Using Supervised Fuzzy Clustering and CWT for Ventricular Late Potentials (VLPs) Detection in High-Resolution ECG Signal

Gholamreza Salimi Khorshidi
School of cognitive sciences, Institute for studies in theoretical physics and mathematics, Tehran, Iran
gsalimi@bme.aut.ac.ir

Ayyoub Jaafari
Faculty of biomedical engineering, Amirkabir University of technology (Tehran Polytechnic), Tehran, Iran
ajafari@bme.aut.ac.ir

Amir Pourmorteza
Control And Intelligent Processing Center Of Excellence, Electrical And Computer Engineering Department, University Of Tehran, Tehran, Iran
amir.pourmorteza@gmail.com

Ali Motie Nasrabadi
Group of biomedical engineering, Faculty of Engineering, Shahed University, Tehran, Iran
nasrabadi@shahed.ac.ir

Mohammadreza Hashemi Golpayegani
Faculty of biomedical engineering, Amirkabir University of technology, Tehran, Iran
hashemi_g@morva.net

Abstract

Ventricular Late Potentials (VLPs) are low-amplitude, high-frequency components that appear at the end of the QRS complex of a High-Resolution ECG (HRECG) record. VLPs are clinically useful for identifying post-MI (Myocardial Infarction) patients prone to Ventricular Tachycardia (VT) and Sudden Cardiac Death (SCD). In this paper, the Continuous Wavelet Transform (CWT) and a supervised fuzzy clustering algorithm are used together to detect VLP. The terminal part of the QRS complex in the Vector Magnitude (VM) waveform is processed with the CWT to extract a feature vector. Resulting time-frequency representation is subdivided into several sub bands, and the sum of the squared decomposition coefficients is computed in each region. Finally, a supervised Fuzzy clustering method, trained by an appropriate set of these feature vectors, is applied to this data in order to identify VLP.

Keywords: ECG, VLP, Fuzzy Clustering, CWT

1 Introduction

Most of the Sudden Cardiac Deaths (SCDs) due to cardiac diseases are thought to be initiated by Ventricular Tachycardia (VT) which is one of the most serious types of cardiac arrhythmia [11]. As the appearance of Ventricular Late Potentials (VLPs) is associated with VT, there is a clinical interest in detection of these signals as a non-invasive diagnosis for post-MI (Myocardial Infarction) patients prone to VT. VLPs are low-amplitude, high-frequency signals which appear at the end of the QRS complex and arise as a result of the late depolarization of damaged myocardium. Because of their very low amplitudes and the noise overlay, VLPs are obscure in a standard electrocardiogram (ECG) but they can be detected by a High-Resolution ECG (HRECG) record acquired using three orthogonal XYZ leads with a minimum sampling frequency of 1000Hz and a resolution of 12 bits [1,9,11].

There are various techniques to improve signal-to-noise ratio (SNR) in VLP analysis. Typically, several heart beats (200-300) are averaged to suppress the background noise and form the Signal Averaged ECG (SAECG). In this paper, a new fuzzy model structure is proposed for the VLP classification problem in which each rule can represent more than one class with different probabilities. Typical fuzzy classifiers consist of interpretable if-then rules with fuzzy antecedents and class labels in the consequent part. The antecedents (if-parts) of the rules partitions the input space into a number of fuzzy regions by means of fuzzy sets, while the consequents (then-parts) describe the output of the classifier in these regions.

Fuzzy logic improves rule-based classifiers by allowing the use of overlapping class definitions and improves the interpretability of the results by providing more insights into the decision making process. The automatic determination of compact fuzzy classifiers rules from data has been approached by several different techniques like; neuro-fuzzy methods, genetic-algorithm (GA) based rule selection, and fuzzy clustering in combination with GA-optimization. Generally, the bottleneck of the data-driven identification of fuzzy systems is the structure identification that requires nonlinear optimization. Thus for high-dimensional problems, the initialization of the fuzzy model becomes very significant.

