

FUZZY RULE BASE DESIGN WITH PROBABILISTIC WEIGHTS

Leonardo G. Melo and Luís A. Lucas and Myriam R. Delgado

Graduate Program on Electrical Engineering and Computer Science

Federal University of Technology - Paraná

venomelo@yahoo.com.br, {lalucas,myriamdelg}@utfpr.edu.br

Abstract –This paper proposes a simple and innovative method to design rule bases for inference systems by joining well-known theories to treat uncertainty, such as probability and fuzzy systems. The rule base design is based on some modifications in the Wang-Mendel method in the sense that all the information obtained from the training set can be considered. The proposed method provides fuzzy rules where each consequent is determined in a probabilistic way. The resulting fuzzy system is applied in a classification problem and has its performance compared with a fuzzy classifier obtained by the original Wang-Mendel method. The results show that the design method being proposed outperforms the traditional WM method, especially when data is noisy.

Keywords – Fuzzy Rules; Probabilistic Weights, Wang-Mendel Method, Automatic Rule Base Design, Classification Problem.

1. Introduction

Since the emergence of fuzzy sets theory by [1], it was supposed to be joined with the Theory of Probability, to produce a better treatment of uncertainty. There is an extensive discussion about the theme in the literature: some authors agree with such combination [2, 3], while others limit themselves to just compare the two theories [4, 5].

Some approaches aim to develop hybrid systems for the treatment of uncertainty as in [2] and [3]. However, these are not the first results joining both theories. Other authors have also previously succeeded in developing fuzzy techniques with probability theory, such as fuzzy random variables [6, 7], which give rise to fuzzy probability distributions [8] and the central limit theorem for these types of variables; as well as the probabilistic fuzzy set described in [9].

In the context of combination or cooperation, [2] formalized the Probabilistic Fuzzy Logic and Probabilistic Fuzzy Logic Systems (PFS). In PFS both types of uncertainty are treated. The rule base is comprised of fuzzy rules whose consequents are distributed in different fuzzy sets, each one associated with a given probability measure. What happens in the system proposed by [2] is that some rules have several consequents with different probabilities of occurrence, which can be seen as a set of rules with the same antecedents and different consequents.

The PFS proposal inspired the methodology presented in this paper since we can also associate probabilistic weights to each possible consequent in the rule. Although our methodology considers that the weights associated with each possible consequent are determined based on a probabilistic measure, we can use the system in a traditional way (i.e. deterministically), considering only the consequent with the highest weight.

The structure of this paper is divided as follows: after this introduction, Section 2 presents a brief review of the classical Wang-Mendel (WM) algorithm as well as an analysis of WM method in the context of fuzzy partition of input variables. Section 3 describes the main operation steps of the proposed method for generating fuzzy rules based on probabilistic weights, detailing the inspiration in the Bayes Theorem for their calculation; besides it presents a simple example to illustrate the methodology. Section 4 is dedicated to the presentation and analysis of the results. Finally, Section 5 concludes the paper and proposes some future works.

2 WANG-MENDEL ALGORITHM

There are several approaches to automatic rule generation, which generally can be classified as evolutionary and non-evolutionary ones. The first type will not be addressed in this paper, nevertheless, an interesting overview of it is given in [10].

One of the most popular non-evolutionary algorithm for rule bases generation is the Wang-Mendel algorithm [11]. Wang-Mendel method is widely used for generating FIS rules with general scope (with applications ranging from control to classification problems).

2.1 A Brief Review of WM Method

The WM approach can be briefly described in the following. Given a training base T , consisting of $size(T)$ pairs (\mathbf{x}, y) where $\mathbf{x} = (x_1, \dots, x_n)$ is the input vector and y is the output; and considering a fuzzy partition for each fuzzy variable X_v , $v = 1, \dots, n$, where L_v is the total of linguistic terms in the fuzzy partition of variable X_v ; a rule R^t is created for each pair (\mathbf{x}, y) . The linguistic term l_v chosen for each variable X_v in the rule R^t is $l_v = \arg \max (\mu_{l_v}(x_v))$, $l_v = 1, \dots, L_v$, i.e., l_v is associated with the membership function that has the highest membership degree at point x_v . Then an initial rule

