following fuzzy inference engines therefore result on each case:

In the case of *max*:

$$y_k(\vec{x}) = max_{j=1}^{R_k}\left\{w_j \prod_{i=1}^{m} \mu_{ji}(x_i)\right\} \tag{7}$$

In the case of *sum*:

$$y_k(\vec{x}) = \frac{\sum_{j=1}^{R_k}\{w_j \prod_{i=1}^{m} \mu_{ji}(x_i)\}}{\sum_{j=1}^{R}\{w_j \prod_{i=1}^{m} \mu_{ji}(x_i)\}} \tag{8}$$

where the index k denotes the corresponding output class, and m is the input dimensionality. Notice as well the rule weighting term w_j is introduced which accounts for the weight associated to the rule R_j in the case of (6), and otherwise $w_j = 1$ which is equivalent to (2) (i.e. when applying the original WM/Chi´s formulation).

Finally, the score associated to each of the rules i involved in a conflict would be here calculated using (3) according to the original proposal in Wang and Mendel [10] and in Chi et al. [11], i.e.:

$$s_i = \bigwedge_{k=1}^{m} \mu_{A_k} = \prod_{k=1}^{m} \mu_{j_{max}k}(x_{pk}) \tag{9}$$

Notice this has been chosen on the matter of allowing an easier comparison with the original method, however, some other approaches are available, such as the ones proposed in Cordón et al. [12] or in Chi et al. [11] (strategy 2).

4 Results

Experimental results are shown in Tables 2-3. For all tables the first column determines the test dataset, and column two describes the KB generation method. Notation is as follows: *Chi* denotes the original WM/Chi's approach, *Chi EvA* denotes the variation in which ET is used to solve rules conflicts under *criterion A* (see Section 2.3), and *Chi EvB* denotes the variation in which ET is used to solve rules conflicts under *criterion B* (see Section 2.3). Columns three and four show respectively the average classification accuracy and the kappa index achieved for TS. Best results for each dataset according to the criterion of the "maximum Kappa score achieved in TS" are shown in bold. When a method achieves worst performance than the baseline reference (Chi), then it is highlighted in italics.

It can be seen from the results in Tables 2-3 that by using the proposed approach to solve conflicts under the ET framework, and for all the experiments carried out, equal or better performance is achieved under *criterion A*. Results under *criterion B* show a less clear trend, sometimes achieving the best absolute performance, and sometimes underperforming the reference baseline. By comparing the two alternative aggregation approaches (*max*, Table 2, and *sum*, Table 3) the results support here the hypothesis of the use of combination aggregation methods (see [12]), for which best performances for all datasets are achieved when using *sum*.

Table 2. Experiment results over TS using the one-winner aggregation approach. Results are shown as mean±std. Best performance on each dataset is highlighted in bold. In italics means the method performs worse than the baseline method (Chi)

Dataset	Method	Accuracy TS	Kappa TS
PIMA	Chi	0.5868±0.0471	0.1410±0.0995
	Chi EvA	0.7114±0.0563	0.3231±0.0904
	Chi EvB	0.7229±0.0627	**0.3255±0.1102**
IRIS	Chi	0.9133±0.0850	0.8588±0.1346
	Chi EvA	0.9133±0.0850	0.8588±0.1346
	Chi EvB	*0.8800±0.0752*	*0.8015±0.1127*
WINE	Chi	0.9234±0.0409	0.8778±0.0638
	Chi EvA	0.9251±0.0406	0.8803±0.0636
	Chi EvB	0.9423±0.0403	**0.9077±0.0643**
CANCER	Chi	0.9428±0.0209	0.8750±0.0433
	Chi EvA	0.9428±0.0209	0.8750±0.0433
	Chi EvB	*0.9410±0.0381*	*0.8730±0.0810*
GLASS	Chi	0.4745±0.1127	0.2785±0.1608
	Chi EvA	0.6209±0.0949	**0.4858±0.1212**
	Chi EvB	0.5887±0.1211	0.4319±0.1443

Table 3. Experiment results over TS using the normalized addition (*sum*) aggregation aggregation approach. Results are shown as mean±std. Best performance on each dataset is highlighted in bold. In italics means the method performs worse than the baseline method (Chi)