Common initialization methods such as grid-type partitioning [26] and rule generation on extreme initialization, result in complex and non-interpretable initial models and the rule-base simplification and reduction steps become computationally demanding. To avoid these problems, fuzzy clustering algorithms were proposed [37]. However, the obtained membership values have to be projected onto the input variables and approximated by parameterized membership functions that deteriorate the performance of the classifier. This decomposition error can be reduced by using eigenvector projection [28], but the obtained linearly transformed input variables do not allow the interpretation of the model. To avoid the projection error and maintain the interpretability of the model, the proposed approach is based on the Gath-Geva (GG) clustering algorithm [22] instead of the widely used Gustafson-Kessel (GK) algorithm [38], because the simplified version of GG clustering allows the direct identification of fuzzy models with exponential membership functions [25]. Neither GG nor GK algorithm utilizes the class labels. Hence, they give suboptimal result if the obtained clusters are directly used to formulate a classical fuzzy classifier.

Hence, there is a need for fine-tuning of the model. This GA or gradient-based fine-tuning, however, can result in overfitting and thus poor generalization of the identified model. Unfortunately, the severe computational requirements of these approaches limit their applicability as a rapid model-development tool. This paper focuses on the design of interpretable fuzzy rule-based classifiers from data with low-human intervention and low-computational complexity. Hence, a new modeling scheme is introduced based only on fuzzy clustering. The proposed algorithm is similar to the Multi-Prototype Classifier technique [18, 33]. The main difference of this approach is that each cluster represents different classes, and the number of clusters used to approximate a given class has to be determined manually, while the proposed approach does not suffer from such problems. Generally, there is a very large set of possible features to compose feature vectors of classifiers. As it is ideal that the training dataset's size should increase exponentially with the feature vector size, it is desired to choose a minimal subset among it. Some generic tips to choose a good feature set include the facts that they should discriminate as much as possible the pattern classes and they should not be correlated or redundant.

2 Simson Method

The conventional time-domain method of VLP detection, developed by Simson [8], is based on feature extraction from the filtered SAECG [8, 9, 13]. Simson's method employs a high-pass filter (cutoff frequency of 25 or 40Hz) to attenuate low-frequency components of averaged XYZ signals (SAECG). To avoid the filter ringing effect in the terminal parts of QRS complex, Simson proposed a bi-directional four-pole Butterworth high-pass filter [8]. After high-pass filtering of the averaged XYZ signals, these signals are combined into a Vector Magnitude (VM) waveform defined by

$$VM = \sqrt{X^2 + Y^2 + Z^2} \qquad (1)$$

After estimating the onset and offset of the filtered QRS complex (the QRS complex in the VM signal), three conventional time-domain features can be measured to detect VLPs [1, 11, 12, 13]:

- QRS_T: Duration of the filtered QRS complex (from the onset to the offset)
- $D_{40:}$ Low-amplitude signal duration (from the offset backward to the point where VM reaches the 40μV)
- $V_{40:}$Root-mean-square value of the last 40ms of the filtered QRS (showed in Fig1)

The criteria to define a VLP positive test are QRS_T>114ms, D_{40}>38ms and V_{40}<20μV [11]. In Fig.1, a plot of a typical filtered QRS complex and the definition of the conventional time-domain features, introduced above, can be viewed.

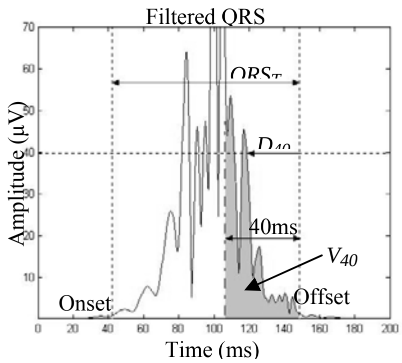

Fig.1 A typical filtered QRS complex and the definition of three features: QRS_T, D_{40} and V_{40}.

3 Feature Extraction

The adopted feature extraction method for VLP consists of five steps:

- Averaging XYZ leads to improve the SNR
- Bidirectional Butterworth filtering of averaged *XYZ* signals consisting of a 4th order high-pass and a 5th order low-pass filters with cutoff frequencies of 40Hz and 250 Hz respectively.
- Combination of the filtered averaged signals into a *VM* waveform using equation (1).
- Applying the CWT to the terminal part of the QRS complex in the *VM* signal.
- Feature extraction from the resulted time-scale plot.