base ($RB_{complete} = \{R^1, R^2, \dots, R^t, \dots, R^{size(T)}\}$ with every rule generated) is obtained. $RB_{complete}$ may contain conflicting rules, i.e. there may exist several rules with the same antecedents and different consequents (i.e. the same premise but different conclusions). So, a final reduced rule base RB_{reduc} is created by reducing $RB_{complete}$: at most one rule is kept for each possible antecedent (premise), and its consequent (conclusion) is chosen based on the firing strength value ($FS^t = \mu_{l_1}(x_1) \mathbf{t} \mu_{l_2}(x_2) \mathbf{t} \dots \mathbf{t} \mu_{l_n}(x_n)$), where \mathbf{t} is a t-norm and $\mu_{l_v}(x_v)$ is the compatibility degree associated with the linguistic term l_v chosen for X_v .

Although the WM method can be considered one of the simplest way to generate rules from data, it can be improved to not discard some important information present in the training set. In the next section, we present an analysis of the WM method based on a fuzzy partition of the input space to support this assumption.

2.2 Analysis of Wang-Mendel Method in a Classification Context

Before presenting the proposed method, it seems necessary a more detailed analysis of claims that the method of Wang Mendel (WM) can disregard important information that may be present in the training dataset.

In WM method for classification problems for example, several points are associated with different classes, and the goal is to build a rule base through these points, considering only the information of firing strength of each generated rule. Instead of analyzing the WM method based only on the points associated with the rules and their firing strengths, here we will analyze WM based on the universes partition.

First we set the number of linguistic terms $\{L_1, L_2, \dots, L_n\}$ for each variable X_v and perform a fuzzy partition of the input space generating a total of Regions Of Interest (RoI^r) equal to $Tot_{RoI} = \prod_{v=1}^n L_v$.

Then we can rewrite WM method in the following way:

1. For a region of interest RoI^r , with $r = 1, 2, \dots, Tot_{RoI}$, resulting from the partition of the input space, create a fuzzy rule associated with every point in that region. In the rule's premise, for each input variable, select the membership function (MF) with the highest compatibility degree. So, the region of interest (RoI^r), which the point belongs to, is based on the support of the MF selected in the premise $\{A_1^r \times A_2^r \times \dots \times A_n^r\}$ of its associated rule. Moreover, it can be noticed that all the rules (points) with the same premise are in the same region of interest (see Fig. 1 (a) for an example with $n = 2$);
2. Calculate the firing strength (FS) of each rule in the region of interest (RoI^r), using an appropriate operator, e.g. t-norm (see Fig. 1 (b)). Here we will consider that the Firing Strength ($FS(\mathbf{x})$) of each rule or point \mathbf{x} can be viewed as the membership degree ($MD(\mathbf{x})$) of that point to the region of interest (the higher is such degree, the more it belongs to the region of interest).
3. Eliminate redundancies and inconsistencies in the region of interest: for each subset of rules with the same premise, only the rule with the highest firing strength (FS) remains. Another way to describe this step is: for each region of interest, choose only one point to represent it (the one with the highest membership degree (MD)).

