Erschienen in: Advances in intelligent data analysis : third International Symposium, IDA-99, Amsterdam, The Netherlands, August 9 - 11, 1999; proceedings / David J. Hand ... (eds.). - Berlin [u.a.] : Springer, 1999. - S. 87-98. - (Lecture notes in computer science ; 1642). - ISBN 3-540-66332-0

Discriminative Power of Input Features in a Fuzzy Model

Rosaria Silipo[1] and Michael R. Berthold[2]

[1] International Computer Science Institute (ICSI)
1947 Center Street, Suite 600, Berkeley, CA 94704, USA
rosaria@icsi.berkeley.edu
[2] Berkeley Initiative in Soft Computing (BISC)
Dept. of EECS, CS Division, 329 Soda Hall
University of California, Berkeley, CA 94720, USA
berthold@cs.berkeley.edu

Abstract. In many modern data analysis scenarios the first and most urgent task consists of reducing the redundancy in high dimensional input spaces. A method is presented that quantifies the discriminative power of the input features in a fuzzy model. A possibilistic information measure of the model is defined on the basis of the available fuzzy rules and the resulting possibilistic information gain, associated with the use of a given input dimension, characterizes the input feature's discriminative power. Due to the low computational expenses derived from the use of a fuzzy model, the proposed possibilistic information gain generates a simple and efficient algorithm for the reduction of the input dimensionality, even for high dimensional cases. As real-world example, the most informative electrocardiographic measures are detected for an arrhythmia classification problem.

1 Introduction

In the last years it has become more and more common to collect and store large amounts of data from different sources [1]. However a massive recording of system's monitoring variables does not grant a better performance of further analysis procedures, if no new information is introduced in the input space. In addition the analysis procedure itself becomes more complicated for high dimensional input spaces and insights about the system's underlying structure more difficult to achieve.

An evaluation of the effectiveness of every input feature in describing the underlying system can supply new information and simplify further analysis. The detection of the most informative input features, that is the features characterizing at best the underlying system, reduces time and computational expenses of any further analysis and makes easier the detection of crucial parameters for data analysis and/or system modeling.

A quite common approach for the evaluation of the effectiveness of the input features defines some feature merit measures, on the basis of a statistical model

of the system [2, 1]. Assuming that a large database is available, the probability estimations, involved in the definition of the feature merit measures, are performed by means of the events frequencies, which require a precise definition of the input parameters and a clear identification of the output classes. In many real world applications, however, estimated frequencies are unavoidably alterated by doubtful members of the output classes and by an inaccurate description of the input parameters. In addition the estimation of a probabilistic model is computationally expensive for high dimensional input spaces.

The concept of fuzzy sets was introduced in [3] with the purpose of a more efficient, though less detailed, description of real world events, allowing an appropriate amount of uncertainty. Fuzzy set theory yields also the advantage of a number of simple and computationally inexpensive methods to model a given training set. Based on the fuzzy set theory, some measures of fuzzy entropy have been established [4, 5] as measures of the degree of fuzziness of the model with respect to the training data. All the defined measures involve the data points into the fuzzy entropy calculation, in order to represent the uncertainty of the model in describing the training data.

In this paper an analysis "a posteriori" of fuzzy systems is proposed, to evaluate the discriminative power of the input features in characterizing the underlying system. A measure of possibilistic information is defined only on the basis of fuzzy rules. The separability of the different membership functions is measured on every input dimension and the input dimension with highest separability defines the most discriminative input feature, at least according to the analyzed fuzzy model. All that is based on the hypothesis that the fuzzy model describes with sufficient accuracy the data of the training set, that is that a sufficiently general training set has been used for the fuzzy rules inference. The main advantage of analyzing fuzzy rules, instead of fuzzy rules and training data as in [4, 5], consists of the highly reduced computational costs for the same amount of information, provided that the fuzzy model faithfully describes the underlying data structure.

