Hyperbox classifiers for ECG beat analysis

G Bortolan[1], II Christov[2], W Pedrycz[3]

[1]Institute of Biomedical Engineering ISIB - CNR, Padova, Italy
[2]Centre of Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria
[3]Dept of Electrical & Computer Engineering, University of Alberta, Edmonton, Canada and Systems Research Institute Polish Academy of Sciences, Warsaw, Poland

Abstract

Hyperbox classifiers have been investigated for the detection and classification of different types of heartbeats in the ECG, which is of major importance in the diagnosis of cardiac dysfunctions. In particular, the learning capacity and the classification ability for normal beats (N) and premature ventricular contractions (PVC) have been tested, with particular interest in the aspect of the interpretability of the results. The MIT-BIH arrhythmia database has been used for testing and validating the proposed method. A total of 26 morphology features have been extracted from ECG and reconstructed VCG signals. Three learning process have been tested combining the fuzzy clustering and the genetic algorithm for identifying the optimal hyperboxes and for testing a family of hyperellipsoid. The results showed that a limited number of hyperboxes increased the geometrical interpretability without a significant reduction of the accuracy.

1. Introduction

ECG signal analysis is a useful way to study and diagnose cardiac dysfunctions. Detection and classification of ECG beats is of considerable importance in real-time critical care or patient monitoring. Pattern recognition algorithms reveal a great tool in the computerized classification techniques. The connectionist approaches, k-Nearest Neighbour and fuzzy clustering are only some examples of a huge amount of pattern classifiers [1-4]. There is an increasing interest in studying and examining specific classification techniques in which the induced feature space possess a simple geometry, which becomes beneficial mainly in the process of interpreting the results.

Hyperbox classifiers are a useful tool for a high level of interpretability of the classification rules and a simple structure of the overall classifier so that one could easily visualize the obtained results. The learning process is developed with hybrid learning strategy which consists of a certain combination of fuzzy clustering, more specifically C-Means and the genetic algorithm. The hyperbox classifiers have been investigated for the detection and classification of different types of heartbeats in the electrocardiogram (ECG). In addition, the family of hyperellipsoids have been studied and tested as a generalization of the geometry supported by hyperbox classifiers. The MIT-BIH arrhythmia database was used for testing and validating the proposed methods.

2. Methods and material

2.1. The ECG database

ECG recordings from the Massachusetts Institute of Technology - Beth Israel Hospital (MIT-BIH) arrhythmia database were used. Each record consists of two leads sampled at 360 Hz for a period of 30 minutes [5]. Two types of heartbeats are considered in this study: Normal (N) and Premature Ventricular Contractions (PVC). A total of 26 morphology features have been extracted from ECG and reconstructed VCG signals and include amplitudes, areas, durations and combinations of them [6]. A set of 22 ECG records out of the 48 present in the MIT-BIH ECG arrhythmia database has been used for testing and validating the proposed method. This selection was made considering the records with highest number of PVC, and no selection based on the quality of the signal was performed. Each record was randomly divided in learning set (66.7%) and test set (33.3%).

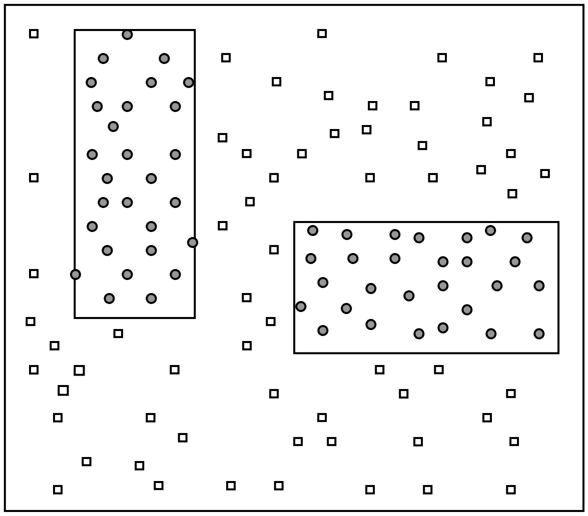

Fig. 1 Example in two dimensional space. Two hyperboxes cover class W_1 (denoted by dots), and class W_2 (squares) is not considered in the classifier.