In recent years, the wavelet analysis has been used widely in biomedical researches [2, 5, 7, and 14]. The wavelet transform is a linear time-frequency transform which is based on decomposition of a signal using a set of basis functions. These basis functions are scaled and shifted versions of a prototype mother wavelet [4, 11]. The CWT produces a time-frequency representation of the signal which is a function of time τ and scale a, like CWT (τ, a) named wavelet decomposition coefficient. The scale can be considered as the inverse of the frequency [9].

The smaller scales bring about a higher resolution in time which is useful to detect VLPs as high-frequency, short-duration signals. In this study, the CWT is adopted, using the MATLAB wavelet toolbox, to the last 40ms of the QRS complex in the *VM* signal after estimating the offset point. To improve the robustness of the method to the error in QRS offset detection, a 8ms right shift is considered for the estimated offset [13] so that the CWT is applied to 48ms interval of the *VM* waveform, with scale range of 1-8 and the Morlet wavelet as the mother wavelet [6]. Then, the resulted time-scale representation is subdivided into 81 regions [6] (nine subdivisions on both the time and scale axes). Finally, the sum of the squared wavelet decomposition coefficients is computed in each region to form a feature vector composed of the 81 elements.

4 Supervised Fuzzy Clustering

The objective of clustering is to partition the identification data Z into *R* clusters. This means, each observation consists of input and output variables grouped into a row vector $Z_k = [x^T_k\ y_k]$, where the *k* subscript denotes the $k = 1\ldots N$th row of the pattern matrix of Z. The fuzzy partition is represented by the $U= [\mu_{i,k}]_{R*N}$ matrix, where the $\mu_{i,k}$ element of the matrix represents the degree of membership, how the Z_k observation is in the cluster $i = 1\ldots R$. The clustering is based on the minimization of the sum of weighted $D^2_{i,k}$ squared distances between the data points and the η_i cluster prototypes that contains the parameters of the clusters.

$$J(Z,U,\eta) = \sum_{i=1}^{R}\sum_{k=1}^{N}(\mu_{i,k})^m D^2(z_k, r_i) \quad (3)$$

In which, m is the fuzzy weighting exponent that determines the fuzziness of the resulting clusters. Usually, the chosen value for m will be m = 2. Classical fuzzy clustering algorithms are used to estimate the distribution of the data. Hence, they do not utilize the class label of each data point available for the identification. Furthermore, the obtained clusters cannot be directly used to build the classifier. In the following a new cluster prototype and the related distance measure will be introduced that allows the direct supervised identification of fuzzy classifiers. As the clusters are used to obtain the parameters of the fuzzy classifier, the distance measure is defined similarly to the distance measure of the Bayes classifier:

$$\frac{1}{D^2_{i,k}(z_k, r_i)} = P(r_i)\prod_{j=1}^{n}\exp\left(-\frac{1}{2}\frac{(x_{j,k}-v_{i,j})^2}{\sigma^2_{i,j}}\right) P(c_j = y_k \mid r_i) \quad (4)$$

This distance measure consists of two terms; the first term is based on the geometrical distance between the v_i cluster centers and the x_k observation vector, while the second is based on the probability that the r_i-th cluster describes the density of the class of the k-th data, P ($c_j = y_k$ | r_i) It is interesting to note, that this distance measure only slightly differs from the unsupervised Gath–Geva clustering algorithm which can also be interpreted in a probabilistic framework [21]. However, the novelty of the proposed approach is the second term, which allows the use of class labels. To get a fuzzy partitioning space, the membership values have to satisfy the following conditions:

$$U \in R^{C\times N},\ \mu_{i,k} \in [0,1],\ \forall i,k;$$

$$\sum_{i=1}^{R}\mu_{i,k} = 1,\ \forall k;\ 0 < \sum_{k=1}^{N}\mu_{i,k} < N,\ \forall i \qquad (5)$$

The minimization of the (6) functional represents a non-linear optimization problem that is subject to constraints defined by (5) and can be solved by using a variety of available methods. The most popular method, is the alternating optimization (AO), which consists of the application of Picard iteration through the first-order conditions for the stationary points of (6), which can be found by adjoining the constraints (5) to J by means of LaGrange multipliers [24] and setting the gradients of J to zero, with respect to Z, U, η and λ.

$$\bar{J}(Z,U,\eta,\lambda) = \sum_{i=1}^{R}\sum_{k=1}^{N}(\mu_{i,k})^m D^2(z_k,r_i) + \sum_{k=1}^{N}\lambda_k\left(\sum_{j=1}^{R}\mu_{i,k} - 1\right) \qquad (6)$$

Hence, similarly to update the Gath-Geva clustering algorithm equations, the following algorithm will result in a solution that satisfies the (6) constraints.