The detection and ranking of the most effective input variables for a given task could represent one of the first steps in any data analysis process. The implementation of a fuzzy model requires generally a short amount of time even in case of very high dimensional input spaces and so does the corresponding evaluation of the discriminative power of the input features. Whenever a more accurate system's representation is wished, the analysis can continue with the application of more sophisticated and more computationally expensive analysis techniques on the most effective input features, pre-screened on the basis of the proposed possibilistic information.

2 Possibilistic Feature Merit Measures

2.1 A Possibility Measure

Given a number m of output classes C_i, $i = 1, \ldots, m$, and an n-dimensional input space, numerous algorithms exist, which derive a set of N_R fuzzy rules [3]

$\{R_k\}$, $k = 1, \ldots, N_R$, mapping the n-dimensional input into the m-dimensional output space. This set of rules models the relationships between the input data $\boldsymbol{x} \in \mathcal{R}^n$ and the output classes C_i. Each input pattern $\boldsymbol{x} = [x_1, \ldots, x_n]^T$ is associated to each output class C_i by means of a membership value $\mu_{C_i}(\boldsymbol{x})$. In figure 1.a an example is reported with a two-dimensional input space $\{x_1, x_2\}$, two output classes C_1 and C_2, and with trapezoids as membership functions $\mu_{C_1}(\boldsymbol{x})$ and $\mu_{C_2}(\boldsymbol{x})$ describing the relationships between the input data and the two output classes.

The membership function $\mu_{C_i}(\boldsymbol{x})$ quantifies the degree of membership of input pattern $\boldsymbol{x}$ to output class C_i. Its volume $V(C_i)$, as defined in eq. 1, therefore represents a measure of the possibility of output class C_i, on the basis of the given input space $D \subset \mathcal{R}^n$. Considering normalized membership functions $\mu_{C_i}(\boldsymbol{x})$, a larger volume $V(C_i)$ indicates a class of the output space with higher degree of possibility. An output class represented by a membership function, which takes value +1 everywhere on the input space, is always possible. A membership function with volume $V(C_i) = 0$ indicates an impossible class.

$$V(C_i) = \int_0^1 \int_{\boldsymbol{x} \in D} \mu_{C_i}(\boldsymbol{x}) \; d\boldsymbol{x} \; d\mu \tag{1}$$

The overall possibility of the whole output space $C = \{C_1, C_2, \ldots, C_m\}$ can be defined through the available fuzzy mapping system $\{R_k\} = \{R_1, R_2, \ldots, R_{N_R}\}$ as the sum of all the class possibilities $V(C_i)$, $i = 1, \ldots, m$. The relative contribution $v(C_i)$ of output class C_i to the whole output space's possibility is given in eq. 2.

$$v(C_i) \;=\; \frac{V(C_i)}{\sum_{j=1}^{m} V(C_j)} \tag{2}$$

In case the output class C_i is described by $Q_i > 1$ fuzzy rules, the possibility of class C_i is given by the possibility of the union of these $q = 1, \ldots, Q_i$ fuzzy subsets of class C_i, each with membership functions $\mu^q_{C_i}(\boldsymbol{x})$. The possibility of the union of membership functions can be expressed as the sum of their possibilities, taking care of including the intersection possibility only once (eq. 3). If trapezoids are adopted as membership functions, the possibility of each fuzzy rule $V_q(C_i)$ becomes particularly simple to calculate [6].

$$\begin{aligned} V(C_i) = V\left(\cup_{q=1}^{Q_i} V_q(C_i)\right) \;&=\; \int_0^1 \int_{\boldsymbol{x} \in D} \cup_{q=1}^{Q_i} \mu^q_{C_i}(\boldsymbol{x}) \; d\boldsymbol{x} \; d\mu \;= \\ = \sum_{q=1}^{Q_i} \left[V_q(C_i) - \sum_{h=q+1}^{Q_i} V_q(C_i) \cap V_h(C_i) \right] \end{aligned} \tag{3}$$

2.2 A Possibilistic Information Measure

The variable $v(C_i)$ quantifies the possibility of class C_i relatively to the possibility of the whole output space and according to the fuzzy rules used to model

the input-output relationships. $v(C_i)$, as defined in eq. 2, can then be adopted as the basic unit to measure the possibilistic information associated with class C_i. With respect to a probabilistic model, the employment of the relative possibility of class C_i, $v(C_i)$, takes into account the possible occurrence of multiple classes for any input pattern $\boldsymbol{x}$ and the calculation of the relative volume $v(C_i)$ is generally easier than the estimation of a probability function.