2.2. Hyperboxes and hyperellipsoids

The hyperboxes (HB) are regions formed in the feature space which are described in the following form:

$$\{x_i \in [x_{iA}, x_{iB}]\} \quad i = 1,.....,N$$

Given this description, one can directly directly translate them into the following well-known format of rule-based classification rules:

"IF feature X_1 assume values in $[X_{1A}, X_{1B}]$ and feature X_2 assume values in $[X_{2A}, X_{2B}]$ … and feature X_N assume values in $[X_{NA}, X_{NB}]$ THEN class W_i"

where the interval limits are the respective edges of the corresponding hyperbox. An example of hyperbox-based classifier formed in a two dimensional feature space is shown in Fig.1, where two hyperboxes cover data of class W_1, whereas the patterns belonging to class W_2 are excluded.

The hyperboxes exhibit some important characteristics:

- the underlying geometry of the hyperboxes is simple and comes with a visual interpretation
- they can be easily interpreted as a set of IF-THEN rules
- they explicitly localize (and describe) the region in the feature space that are pertinent to the given class of patterns

An interesting and intuitively appealing generalization og hyperboxes comes in the form hyperellipsoids, a family of shapes formed from the spherical product of two superquadratic curves. This family can be used to model a wide range of shapes including sphere, ellipsoids, and parallelepipeds, as well all shapes in-between [7].

For example, in the case of three dimensional space, a hyperellipsoid is described by the following equation:

$$\left(\frac{x}{x_r}\right)^m + \left(\frac{y}{y_r}\right)^m + \left(\frac{z}{z_r}\right)^m = 1$$

where x_r, y_r and z_r are the equatorial and polar radii, and the parameter **m** determines the form (shape) of the object. In particular, we encounter:

m = 1 diamond
m = 2 sphere ($x_r = y_r = z_r$)
m = 2 ellipsoid
m = 4 hyperellipsoid
m = inf hyperbox

2.3. The learning process

In order to characterize and to search in the feature space the optimal HB that better characterize the considered class, different learning processes have been developed with the use and the combination of fuzzy clustering and genetic algorithms. In particular, the following three methods have been considered and tested:

1. (HB-2PH) The learning process is developed with a hybrid architecture in two phases. The fuzzy clustering (Fuzzy C-Means) represent the first step, in which the clusters are localized and characterized by the corresponding prototypes. In the second step, hyperboxes will be defined around the prototypes, and optimized with the use of genetic algorithm, which are very suitable for problems with high dimensions [8].
2. (HB-GA) The learning process is developed using only the genetic algorithm, and the optimal hyperboxes will be characterized by the ability of the GA to find the optimal solution of the HB.
3. (HB-SE) The GA has been used for determining the optimal generalized hyperbox that better represents and discriminate the analyzed classes. For this purpose, a family of hyperellipsoids was sought. For this experiment, in order to have a visual interpretability of the results, the case of 3-D was considered.

Two sets of features were taken into account in the experiments:

A. all the 26 features were considered in the HB-2PH learning strategy for testing the global capacity of the classifier.

B. a limited number of features was considered in the HB-GA and HB-SE learning strategies were implemented In this way we are at position to better study the aspects of geometrical interpretability of the classifier. In particular, maximal positive and negative peak and the area of QRS complex in lead 1 were selected.

3. Results

Different experiments were performed in order to test and validate the capacity of the proposed technique. The first experiment considers a two-phase learning strategy (HB-2PH). The first step of fuzzy clustering was useful for the determination of the prototypes and consequently to delimit the search region when running the genetic algorithm. All the 26 parameters were used for testing the potential of the proposed method. Different number of hyperboxes (1, 2 and 7) was considered. The mean sensitivity and specificity are reported on Table 1 for the classification of N beats. Table 2 shows the results obtained for the classification of PVC beats.

With one hyperbox the mean sensitivity and specificity were 97.8% and 97.8% for N and 79.7% and 99.7% for PVC, whereas two hyperboxes produced 98.0% and 99.2% for N and 78.3% and 98.2% for PVC. The use of seven hyperboxes improved the results, but not significantly, and the drawback of limiting the interpretability and increasing the complexity of the classifier. These findings indicate that a low number or hyperboxes are quite sufficient to describe the feature space of N or PVC classes.