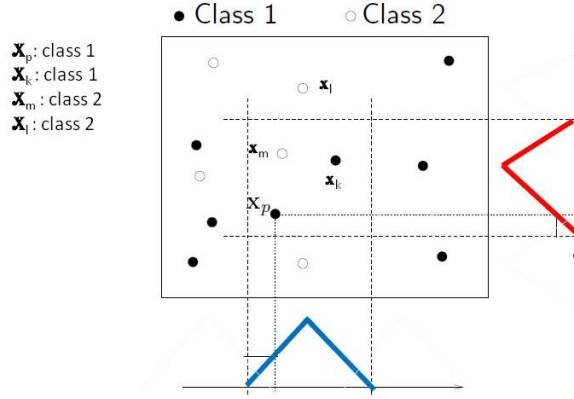
Fig. 1 (a) and Fig. 1 (b) illustrate steps 1 and 2, respectively. In Fig. 1 (a), three points \mathbf{p} , \mathbf{k} and \mathbf{m} in the region of interest will generate three fuzzy rules. In Fig. 1 (b) we can see that the FS of each rule depends on the MF selected to be its premise. Now imagine that the premises of the rules associated with \mathbf{p} and \mathbf{k} appear also in 10 other rules of that region. That is, 10 points assigned to Class 1 and used for training, activate the same membership functions in the premise and therefore are in the same region of interest. Now suppose that the premise of the rule associated with \mathbf{m} appears only in one rule in this same region of interest. That is, this point, is the unique one associated with class 2 in the same region, but it has a slightly higher membership degree (MD), or in other words, the rule associated with \mathbf{m} has a slightly higher FS. What will happen is that the rule associated with \mathbf{m} will remain in the base as the unique representant of that region of interest, but perhaps it does not correctly represent all the information available.

Based on this new analysis we can see that a region will be represented only by one point (the one with the highest MD), ignoring all the remaining information available in this region. Then, we can conclude that the WM method may ignore a lot of information in its rule-base data driven process. So, in the next section we will propose an improvement in the classical WM method to deal with such problems.

3 RB DESIGN AND PROBABILISTIC WEIGHTS

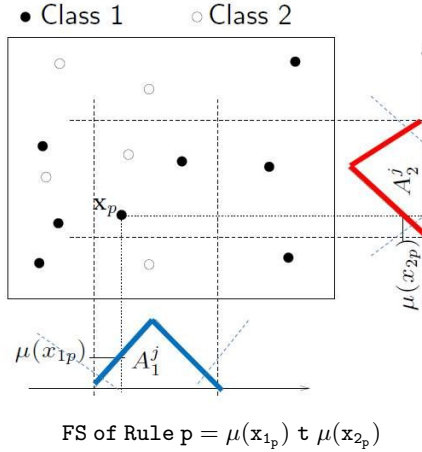
Aiming to overcome the problems discussed in the previous section, and to present a method for the automatic generation of fuzzy rules based on probabilistic weights, this section presents an alternative to the WM method, summarized in the following steps:

1. Set the number of linguistic terms and perform the fuzzy partition of the input space (based on the expert knowledge or any automatic method as for example a clustering-based method). Let C be the total of classes (or the total of linguistic terms in the consequent) for the considered application domain.
2. For every region of interest RoI^r resulting from the input space partition, associate a fuzzy rule with each point in that region as described in step 1 of Section 2.2.



Rule p : If x_1 is A_1^r and x_2 is A_2^r then class is class1
 Rule m : If x_1 is A_1^r and x_2 is A_2^r then class is class2
 Rule k : If x_1 is A_1^r and x_2 is A_2^r then class is class1

(a)WM: Step 1. Resulting rules in the RoI $A_1^r \times A_2^r$



$$\text{FS of Rule } p = \mu(x_{1p}) \tau \mu(x_{2p})$$

(b)WM: Step 2. Firing Strength (FS) of Rule p

Figure 1: Classic WM Algorithm: Steps 1 and 2

3. Calculate the firing strength $FS_i(\mathbf{x})$ for each rule (or point \mathbf{x}) associated with class K_i , as described in step 2 of Section 2.2. $FS_i(\mathbf{x})$ can be viewed as a measure of how much the point \mathbf{x} of class K_i belongs to the region of interest RoI^r : $FS_i(\mathbf{x}) = MD_i(\mathbf{x})$, where $MD_i(\mathbf{x})$ is the membership degree of a point \mathbf{x} with class K_i to the region RoI^r and can be calculated using an appropriate t-norm.
4. Calculate for every class or consequent K_i , $i = 1, \dots, C$, $N_i =$ total number of points of class K_i in RoI^r . N_i can be viewed as a measure of how many times the rule "If x_1 is A_1^r and x_2 is A_2^r then y is K_i " appears.
5. Calculate $N =$ total number of points in RoI^r . N can be viewed as a measure of how many times the premise "If x_1 is A_1^r and x_2 is A_2^r " appears.
6. Calculate $P(K_i) = N_i/N$, the initial probability of occurrence of the consequent K_i in RoI^r .
7. Define $H_{Best_i} = \max(MD_i(\mathbf{x}))$, where $Best_i = \{\mathbf{x} \mid \text{class of } \mathbf{x} \text{ is } K_i \text{ and } MD_i(\mathbf{x}) = H_{Best_i}\}$, i.e. $Best_i$ is the set of points in RoI^r with class K_i for which the membership degree is maximum or $Best_i$ is the set of best points of class K_i in RoI^r . Finally, define $Best = \bigcup Best_i$, i.e. $Best$ is the set of best points in RoI^r , independently of their class.
8. Inspired in the Bayes Theorem combine the frequency information $P(K_i)$ with the quality information H_{Best_i} to obtain the final probabilistic weight: $w_{K_i} = P(K_i) \cdot H_{Best_i} / P(Best)$.