As in the traditional information theory, the goal is to produce a possibilistic information measure, that is [1]:

1. at its maximum if all the output classes are equally possible, i. e. $v(C_i) = \frac{1}{m}$ for $i = 1, \ldots, m$, m being the number of output classes;
2. at its minimum if only one output class C_i is possible, i. e. in case $v(C_j) = 0$ for $j \neq i$;
3. a symmetric function of its arguments, because the dominance of one class over the others must produce the same amount of possibilistic information, independently of which the favorite class is.

In order to produce a measure of the global possibilistic information $I(C)$ of the output space $C = \{C_1, \ldots, C_m\}$, the traditional functions employed in information theory – as the entropy function $I_H(C)$ (eq. 4) and the Gini function $I_G(C)$ (eq. 5) [1, 2] – can then be applied to the relative possibilities $v(C_i)$ of the output classes.

$$I_H(C) \;=\; -\sum_{i=1}^{m} v(C_i)\ \log_2\left(v(C_i)\right) \tag{4}$$

$$I_G(C) \;=\; 1 \;-\; \sum_{i=1}^{m} \left(v(C_i)\right)^2 \tag{5}$$

In both cases, entropy and Gini function, $I(C)$ represents the amount of possibilistic information intrinsically available in the fuzzy model. In particular not all the input features are effective the same way in extracting and representing the information available in the training set through the fuzzy model. The goal of this paper is to make explicit which dimension of the input space is the most effective in recovering the intrinsic possibilistic information $I(C)$ of the fuzzy model.

2.3 The Information Gain

Given a fuzzy description of the input space $\{R_k\}$ with intrinsic possibilistic information $I(C)$, a feature merit measure must describe the information gain derived by the employment of any input feature x_j in the model. Such information gain is expressed as the relative difference between the intrinsic information of the system before, $I(C)$, and after using that variable x_j for the analysis, $I(C|x_j)$, (eq. 6). The x_j input features producing the highest information gains

are the most effective in the adopted model to describe the input space, and therefore the most informative for the proposed analysis.

$$g(C|x_j) = \frac{I(C) - I(C|x_j)}{I(C)} \tag{6}$$

Let us suppose that the input variable x_j is related to the output classes by means of a number N_R of membership functions $\mu^q_{C_i}(x_j)$, with $q = 1, \ldots, Q_i$ membership functions for every output class C_i, for $i = 1, \ldots, m$ output classes, and $N_R = \sum_{i=1}^{m} Q_i$. The use of input variable x_j for the final classification consists of the definition of an appropriate set of thresholds along input dimension j, that allow the best separation of the different output classes. A set of cuts is then created on the j-th input dimension, to separate the $F \leq N_R$ contiguous trapezoids related to different output classes.

If trapezoids are adopted as membership functions of the fuzzy model, the optimal cut between two contiguous trapezoids is located at the side intersection, if the trapezoids overlap on the sides; at the middle point of the overlapping flat regions, if the trapezoids overlap in their flat regions; at the middle point between the two trapezoids, if they do not overlap.

Between two consecutive cuts, a linguistic value L_k $(k = 1, \ldots, F)$ can be defined for parameter x_j. Considering $x_j = L_k$ corresponds to isolating one stripe c_k on the input space. In stripe c_k new membership functions $\mu^q(C_i|x_j = L_k)$ to the output classes C_i are derived as the intersections of the original membership functions $\mu^q_{C_i}(\boldsymbol{x})$ with the segment $x_j = L_k$. Each stripe c_k is characterized by a local possibilistic information $I(c_k) = I(C|x_j = L_k)$ (eq. 4 or 5). The average possibilistic information $I(C|x_j)$, derived by the use of variable x_j in the fuzzy model, corresponds to the averaged sum of the local possibilistic information of stripes c_k (eq. 7).