Initialization: Given a set of data Z specify R, choose a termination tolerance ε>0. Initialize the U= [$\mu_{i,k}$] RεN partition matrix randomly, where $\mu_{i,k}$ denotes the membership that the Z_k data is generated by the i'th cluster. Repeat for l = 1, 2….

Step1: Calculate the parameters of the clusters. Calculate the centers and standard deviation of the Gaussian membership functions (the diagonal elements of the F_i covariance matrices) (7) and estimate the consequent probability parameters (8) and *a priori* probability of the cluster and the weight (impact) of the rules (9).

$$v_i^{(l)} = \frac{\sum_{k=1}^{N}(\mu_{i,k}^{(l-1)})^m x_k}{\sum_{k=1}^{N}(\mu_{i,k}^{(l-1)})^m}$$

$$\sigma_{i,j}^{2(l)} = \frac{\sum_{k=1}^{N}(\mu_{i,k}^{(l-1)})^m (x_{j,k} - v_{j,k})^2}{\sum_{k=1}^{N}(\mu_{i,k}^{(l-1)})^m} \qquad (7)$$

$$p(c_i \mid r_j) = \frac{\sum_{k \mid y_k = c_i}(\mu_{j,k}^{(l-1)})^m}{\sum_{k=1}^{N}(\mu_{j,k}^{(l-1)})^m}$$

$$1 \le i \le C,\ 1 \le j \le R \qquad (8)$$

$$P(r_i) = \frac{1}{N}\sum_{k=1}^{N}(\mu_{i,k}^{(l-1)})^m$$

$$w_i = P(r_i)\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi\sigma_{i,j}^2}} \qquad (9)$$

Step2: Compute the distance measure $D^2_{i,k}$ by (4).

Step3: Update the partition matrix

$$\mu_{i,k}^{(l)} = \frac{1}{\sum_{j=1}^{R}(D_{i,k}(z_k,r_i)/D_{j,k}(z_k,r_j))^{2/(m-1)}},$$

$$1 \le i \le R,\ 1 \le k \le N \quad untill \ \|U^{(l)} - U^{(l-1)}\| \langle \varepsilon \qquad (10)$$

5 Feature Selection

Using too many input variables may result in difficulties in the interpretability capabilities of the obtained classifier. Hence, selection of the relevant features is usually necessary. In this paper, the modified Fischer interclass separability method is used which is based on statistical properties of the data. The interclass separability criterion is based on the F_B between-class and the F_W within class covariance matrices that sum up to the total covariance of the data F_T=F_B+F_W, where:

$$F_W = \sum_{l=1}^{R} P(r_l) F_l \ ,$$

$$F_B = \sum_{l=1}^{R} P(r_l)(v_l - v_0)^T (v_l - v_0),$$

$$v_0 = \sum_{l=1}^{R} P(r_l) v_l \qquad (11)$$

$$J = \frac{\det(F_B)}{\det(F_w)} \qquad (12)$$

The feature interclass seperatibility selection criterion is a trade-off between F_W and F_B (Eq.12). The importance of a feature is measured by leaving out the interested feature and calculating *J* for the reduced covariance matrices. The feature selection is a step-wise procedure, when in every step the least needed feature is deleted from the model [29].