Dataset	Method	Accuracy TS	Kappa TS
PIMA	Chi	0.7137±0.0485	0.3548±0.1066
	Chi EvA	0.7375±0.0583	**0.3901±0.1034**
	Chi EvB	0.7274±0.0690	*0.3249±0.1192*
IRIS	Chi	0.9600±0.0328	0.9323±0.0559
	Chi EvA	0.9600±0.0328	0.9323±0.0559
	Chi EvB	*0.9333±0.0424*	*0.8888±0.0706*
WINE	Chi	0.9483±0.0532	0.9205±0.0803
	Chi EvA	0.9489±0.0535	0.9213±0.0807
	Chi EvB	0.9503±0.0459	**0.9222±0.0696**
CANCER	Chi	0.9535±0.0357	0.8979±0.0766
	Chi EvA	0.9535±0.0357	0.8979±0.0766
	Chi EvB	*0.9451±0.0398*	*0.8810±0.0854*
GLASS	Chi	0.5696±0.0994	0.3829±0.1428
	Chi EvA	0.6146±0.0781	0.4499±0.0925
	Chi EvB	0.6348±0.1314	**0.4961±0.1562**

5 Discussion

In this work we present a method for handling rule conflicts that appear during the application of the WM algorithm for the induction of fuzzy IF-THEN rules. Specifically we have focused on the context of classification problems, following the extension of the WM method described by Chi et al. [11], and proposed an approach by reinterpreting the conflict resolution mechanism as an evidence combination problem. For this purpose the Evidential Theory framework proposed by Dempster and Shafer has been used [13].

Although the results here exposed should still be regarded as preliminary, they are encouraging toward the usefulness of the proposed method. Two possible conflict resolution criteria (here referred as *A* and *B*) have been proposed. The results among our classification benchmark have shown equal or better performance in all the datasets under *criterion A*. Under *criterion B* we were able to achieve the best absolute performance in some datasets, however, for other datasets the method underperformed the reference baseline.

Further testing is indeed needed. Whether it is possible to determine beforehand the exact circumstances over which the approach presented would effectively contribute to achieve better classification performance than the classical WM/Chi approach still needs to be investigated. In general results might not generalize for all situations, and possibly the contribution of the conflict resolution mechanism to the overall performance might become blurred by mightier configuration choices, such as the parameterization of the inference engine, the implementation of the fuzzy logic operators, or the particularities of the input dataset.

We expect as well the dimensionality of the feature space to be a factor of consideration. At this respect, in our analyses –as well as for all the referenced approaches in the literature- equally spaced triangular fuzzy sets were used for the partition of the input space. Even though this kind of partition is not optimal, literature has shown that it provides reasonably good performance when the number of input fuzzy sets has been chosen correctly. Our choice to use here three equally spaced input fuzzy sets obeys, in one hand, to the motivation of keeping the experimentation procedures rather simple. On the other hand the number of rule conflicts is expected to increase when lowering the dimensionality of the feature input space (more examples will fall under the same fuzzy input region). As we were interested in evaluating the effects of varying the conflict resolution mechanism over the performance, keeping of a low dimensional feature input space seemed as of being a proper choice. Future work to objectively quantify the amount of rule conflict occurring on each case will be valuable to assess the actual impact of each of the conflict resolution approaches.

Further understanding of the full possibilities that Evidential Theory might bring into this context is also needed. We might be, in fact, not yet taking advantage of the full capabilities of this model. Specifically we are currently ignoring the amount of belief that is shared among all the outcomes, i.e. $m(\Omega)$, which cannot be assigned to specific focal elements. It might be useful to consider this as a relevant piece of information to some ending, such as carrying out some sort of output weighting or normalization. In addition, notice that according to the formulation presented in Section 2.3, we are in fact dealing with a simplified version of ET in which, in practice, each rule provides evidence supporting only one of the possible classes/hypothesis of *U*. In contrast, Evidential Theory could be applied to allow rules to point to several outputs at the same time. In other words, we are currently limiting to a Multiple-Input Single-Output scenario, however benefits might derive when considering Multiple-Input Multiple-Output rules of the form $IF\ x_1\ is\ A_1\ \wedge \ldots \wedge\ x_m\ is\ A_m\ THEN\ C_1 \wedge \ldots \wedge C_k$. As Dempster-Shafer theory has as a special case the probability theory, it would be interesting as well to explore the links to probabilistic conflict resolution. In particular a parallelism could be established between rules (6) and the rules of probabilistic fuzzy classifiers [20].