$$I(C|x_j) = \frac{1}{F} \sum_{k=1}^{F} I(C|x_j = L_k) \tag{7}$$

The less effective the input feature x_j is in the original set of fuzzy rules, the closer the remaining $I(C|x_j)$ is to the original possibilistic information $I(C)$ of the model and the lower the corresponding information gain is, as described in eq. 6. Every parameter x_j produces an information gain $g(C|x_j)$ expressing its effectiveness in performing the required classification on the basis of the given fuzzy model. The proposed information gain can be adopted as a possibilistic feature merit measure.

2.4 An Example

In figure 1 an example is shown for a two-dimensional input space, two output classes, and with trapezoids as membership functions. The corresponding intrinsic possibilistic information of the original model $I(C)$ is reported in table 1. The average information of the system, $I(C|x_1)$ and $I(C|x_2)$, respectively after dimension x_1 and x_2 have been used for the classification, are reported in table 2 together with the corresponding information gains $g(C|x_1)$ and $g(C|x_2)$.

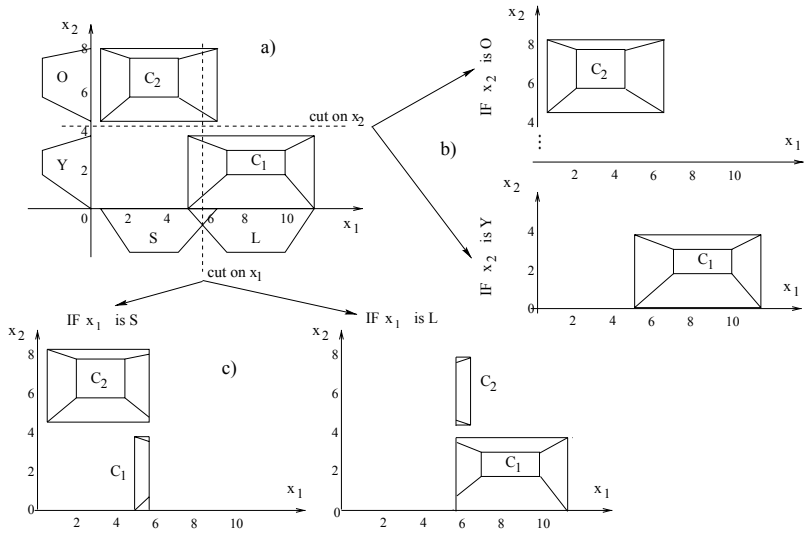

Fig. 1. New data spaces cutting on variable b) x_2 and c) x_1

A cut between the two membership functions on dimension x_2 (Fig. 1.b) produces a better separation than a cut on dimension x_1 (Fig. 1.c). That is the analysis on dimension x_2 offers a higher gain in information than the analysis on dimension x_1. This is indicated by $g(C|x_1) < g(C|x_2)$ either considering $I()$ as the entropy or the Gini function (Tab. 2). From the comparison of the information gains, $g(C|x_1)$ and $g(C|x_2)$, the analysis on variable x_2 supplies more of the information available in the fuzzy model than the analysis carried on variable x_1. The same conclusion could have been reached using $I(C|x_1) > I(C|x_2)$, but an information description through the gain function produces more clear results than using directly the possibilistic information parameter $I(C|x_j)$.