5 Results and Discussion

To evaluate the method used in this study a HRECG database consisting of two groups of signals was selected. The first group contained 50 healthy volunteers' HRECG records acquired by a digital data acquisition system, ML785 PowerLab/8SP, with a sampling frequency of 2000Hz and a 16-bit analog-to-digital converter (ADC). The second group was consisted of semi-simulated HRECG signals with VLPs. In order to simulate each of these signals, three basic simulated waveforms resembling the VLP characteristics were added to *XYZ* leads of a basic HRECG record, a HRECG record without VLP. VLPs are low-amplitude signals (~1-20μV) with short duration (~5-50ms) and broadband spectrum (~40-250Hz) [11]. According to these characteristics, VLPs were simulated as colored Gaussian processes resembling better the real world signals. The basic VLP waveforms were added to the end part of the QRS complex of every heart beat of the *XYZ* leads belonging to the basic HRECG records. The position of the VLPs was varied randomly from beat to beat with respect to the fiducial mark, QRS peak [10, 11]. This HRECG database was divided into a training set, including thirty HRECG signals with VLPs and thirty without, and a test set consisting of 20 records without VLPs and 20 with. For a better training of the neural network and preserving its generalization, the training set was expanded; five sets containing 300 heart beats were selected from every HRECG record of training set randomly. Because of the fact that every HRECG record had at least 350 heart beats, the beat selection was done without replacement for each set. Therefore, an expanded training set consisting of 300 patterns was obtained. The performance of the VLP detection method was measured using conventional criteria i.e. the accuracy *ACC*, sensitivity *SE*, and specificity *SP* defined by

$$ACC = 100 \times (TP + TN) / N$$
$$SE = 100 \times TP / (TP + FN)$$
$$SP = 100 \times TN / (TN + FP) \qquad (13)$$

Where *N, TP, TN, FP*, and *FN* are respectively the total number of patterns, the number of true positive, the number of true negative, the number of false positive, and the number of false negative [6]. Using the expanded training set and the test set, the method based on the CWT and the Fuzzy Supervised Clustering system, introduced in Fig.2, was evaluated that showed good results for the test set. To investigate the performance of the VLP detection method proposed in this work, the conventional time-domain method (Simson's method) was applied to the test set; also, a method based on applying our system to the conventional time-domain features [4,13] was used to detect VLP. Table1 presents the results of the proposed method in comparison with Simson's method and applying a Proposed Fuzzy System to the conventional time-domain features, for the test set. In Simson's method, the balance between *SE* and *SP* can be controlled by choosing one, two, or three of the positive VLP criteria (QRS_T>114ms, D_{40}>38ms, and V_{40}<20μV) at the same time. For example, if higher *SE* is desired, only one of the three criteria is chosen; in contrast, all of the criteria must be satisfied at the same time to achieve higher *SP*. This fact can be seen in Table 1. In most cases, two of the three criteria are used to obtain the balance between *SE* and *SP* [13]. These results obtained with a Fuzzy system with 5 rules.

Table1: VLP detection results. The comparison between the proposed method, Simson's method and applying a Supervised Fuzzy Clustering system to the features (for the test set).

Method		ACC(%)	SE(%)	SP(%)
Simson's Method	One Criterion	80	100	60

Two Criterion	75	80	70
Three Criterion	72	60	85
Applying Neural Network Method to mVM_CWT1 Features	92.5	91	94
Applying Supervised Fuzzy Clustering to the conventional time-domain features	82	81.26	83.66
Applying Supervised Fuzzy Clustering to VM_QRS40 Features	84	73.76	94.13
Applying Supervised Fuzzy Clustering to mVM_QRS40 Features	90	89.88	91.16
Applying Supervised Fuzzy Clustering to mVM_CWT1 Features	92	88.83	95.16

6 Conclusion

The aim of this work was to investigate the capability of a method based on the continuous wavelet transform and a Supervised Fuzzy Clustering, in order to extract features and classify for VLP detection problem in HRECG. The results show good improvements in sensitivity and discriminancy compared to Simson's method and Supervised Fuzzy Clustering to the conventional time-domain features (see Table1). Another possible advantage of the proposed method can be the ability of the VLP detection in patients with bundle branch block, while Simson's method can not be used with these patients because the bundle branch block causes the QRS duration to be extended and consequently increases QRS_T. Simson's method is very sensitive to QRS offset, and even a few errors in estimating of this point can result in wrong diagnostic. However the method used in this study is robust to the error in QRS offset detection. The results represented in Table 1 shows that the CWT based features, used in this study, are more capable than the conventional time-domain features to detect VLPs using Supervised Fuzzy Clustering Systems. According to the above advantages, the proposed method may be an appropriate alternative method for the clinical detection of VLP, if it is evaluated completely. Due to the lack of real HRECG for VLP in this research, the proposed method must be applied to a larger HRECG database consisting of real signals with and without VLPs to complete the evaluation. Form Table1, Neural Network Method is a little better than Our Supervised Clustering method but our system is based on only five fuzzy rules and so the training of system is much faster than a MLP Neural Network and can be used in practical applications.