As in the classical WM method, regions with no points (null RoIs) will not produce any rule. This is because we consider $w_{K_i} = 0$, for all i , since $P(K_i) = 0$, for $i = 1, \dots, C$. Another point is that more information is being considered since not only the quality of the points but also their frequency is taken into account to calculate w_{K_i} , i.e., to define the rule's consequent (the one with the highest w_{K_i}).

The complete step-by-step description for the proposed algorithm will be detailed as an example in Section 3.2. The final step (step 8), which used Bayes Theorem as an inspiration to define the weight associated with each consequent will be detailed in the next section.

3.1 Probabilistic Weights and the Bayes Theorem

Step 8 of the proposed algorithm, instead of using only a frequency analysis ($P(K_i)$) based on the number of points of a class in the region of interest, allows us to combine it with the information of the points' quality, measured by the membership degrees of the points to the region i.e. the membership of points to the fuzzy sets that define the region of interest. In this section, we will detail the motivation for the use of the Bayes Theorem to combine frequency and quality information in the calculus of each weight.

Generally speaking, when we have a fuzzy system's rule base obtained by an automatic generation algorithm like the one proposed in this paper, each region of interest RoI^r , which will be further represented by a fuzzy rule with multiple classes (consequents), could be understood as a sample space ($\Omega = RoI^r$), that can be partitioned in different classes (K_i). The aim here is to obtain a reduced rule base with a weight (w_{K_i}) associated with each class K_i of the rule that will be chosen to represent the region RoI^r . Then, we can use the frequency information to calculate this weight. The more points are associated with a class K_i , the higher is the weight w_{K_i} . Nevertheless, it is not interesting that we take into account only the frequency of occurrence of consequent K_i since points whose membership degrees $MD_i(x)$ are low, are not very representative of the region. So, they should have lower importance while defining the weight associated with class K_i .

Since we assume that RoI^r is the sample space, in the rest of this section, we will describe each idea behind step 8, omitting the notation RoI^r , remembering that for the computation of the final rule base, the whole process (steps 2 to 8) must be repeated for each RoI^r in the fuzzy partition considered (see Section 3.2 for more details).

The principal goal of step 8 is to calculate the probability of occurrence of consequent K_i , given that we got the best information ($Best$) in the region considered. Thus it is possible to rewrite the Bayes' theorem as:

$$P(K_i|Best) = \frac{P(K_i \cap Best)}{P(Best)}, \quad (1)$$

where:

$P(K_i|Best)$: final probability of the consequent being K_i , given that we got the points with the highest membership degree.

$P(K_i \cap Best)$: probability of getting at the same time the points associated with the consequent K_i and the ones with the highest membership degree.

$P(Best)$: Probability of getting the best points.

By using Conditional Probability Theorem, the intersection in equation 1 can be rewritten as

$$P(K_i|Best) = \frac{P(K_i) \cdot P(Best|K_i)}{P(Best)}, \quad (2)$$

where:

$P(K_i)$: initial probability of consequent K_i , calculated as described in algorithm's step 6.

$P(Best|K_i)$: probability of getting the best points given that we got points associated with the consequent K_i .