3 Real World Applications

The results in the previous section show the efficiency of the proposed possibilistic feature merit measures in detecting the input dimensions with maximum information content. In this section some experiments on real world databases are

Table 1. The fuzzy information measures for the two dimensional example

C_1	C_2	$I_H(C)$	$I_G(C)$
$V(C_1) = 13.0$	$V(C_2) = 12.6$	0.99	0.49
$v(C_1) = 0.51$	$v(C_2) = 0.49$		

Table 2. $I(C|x_j)$ and $g(C|x_j)$

$x_1 = S$	$x_1 = L$	$x_2 = Y$	$x_2 = O$
$V(C_1\|x_1) = 0.53$	$V(C_1\|x_1) = 13.0$	$V(C_1\|x_2) = 13.0$	$V(C_1\|x_2) = 0.00$
$V(C_2\|x_1) = 12.6$	$V(C_2\|x_1) = 0.53$	$V(C_2\|x_2) = 0.00$	$V(C_2\|x_2) = 12.6$
$v(C_1\|x_1) = 0.04$	$v(C_1\|x_1) = 0.96$	$v(C_1\|x_2) = 1.0$	$v(C_1\|x_2) = 0.00$
$v(C_2\|x_1) = 0.96$	$v(C_2\|x_1) = 0.04$	$v(C_2\|x_2) = 0.00$	$v(C_2\|x_2) = 1.0$
$I_H(C\|x_1) = 0.24$		$I_H(C\|x_2) = 0.00$	
$I_G(C\|x_1) = 0.07$		$I_G(C\|x_2) = 0.00$	
$g_H(C\|x_1) = 0.76$		$g_H(C\|x_2) = 1.0$	
$g_G(C\|x_1) = 0.84$		$g_G(C\|x_2) = 1.0$	

performed and the corresponding results reported, in order to observe whether these possibilistic feature merit measures are actually capable to detect the database features which controls the maximum information even on real-world data.

3.1 The IRIS Database

The first experiment is performed on the IRIS database. This is a relatively small database, containing data for three classes of iris plants. The first class is supposed to be linearly separable and the last two classes non linearly separable. The plants are characterized in terms of: 1) sepal length 2) sepal width 3) petal length and 4) petal width.

Both possibilistic information gains are very high for the third and the fourth input parameter, and almost zero for the first two input features (Tab. 3). In [8], where a detailed description of the parameters adopted in the IRIS database is produced, the sepal length and sepal width – parameter 1 and 2 – are reported to be more or less the same for all the three output classes, i. e. uninformative. Thus input parameter 1 and 2 should not contribute to the correct discrimination of the output classes. On the opposite, the petal features – parameters 3 and 4 – characterize very well the first class of iris (iris setosa) with respect to the other two.

In this case, the proposed possibilistic feature merit measures produce a very reliable description of the informative power of every input parameter. Hence parameters 1 and 2 could be removed and the analysis performed solely on the basis of parameters 3 and 4 without a relevant loss of information. The class correlation, reported in [9], is also very high for parameters 3 and 4 and much lower for the first two parameters. That confirms the results from the possibilistic feature merit measures.

3.2 Arrhythmia Classification

A very suitable area for fuzzy – or more generally imprecise – decision systems consists of medical applications. Medical reasoning is quite often a qualitative

Table 3. Information gain $g(C)$ of the iris features in the IRIS database

$I(C)$		x_1	x_2	x_3	x_4
$I_H(C) = 1.44$	$g_H(C)$	0.10	0.06	0.82	0.81
$I_G(C) = 0.61$	$g_G(C)$	0.10	0.06	0.84	0.79

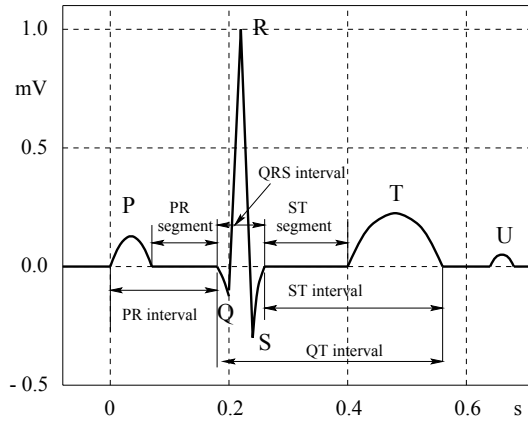

Fig. 2. The ECG waveshape.

and approximative process, so that the definition of precise diagnostic classes with crisp membership functions can sometimes lead to inappropriate conclusions. One of the most investigated fields in medical reasoning is the automatic analysis of the electrocardiogram (ECG), and inside that the detection of arrhythmic heart beats.