References

[1] I.C.Baykal and A.Yilmaz, Detection of Late Potentials in Electrocardiogram Signals in both Time and Frequency Domains Using Artificial Neural Networks, *44th IEEE Midwest Symposium on Circuits & Systems*, 2001, pp.576-579.

[2] S.W.Chen, A Wavelet-Based Heart Rate Variability Analysis for the Study of Nonsustained Ventricular Tachycardia, IEEE Transactions on Biomedical Engineering, Vol.49, No.7, 2002, pp.736-742.

[3] A.Cohen , Biomedical Signal Processing, CRC Press, VolumeII, 1986.

[4] A.Mousa and A.Yilmaz, Neural Network Detection of Ventricular Late Potentials in ECG Signals Using Wavelet Transform Extracted Parameters, IEEE Proceedings of the 23rd Annual EMBS International Conference, Turkey, 2001, pp.1668-1671.

[5] A.Mousa and A.Yilmaz, A Method Based on Wavelet Analysis for the Detection of Ventricular Late Potentials in ECG Signals, 44th IEEE Midwest Symposium on Circuits & Systems, 2001, pp.497-500 .

[6] A.Rakotomamonjy, B.Migeon, and P.Marche, Automated Neural Network Detection of Wavelet preprocessed Electrocardiogram Late Potentials, Med.Biol.Eng.Comput.,Vol.36, 1998, pp.346 - 350.

[7] M.B.Simson, Use of Signals in The Terminal QRS Complex to Identify Patients with Ventricular Tachycardia after Myocardial Infarction, Circulation, Vol.64, No.2, 1981, pp. 235-242.

[8] A.Spaargaren and M.J.English, Detecting Ventricular Late Potentials Using the Continuous Wavelet Transform, Proceedings of Computers in Cardiology,IEEE Comput.Soc.,Vol.26, 1999, pp.5-8.

[9] A.Taboada-Crispi, J.V.Lorenzo-Ginori, and D.F.Lovely, Adaptive Line Enhancing Plus Modified Signal Averaging for Ventricular Late Potential Detection, Electronics Letters, Vol.35, No.16, 1999, pp.1293-1295.

[10] A.Taboada-Crispi, Improving Ventricular Late Potentials Detection Effectiveness, Doctoral Thesis, University of New Brunswick,2002.

[11] S.Wu, Y.Qian, Z.Gao, and J.Lin, A Novel Method for Beat-to-Beat Detection of Ventricular