Assuming that $\Omega = RoI^r = K_1 \cap K_2 \cap \dots \cap K_C$, which is true since one point can be associated with only one class, we calculate $P(Best)$ based on the total probability as

$$P(Best) = \sum_i P(K_i) \cdot P(Best|K_i). \quad (3)$$

In this paper we considered an heuristic assumption to associate $P(Best|K_i)$ to H_{Best_i} . Since we can assume that, the higher is the membership degree of a point, the higher is the probability of it being member of the set $Best$, we considered here that $P(Best|K_i) = H_{Best_i} = \max(MD_i(x))$, where $MD_i(x)$ is the membership degree to RoI^r of a point x with consequent K_i .

We obtain therefore a normalized probability:

$$P(K_i|Best) = \frac{P(K_i) \cdot H_{Best_i}}{P(Best)}, \quad (4)$$

which combines both the membership degree that is used in the traditional method of Wang-Mendel (although there they are not separated by class) with the initial probability (calculated by the number of points associated with each class). Moreover, we conclude that in the proposed approach, the rule chosen to represent the region of interest considers information in a probabilistic and 'possibilistic' (fuzzy) way.

3.2 A Simple Example

Let $A_1^q, q = 1, \dots, L_1$, and $A_2^q, q = 1, \dots, L_2$ be the labels that define the fuzzy partition of the input space $X_1 \times X_2$. Consider three regions of interest $RoI^r, r = 1, \dots, 3$, where $RoI_1 = \text{support}(A_1^1) \times \text{support}(A_2^1)$, $RoI_2 = \text{support}(A_1^2) \times \text{support}(A_2^2)$ and $RoI_3 = \text{support}(A_1^3) \times \text{support}(A_2^3)$. Let the set of rules and their respective firing strengths (or membership degrees MD) obtained by the proposed algorithm (step 2 and step 3), as shown in Table 1.

Table 1: Complete Rule Base for the example

RoI^r	If x_1 is	and x_2 is	then Y is	$MD_i(\mathbf{x})$
1	A_1^1	A_2^1	1	0.8
1	A_1^1	A_2^1	1	0.9
1	A_1^1	A_2^2	2	0.7
2	A_1^2	A_2^2	2	0.9
2	A_1^2	A_2^2	2	0.6
2	A_1^3	A_2^3	1	0.3
3	A_1^3	A_2^3	3	0.5
3	A_1^3	A_2^3	3	0.4

Then we use steps 4, 5, and 6 to compute the initial probabilities ($P(K_i)$) for each class in each region of interest. Thereafter we aggregate redundant rules representing them by only one rule with the maximum membership degree associated (H_{Best_i}), according to step 7. Then, the rule base can be reduced to the form of Table 2.

Table 2: The reduced rule base for the example

RoI^r	If x_1 is	and x_2 is	then Y is	$P(K_i)$	H_{Best_i}
1	A_1^1	A_2^1	1	2/3	0.9
1	A_1^1	A_2^1	2	1/3	0.7
2	A_1^2	A_2^2	2	2/3	0.9
2	A_1^2	A_2^2	1	1/3	0.3
3	A_1^3	A_2^3	3	1	0.5

Now, we calculate the total probabilities $P(Best)$ for each region $RoI^r, r = 1, \dots, 3$, according to equation 3.

$$RoI_1 : P(Best) = 2/3 \cdot 0.9 + 1/3 \cdot 0.7 = 0.83;$$

$$RoI_2 : P(Best) = 2/3 \cdot 0.9 + 1/3 \cdot 0.3 = 0.70;$$

$$RoI_3 : P(Best) = 1 \cdot 0.5 = 0.50.$$

Finally, in step 8, the probabilistic weights are calculated according to equation 4. The final reduced rule base with the possible consequents is given in Table 3.