Last but not least, even though in this work we have focused on classification problems, the methodology here proposed can be directly applied as well to regression problems, and therefore to the original WM formulation [10]. Possible benefits of the approach into this context need still to be assessed.

Future work will be carried out in order to provide answers to the aforementioned questions.

References

1. J.-S. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man and Cybernetics,* vol. 23, no. 2, pp. 665-685, 1993.
2. W. Yu and X. Li, "Fuzzy identification using fuzzy neural networks with stable learning algorithms," *IEEE Transactions on Fuzzy Systems,* vol. 12, no. 3, pp. 411-420, 2004.
3. D. Alvarez-Estevez, J. Principe and V. Moret-Bonillo, "Information theoretic fuzzy modeling for regression," in *IEEE World Congress on Computational Intelligence*, Barcelona, Spain, 2010.
4. S. Chiu, "Fuzzy model identification based on cluster estimation," *Journal of Intelligent &*

Fuzzy Systems: Applications in Engineering and Technology, vol. 2, no. 3, pp. 267-278, 1994.

5. J. Bezdek, R. Ehrlich and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences,* vol. 10, no. 2-3, pp. 191-203, 1984.
6. D. Nauck and R. Kruse, "Obtaining interpretable fuzzy classification rules from medical data," *Artificial Intelligence in Medicine,* vol. 16, pp. 149-169, 1999.
7. S. Pal and S. Mitra, "Multilayer perceptron, fuzzy sets and classification," *IEEE Transactions on Neural Networks,* vol. 3, no. 5, pp. 683-697, 1992.
8. D. Alvarez-Estevez, J. Principe and V. Moret-Bonillo, "Neuro-fuzzy classification using the Correntropy criterion: application to sleep depth estimation," in *2010 International Conference on Artificial Intelligence*, Las Vegas, NV, 2010.
9. J. Alonso and L. Magdalena, "HILK++: an interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers," *Soft Computing,* no. 15, pp. 1959-1980, 2011.
10. L.-X. Wang and J. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 22, no. 6, pp. 1414-1427, 1992.
11. Z. Chi, H. Yan and T. Pham, Fuzzy algorithms: with applications to image processing and pattern recognition, World Scientific Publishing Co. Pte. Ltd., 1996.
12. O. Cordon, M. del Jesus and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems," *International Journal of Approximate Reasoning,* vol. 20, no. 1, pp. 21-45, 1999.
13. G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
14. L.-X. Wang, "The WM method completed: a flexible fuzzy system approach to data mining," *IEEE Transactions on Fuzzy Systems,* vol. 11, no. 6, pp. 768-782, 2003.
15. H. Ishibuchi, Y. Kaisho and Y. Nojima, "Design of linguistically interpretable fuzzy rule-based classifiers: a short review and open questions," *Journal of Multi-valued Logic and Soft Computing,* vol. 17, pp. 101-134, 2011.
16. J. Sanz, A. Fernández, H. Bustince and F. Herrera, "Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning," *Information Sciences,* vol. 180, no. 19, pp. 3674-3685, 2010.
17. M. Elkano, M. Galar, J. Sanz and H. Bustince, "CHI-BD: A fuzzy rule-based classification system for Big Data classification problems," *Fuzzy Sets and Systems,* 2017.
18. E. Shortliffe and B. Buchanan, "A model of inexact reasoning in medicine," *Mathematical Biosciences,* vol. 23, 1975.
19. M. Lichman, 2013. [Online]. Available: http://archive.ics.uci.edu/ml.
20. J. van den Berg, U. Kaymak and W. van de Bergh, "Fuzzy classification using probability-based rule weighting," in *Proceedings of the 11th IEEE International Conference on Fuzzy Systems*, Hawaii, USA, 2002.