Late Potentials, IEEE Transactions on Biomedical Engineering,Vol.48,No.8,2001, pp.931-935.
[12] Q.Xue and B.R.S.Reddy, Late Potential Recognition by Artificial Neural Networks, IEEE Transactions on Biomedical Engineering, Vol.44, No.2, 1997, pp.132-143.
[13] Janos Abonyi and Ferenc Szeifert , Supervised Fuzzy Clustering for the Identification Of Fuzzy Classifiers , University of Veszprem, Dep. of Process Engineering,
[14] Detection of Ventricular Late Potentials in High-Resolution ECG Signals by a Method Based on the Neural Network and Continuous Wavelet Transform , Ali Shahidi Zandi , Mohammad Hassan Moradi
[15] Baraldi A. and Blonda P. (1999) A survey of fuzzy clustering algorithms for pattern recognition - Part I, IEEE Transaction On Systems, Man and Cybernetics Part B29 (6): 778-785.
[16] Bezdek J.C., Hathaway R.J., Howard R.E., Wilson C.A. and Windham M.P., (1987) Local Convergence Analysis of a Grouped Variable Version of Coordinate Descent. Journal of Optimization Theory and Applications, 71:471–477.
[17] Biem A., Katagiri S., McDermott E., Juang BH. (2001) An application of discriminative feature extraction lo filter-bank-based speech recognition.IEEE Transactions On Speech And Audio Processing 9 (2): 96-110.
[18] Campos T. E., Bloch I., Cesar R. M. Jr. (2001) Feature Selection Based on Fuzzy Distances Between Clusters: First Results on Simulated Data.Lecture Notes in Computer Science, Springer-Verlag press, ICAPR'2001 - International Conference on Advances in Pattern Recognition, Rio de Janeiro, Brazil, May.16
[19] Cios K.J., Pedrycz W., Swiniarski R.W. (1998) Data Mining Methods forKnowledge Discovery. Kluwer Academic Press, Boston.
[20] Corcoran A.L., Sen S. (1994) Using real-valued genetic algorithms to evolve rule sets for classification. In IEEE-CEC, June 27-29, 120–124, Orlando, USA.
[21] Gath I. and Geva. A.B. (1989) Unsupervised Optimal Fuzzy Clustering,IEEE Transactions on Pattern Analysis and Machine Intelligence 7 773–781.
[22] Gustafson, D.E., Kessel, W.C. (1979) Fuzzy Clustering With a Fuzzy Covariance Matrix, In Proc. IEEE CDC, San Diego, USA.
[23] Hathaway R.J. and Bezdek J.C. (1993) Switching Regression Models and Fuzzy Clustering. IEEE Transactions on Fuzzy Systems, 1:195–204.
[24] Hoppner F., Klawonn F., Kruse R. and Runkler T. (1999) Fuzzy Cluster Analysis – Methods for Classification, Data Analysis and Image Recognition,John Wiley and Sons.
[25] Ishibuchi H., Nakashima T., Murata T. (1999) Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems.IEEE Trans. SMC–B 29, 601–618.
[26] Kambhatala N. (1996) Local Models and Gaussian Mixture Models for Statistical Data Processing, Ph.D. Thesis, Oregon Gradual Institute of Science and Technology.
[27] Kim E., Park M., Kim S. and Park M. (1998) A Transformed Input–Domain Approach to Fuzzy Modeling. IEEE Transactions on Fuzzy Systems, 6:596–604.
[28] Loog L.C. M., Duin R.P.W. ,Haeb-Umbach R. (2001) Multiclass linear dimension reduction by weighted pairwise Fisher criteria, IEEE Trans. On PAMI, vol. 23, no. 7, pp. 762-766.
[29] Nauck D. and Kruse R. (1999) Obtaining interpretable fuzzy classification rules from medical data. Artificial Intelligence in Medicine, 16: 149–169.
[30] Pe˜na-Reyes C. A. and Sipper M. (2000) A fuzzy genetic approach to breast cancer diagnosis. Artificial Intelligence in Medicine, 17: 131–155.
[31] Quinlan J. R. (1996) Improved Use of Continuous Attributes in C4.5, Journal of Artificial Intelligence Research, 4: 77–90.
[32] Rahman A.F.R. and Fairhurst M.C. (1997) Multi-prototype classification: improved modelling of the variability of handwritten data using statistical clustering algorithms. Electron. Lett. 33 14, pp. 12081209.17
[33] Roubos J.A., Setnes M. (2000) Compact fuzzy models through complexity reduction and evolutionary optimization. In FUZZ-IEEE, pp762-767, May 7-10, San Antonio, USA.
[34] Roubos J.A., Setnes M. and Abonyi J. (2001) Learning fuzzy classification rules from data. Developments in Soft Computing, eds. John, R. and Birkenhead, R., Springer - Verlag Berlin/Heidelberg, 108-115.
[35] Setiono R. (2000) Generating concise and accurate classification rules for breast cancer diagnosis. Artificial Intelligence in Medicine, 18: 205–219.
[36] Setnes M., Babuˇska R. (1999) Fuzzy Relational Classifier Trained by Fuzzy Clustering, IEEE Trans. SMC–B, 29, 619–625
[37] Setnes M., Babuˇska R., Kaymak U., van Nauta Lemke H.R. (1998) Similarity measures in fuzzy rule base simplification. IEEE Trans. SMC–B 28, 376–386.
[38] Takagi T., Sugeno M. (1985) Fuzzy identification of systems and its application to modeling and control. IEEE Trans. SMC 15, 116–132.
[39] Valente de Oliveira J. (1999) Semantic constraints for membership function optimization. IEEE Trans. FS 19, 128–138.