Table 3: The rule base (with all possible consequents) for the example

If x_1 is	and x_2 is	then Y is	w_{K_i}
A_1^1	A_2^1	1	$(2/3 \cdot 0.9)/0.83$
A_1^1	A_2^1	2	$(1/3 \cdot 0.7)/0.83$
A_1^2	A_2^2	2	$(2/3 \cdot 0.9)/0.70$
A_1^2	A_2^2	1	$(1/3 \cdot 0.3)/0.70$
A_1^3	A_2^3	3	$(1 \cdot 0.5)/0.5$

The most simple way to use the proposed approach is a deterministic way: a classical FIS where the consequent considered in each rule is the one with the highest weight.

The main question that arises when we use the data driven rule design proposed could be: if we discard the probability information (i.e. if we do not consider the conflicts in the training data) what is the contributions of the proposed approach? We can emphasize that the weights computation alone represents a contribution, since unlike other approaches it considers both types of information (quantitative and qualitative) and even if we do not use the system in a stochastic way, the contributions emerge as we will show in the experiments described in the next Section.

Table 4: The final rule base obtained

If x_1 is	and x_2 is	then Y is	w_{K_i}
A_1^1	A_2^1	1	$w_{K_1} = 0.7189$
A_1^2	A_2^2	2	$w_{K_2} = 0.8571$
A_1^3	A_2^3	3	$w_{K_3} = 1$

4 EXPERIMENTS AND RESULTS

The experiments conducted concern the application of the method to a classification problem considering different levels of separability among classes. We intend to compare efficiencies of the WM algorithm, and the probabilistic method proposed.

In order to determine how different the methods behave when we decrease the “separability” in data, two instances of classification problems were produced with different means (μ) and variances (σ) in the Gaussian used to generate the training and the testing points. Figures 2 and 3 illustrate the points generated for instances a and b with μ for classes (1,2,3) = (10,30,50) and and μ for classes (1,2,3)= (10,20,30), both with $\sigma=9$.

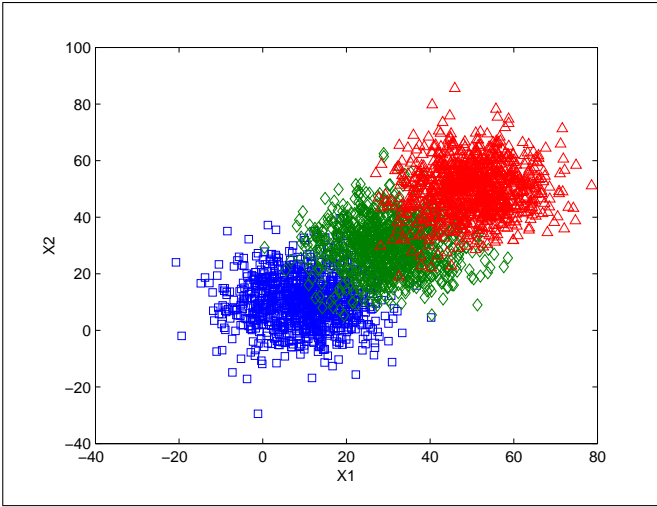


Figure 2: Points Distribution for the Instance a

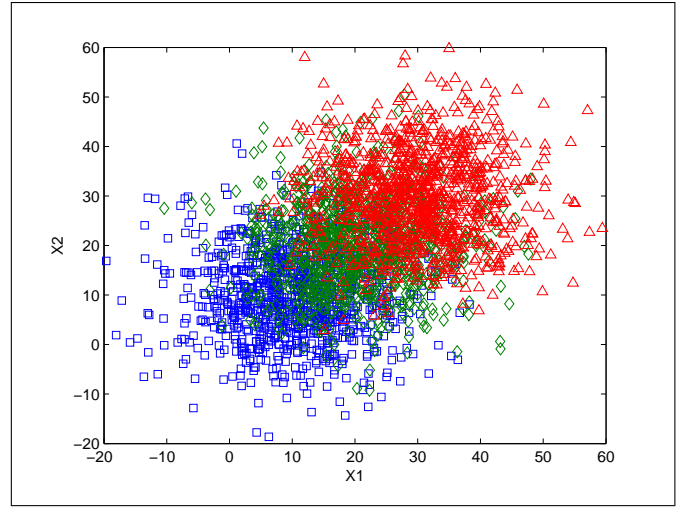


Figure 3: Points Distribution for the Instance b

Since the goal of this study does not include the automatic partition of the universe, the functions used to generate the points were the same used in the partitions (for all the methods being compared).