Some cells (the sino-atrial node) in the upper chambers (the atria) of the cardiac muscle (the myocardium) spontaneously and periodically change their electrical polarization, which progressively extends to the whole myocardium. This periodic and progressive electric depolarization of the myocardium is recorded as small potential differences between two different locations of the human body or with respect to a reference electrode. An almost periodic signal, the ECG, that describes the electrical activity of the myocardium in time, is the result. Each time period consists of a basic waveshape, whose waves are marked with the alphabet letters P, Q, R, S, T, and U (Fig. 2). The P wave describes the depolarization process of the two upper myocardium chambers, the atria; the QRS complex all together the depolarization of the two lower myocardium chambers, the ventricula; and the T wave the repolarization process at the end of each cycle. The U wave is often absent from the beat waveshape and, however, its origin is controversial. The heart contraction follows the myocardium depolarization phase. Anomalies in the PQRST waveshape are often connected to misconductions of the electrical impulse on the myocardium.

Table 4. Set of measures characterizing each beat waveshape.

RR	RR interval (ms)
RRa	average of the previous 10 RR intervals
QRSw	QRS width (ms)
VR	Iso-electric level (μV)
pA	Positive amplitude of the QRS (μV)
nA	Negative amplitude of the QRS (μV)
pQRS	Positive area of the QRS (μV * ms)
nQRS	Negative area of the QRS (μV * ms)
pT	positive area of the T wave (μV * ms)
nT	negative area of the T wave (μV * ms)
ST	ST segment level (μV)
STsl	slope of the ST segment (μV/ms)
P	P exist (yes 0.5, no -0.5)
PR	PR interval (ms)

A big family of cardiac electrical misfunctions consists of arrhythmic heart beats, deriving from an anomalous (ectopic) origin of the depolarization wavefront in the myocardium. If the depolarization does not originate in the sino-atrial node, a different path is followed by the depolarizing wavefront and therefore a different waveshape appears in the ECG signal. Arrhythmia are believed to occur randomly in time and the most common types have an anomalous origin in the atria (SupraVentricular Premature Beats, SVPB) or in the ventricula (Ventricular Premature Beats, VPB). With the development of automatic systems for the detection of QRS complexes and the extraction of quantitative measurements, large sets of data can be generated from hours of ECG signal. A larger number of measures though does not guarantee better performances of the upcoming classifier, if no significant new information is added. A pre-screening of the most significant measures for the analysis has the double advantage of lowering the input dimension and of improving the classifier's performance when poor quality measures are discarded.

The MIT-BIH database [10] represents a standard in the evaluation of methods for the automatic classification of the ECG signal, because of the wide set of examples of arrhythmic events provided. The MIT-BIH ECG records are two-channel, 30 minutes long and sampled at 360 samples/s. Two records (200 and 233) from the MIT-BIH database are analyzed in this study, because of their high number of arrhythmic beats. QRS complexes are detected and for each beat waveshape a set of 14 measures [11] is extracted by using the first of the two channels in the ECG record (Tab. 4). The first 2/3 of the beats of each record are used as training set and the last 1/3 as test set. A two-class, normal (N) vs. ventricular premature beats (VPB) is considered for record 200 and a three-class problem (N, VPB, and SVPB) for record 233, in order to quantify the discriminative power of the input features for both classification tasks.

Table 5. Information gain for different ECG beat measures (record 200). The amounts of correctly classified N and VPB and of uncertain beats are expressed in %.