We used a total 1000 points for each class, dividing them into five groups of 200 points each. Then, we proceeded with cross-validation (more specifically 5-fold cross validation [12]). This way, the last four groups, totaling 800 points were used for training the algorithms, while the initial 200 points group was used for testing. Then another group was used for testing and the other for training and so on, until all groups have had the chance to participate in the testing phase. In the training phase, the rule base was obtained by each specific method in each fold. Table 5 illustrates the rule base obtained by both methods (WM and Probabilistic) for the instances a and b, fold 1. The results, in terms of performance, are presented in Table 6 considering the means obtained from five folds of the cross-validation procedure. As a criterion for comparison we used the percentage of correct outputs

$$PCO = \frac{PC}{PT} \cdot 100, \quad (5)$$

where PC = total number of test points correctly classified and PT = total test points.

After obtaining the fuzzy rules by each method, these rules were implemented as a classifier. In the case of the proposed method, since we have a probabilistic weight associated with each possible label, the rule consequent was defined as the label (or fuzzy set) associated with the highest probabilistic weight (the ones emphasized in Table 5). Then we use the conventional inference for fuzzy classifiers. After each rule infers its class output associated with its activation degree, these labels are aggregated by the maximum, ie, the class indicated on system output will be the one whose rule has the highest activation degree. Since there is no defuzzification in fuzzy classifiers, the output classification is the label obtained by the suggested system. In the test phase, the actual values were compared with those estimated by the systems being compared.

As already mentioned, the percentages of correct answers in each experiment are summarized in Table 6. We can use the standard deviation indicated in Table 6 to construct the limits of the confidence interval (95%), so that in the first instance, we have (80.4689%, 81.7311%) for the WM method, and (82.5269%, 84.2731%) for the proposed probabilistic method. So, we can conclude that the proposed method outperformed WM (with 95% of confidence).

For the second instance, there is higher difference between results of the two methods. After the confidence intervals were constructed, (56.5831%, 59.3503%) for WM and (65.7845%, 68.2155%) for Probabilistic, it is also possible to affirm with 95%

Table 5: Rule Base obtained by the WM and Probabilistic Methods: Instances a and b, Fold1

If X1 is	and X2 is	then Class is	Probabilistic		WM	
			w_{K_i}		class	
			Instance a	Instance b	Instance a	Instance b
1	1	1	0.9764	0.8462	1	1
1	1	2	0.0236	0.1458		
1	1	3	0	0.0080		
1	2	1	0.5364	0.5937	1	1
1	2	2	0.4636	0.3694		
1	2	3	0	0.0368		
1	3	1	0	0.1979		
1	3	2	1.0	0.5631	2	2
1	3	3	0	0.2390		
2	1	1	0.5411	0.6131		1
2	1	2	0.4589	0.3484	2	
2	1	3	0	0.0385		
2	2	1	0.0317	0.1924		1
2	2	2	0.9286	0.6626	2	
2	2	3	0.0397	0.1450		
2	3	1	0	0.0295		
2	3	2	0.4643	0.3750	2	
2	3	3	0.5357	0.5955		3
3	1	1	0	0.2090		
3	1	2	1.0	0.5973	2	2
3	1	3	0	0.1937		
3	2	1	0	0.0367		
3	2	2	0.4710	0.3446	2	
3	2	3	0.5290	0.6187		3
3	3	1	0	0.0019		
3	3	2	0.0294	0.1791		
3	3	3	0.9706	0.8190	3	3

Table 6: Correct Classification Percentage (P_{CO})

Instance	Approach	Fold ₁	Fold ₂	Fold ₃	Fold ₄	Fold ₅	Average	Std. dev.
a	Prob	83.5000%	84.1667%	82.3333%	83.8333%	83.1667%	83.40001%	0.7032%
a	WM	81.3333%	80.8333%	80.5000%	81.8333%	81.0000%	81.1000%	0.5083%
b	Prob	65.5000%	66.6667%	68.0000%	67.1667%	67.6667%	67.0000%	0.9789%
b	WM	58.0000%	56.1667%	58.0000%	58.5000%	59.1667%	57.9667%	1.1143%

of confidence that, although all the results deteriorated when compared with the first experiment, the probabilistic method outperformed the Wang-Mendel based approach. The reasons why the results deteriorated in the second instance can be explained by the new input partition space since the antecedents have higher intersection levels giving rise to activation of more fuzzy rules than in the first case.