	RR	RRa	QRSw	VR	pA	nA	pQRS	nQRS	pT	nT	ST	STsl	P	PR	N	VPB	unc.
g_H	.40	.00	**.78**	.09	.07	.00	.08	.04	.00	.61	.57	.01	.00	.00	99	97	1
g_G	.42	.01	**.80**	.11	.09	.01	.10	.05	.00	.63	.59	.02	.00	.00			
g_H	**.47**	-	.42	.14	.25	.25	.38	.21	-	.15	.36	.17	-	-	99	96	1
g_G	**.53**	-	.44	.18	.25	.27	.43	.26	-	.16	.38	.17	-	-			
g_H	**.74**	-	.08	-	.44	.42	.49	.28	-	-	.03	-	-	-	100	97	1
g_G	**.81**	-	.09	-	.48	.43	.55	.31	-	-	.04	-	-	-			
g_H	.52	-	-	-	.59	**.78**	.71	-	-	-	-	-	-	-	100	97	1
g_G	.56	-	-	-	.60	**.78**	.74	-	-	-	-	-	-	-			
g_H	-	-	-	-	.56	.38	**.59**	-	-	-	-	-	-	-	100	97	1
g_G	-	-	-	-	.57	.41	**.61**	-	-	-	-	-	-	-			
g_H	-	-	-	-	.41	-	**.72**	-	-	-	-	-	-	-	98	95	1
g_G	-	-	-	-	.45	-	**.76**	-	-	-	-	-	-	-			
g_H	-	-	-	-	.44	**.55**	-	-	-	-	-	-	-	-	100	97	1
g_G	-	-	-	-	.46	**.57**	-	-	-	-	-	-	-	-			
g_H	-	-	-	-	-	-	.29	-	-	-	-	-	-	-	74	50	0
g_G	-	-	-	-	-	-	.33	-	-	-	-	-	-	-			
g_H	-	-	.32	-	-	-	-	-	-	-	-	-	-	-	56	56	0
g_G	-	-	.38	-	-	-	-	-	-	-	-	-	-	-			
g_H	.48	-	-	-	-	-	-	-	-	-	-	-	-	-	89	31	0
g_G	.52	-	-	-	-	-	-	-	-	-	-	-	-	-			
g_H	-	-	-	-	.53	-	-	-	-	-	-	-	-	-	96	93	0
g_G	-	-	-	-	.57	-	-	-	-	-	-	-	-	-			
g_H	-	-	-	-	-	.60	-	-	-	-	-	-	-	-	95	95	0
g_G	-	-	-	-	-	.65	-	-	-	-	-	-	-	-			

At first all 14 measures are used for classification. The corresponding information gains $g_H(C)$ and $g_G(C)$ are listed in table 5, together with the percentages of correctly classified and uncertain beats on the test set, for record 200. Beats are labeled as uncertain if they are not covered by any rules of the fuzzy model. The percentage of uncertain beats (unc.) is defined with respect to the number of beats in the whole test set. The parameter with highest information gain is marked in bold. The ECG measures with smallest information gains are then progressively removed from the classification process. A similar table can be obtained for record 233.

Ventricular arrhythmia are mainly characterized by alterations in the QRS complex and T wave rather than in the PR segment. VPBs usually present a larger and higher QRS complex, and to a lower extent an alterated ST segment. In table 5 some ECG measures produce from the very beginning no information gain, such as the presence of the P wave (P), the average RR interval of the

previous 10 beats (RRa), and the PR interval (PR) as it was to be expected. Only 4-6 ECG features are characterized by a high information gain, that is are relevant for the classification process. An almost constantly used feature is the RR interval, that quantifies the prematurity of the beat and it is usually a sign for general arrhythmia. Also the QRS width, the positive and negative amplitude of the QRS complex and the corresponding areas, all parameters related to the QRS complex shape, play an important role, individually or together, in the classification procedure. If many input parameters are used, T wave features provide helpful information for classification, but they loose importance if no redundant input information is supplied. The low informative character of the past RR intervals, through the low information gain of the RRa parameter, confirms the unpredictability of VPBs. Individually, the positive and negative amplitude of the QRS complex present the highest information gain, confirmed by the highest performance on the test set, followed by the RR interval, the QRS width, and the QRS positive area.

All the estimated discriminative powers in table 5 find positive confirmation in clinical VPB diagnostics. The redundant or uninformative character of the input features with lowest information gain is proven by the fact that their removing does not affect the final performance on the test set, as long as at least two of the most significant ECG measures are kept. Indeed the same performance on the test set are observed both with the full input dimensionality and removing the least significant ECG measures.