Results show that the proposed method seems to contribute less when the classes are more separated (easy defined). However, in the presence of noise, the use of probabilistic information in the proposed approach seems to turn it less influenced by the noise, since the proposed method :

- achieved better results than the Wang-Mendel (with less deteriorated results) according to the results shown Table 6;
- provided the same rules in both instances, according to the results shown in Table 5.

5 CONCLUSIONS

The paper presented an analysis of Wang- Mendel method based in the input fuzzy partition. In order to overcome the detected problem in WM of loosing information in the rule base design based on a training data set, we proposed a new approach to generate fuzzy rules, by means of probabilistic weights. The proposed method was presented (including a simple example) with a special attention to its relation with the Bayes Theorem.

The effectiveness of the new method was tested in a classification problem where the proposed algorithm was compared with the classic WM. Apparently the new method, while presenting itself better, differed little from the WM algorithm, when the

experiment considered that classes to be identified are reasonably distinct from each other. However, in situations where the classes are less distinct, the probabilistic method has achieved a gain of almost 16% over the comparison algorithm.

In the future we intend to test the proposed approach in real problems to evaluate its effectiveness when compared with fuzzy and non-fuzzy based approaches.

Acknowledgments.

Myriam Regattieri Delgado acknowledges the CNPq grant 307735/2008-7, Fundação Araucária project 400/09 - 10705 for the partial financial support.

REFERENCES

- [1] L. A. Zadeh. "Fuzzy Sets". *Inform. and Control* 8, pp. 338–353, 1965.
- [2] A. H. Meghdadi and M.-R. Akbarzadeh-T. "Probabilistic Fuzzy Logic and Probabilistic Fuzzy Systems". *10th IEEE International Conference on Fuzzy Systems*, vol. 3, pp. 1127–1130, 2001.
- [3] G. Coletta and R. Scozzafava. "Conditional probability and fuzzy information". *Computational Statistics & Data Analysis* 51, pp. 115–132, 2006.
- [4] M. Laviolette and J. W. S. Jr. "The efficacy of fuzzy representations of uncertainty". *IEEE Transactions on Fuzzy Systems*, vol. 2, 1994.
- [5] N. Wilson. "Vagueness and Bayesian Probability". *IEEE Transactions on Fuzzy Systems*, vol. 2, pp. 34–36, 1994.
- [6] H. Kwakernaak. "Fuzzy Random Variables-I. Definitions and Theorems." *Inf. Sci.* 15, pp. 1–29, 1978.
- [7] H. Kwakernaak. "Fuzzy Random Variables II. Algorithms and examples for the discrete case." *Inf. Sci.* 17, vol. 3, pp. 253–278, 1979.
- [8] M. S. Taylor and S. B. Boswell. "A central limit theorem for fuzzy random variables". *Fuzzy Sets and Systems*, vol. 24, pp. 331–344, 1987.
- [9] K. Hirota. "Concepts of probabilistic set". *Fuzzy sets and systems*, vol. 5, pp. 31–46, 1981.
- [10] F. Herrera. "Genetic Fuzzy Systems: Taxonomy, Current Research Trends and Prospects." *Evolutionary Intelligence*, vol. 1, pp. 27–46, 2008.
- [11] L. X. Wang and J. M. Mendel. "Generating Fuzzy Rules by Learning from Examples". *IEEE Trans. Systems Man and Cybernetics*, vol. 22, pp. 1414–1427, 1992.
- [12] S. M. WEISS and C. KULIKOWSKI. *Computer Systems That Learn*. Morgan Kaufmann, 1991.