Record 233 presents a new class of premature beats with supraventricular origin (SVPB) and a more homogeneous class of VPBs. Supraventricular arrhythmia can be differentiated from normal beats mainly by means of the RR interval and the PR segment, whenever the P wave can be reliably detected. Consequently the analysis of record 233, with respect to the analysis of record 200, shows a high information gain also for the PR measure, besides the negative amplitude and area of the QRS complex and the RR interval already used for VPB classification. However, if considered individually, none of the ECG measures produces a high information gain and good performance on the test set for all classes of beats. The PR interval shows to be useless if used alone for SVPB classification, but it gains a high discrimination power if any other significant ECG measure is added. The negative amplitude of the QRS complex and the RR interval alone show to be still highly discriminative for N/VPB classification, but helpless for SVPB recognition.

4 Conclusions

A methodology to estimate the discriminative power of input features based on an underlying fuzzy model is presented. Because of the approximative nature of fuzzy models, many algorithms exist to construct such models quickly from example data. Using properties of fuzzy logic, it is easy and computationally inexpensive to determine the possibilistic information gain associated with each input feature. The algorithm capability is illustrated by using an artifi-

cial example and the well-known IRIS data. The real-world feasibility was then demonstrated on a medical application.

The defined information gain provides a description of the class discriminability inside the adopted fuzzy model. This is related with classification performances, only if the fuzzy model was built on a sufficiently general set of training examples. The proposed algorithm represents a computationally inexpensive tool to reduce high-dimensional input spaces as well as to get insights about the system through the fuzzy model. For example, it can be used to determine which input features are exploited by fuzzy classifiers with better performance.

We believe that especially for large scale data sets in high dimensional feature spaces, such quick approaches to gain first insights into the nature of the data will become increasingly important to successfully find the underlying regularities.

5 Acknowledgments

The authors would like to thank Wei Zong, George. Moody, and prof. R.G. Mark from Harvard-MIT Division of Health Sciences and Technology M.I.T. (USA) for the ECG measures.

References

[1] V. Cherkassky and F. Mulier, "Learning from data", John Wiley and Sons Inc., 1998.

[2] C. Apte, S.J. Hong, J.R.M. Hosking, J. Lepre, E.P.D. Pednault, and B. K. Rosen, "Decomposition of heterogeneous classification problems", *Intelligent Data Analysis*, Vol. 2, n. 2, 1998.

[3] L.A. Zadeh, "A fuzzy-algorithmic approach to the definition of complex or imprecise concepts", *Int. J. Man-Machine Studies*, **8:** 249-291, 1976.

[4] A. De Luca, and S. Termini, "A definition of nonprobabilistic entropy in the setting of fuzzy sets theory",

[5] C.Z. Janikow, "Fuzzy Decision Trees: Issues and Methods", *IEEE Trans. Syst. Man and Cyb. PartB: Cybernetics*, **28:** 1-14, 1998.

[6] M. R. Berthold, K.P. Huber, "Comparing Fuzzy Graphs", Proc. of Fuzzy-Neuro Systems, pp. 234-240, 1998.

[7] M.R. Berthold, J. Diamond, "Constructive Training of Probabilistic Neural Networks", *Neurocomputing* 19: 167-183, 1998.

[8] R.A. Fisher, "The use of multiple measurements in taxonomic problems", *Annual Eugenics, II*, John Wiley, NY. **7**:179-188, 1950.

[9] C. Blake, E. Keogh, and C.J. Merz. UCI Repository of machine learning databases [http://www.ics.uci.edu/∼mlearn/MLRepository.html]. Irvine, CA: University of California, Department of Information and Computer Science, 1998.

[10] MIT-BIH database distributor, Beth Israel Hospital, Biomedical Engineering, Division KB-26, 330 Brookline Avenue, Boston, MA 02215, USA.

[11] W. Zong, D. Jiang. "Automated ECG rhythm analysis using fuzzy reasoning", Proc. of Computers in Cardiology, pp. 69-72, 1998.