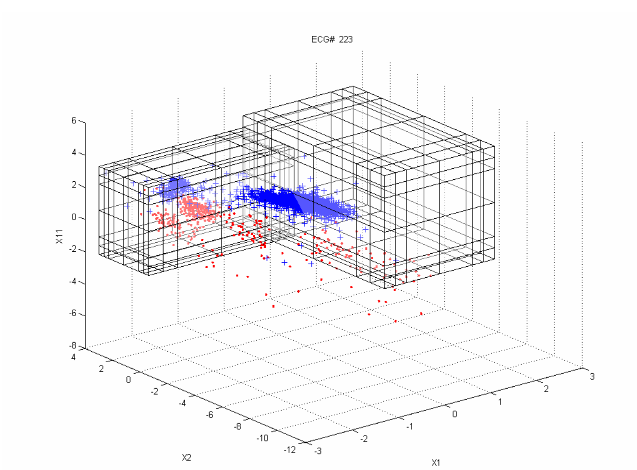

Fig. 2. Example of classification of N class with HB-2PH method with two hyperboxes (PVC: red dots, N blue + marks)

Table 1. Classification of N class with HB-2PH.

# of HB	Se	Sp
1	97.8%	97.8%
2	98.0%	99.2%
7	98.2%	98.9%

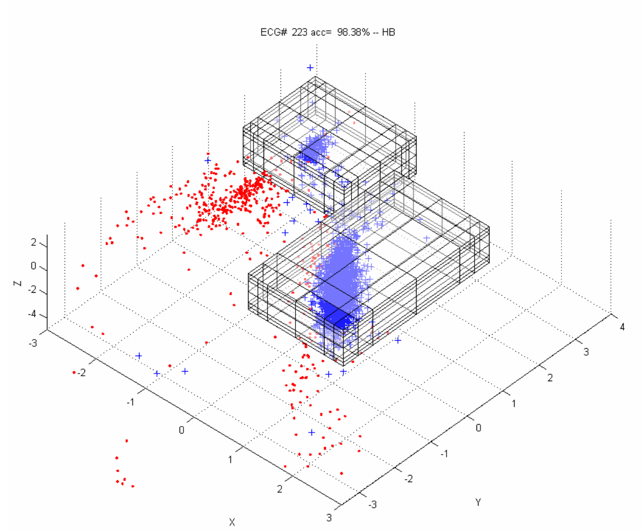

Fig. 3. Example of classification of N class with HB-GA method with two hyperboxes (PVC: red dots, N blue + marks)

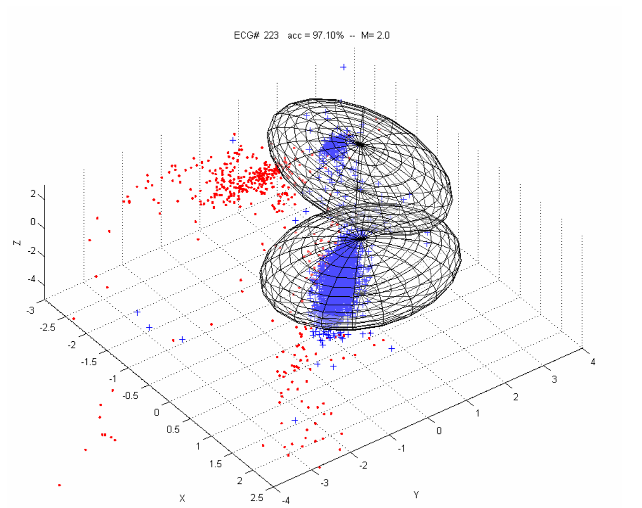

Fig. 4. Example of classification of N class with HB-SE method with two hyperellipsoid (m=2) (PVC: red dots, N blue + marks).

Table 2. Classification of PVC class with HB-2PH.

# of HB	Se	Sp
1	79.7%	99.7%
2	78.3%	98.2%
7	82.3%	97.2%

Table 3. Classification of N class with HB-GA.

# of HB	Se	Sp
1	98.6%	94.0%
2	98.6%	96.8%
3	98.6%	96.6%

Table 4. Classification of N class with HB-SE and m=2.

# of HB	Se	Sp
1	98.6%	94.6%
2	98.8%	96.4%
3	98.9%	96.9%

Figure 2 reports an example (ECG #223) with two HB, and visualizing only three features. The discriminant capacity of the two hyperboxes is visible only partially due to the imposed reduction of the feature space.

In the second experiment, the learning procedure used only the Genetic Algorithm (HB-GA). In this case 3 parameters were used for a more direct way to evaluate the geometrical interpretability of the results. The results are reported in Table 3 for the classification of N beats, considering 1, 2 and 3 HB. Two HB produces a sensitivity of 99.5% and a specificity of 96.8%. Figure 3 reports the same ECG example, and in this case the discriminant ability of the two HB is more evident.

The last experiment considers the use of GA for searching the optimal hyperellipsoid (HB-SE), with the use of three features. Table 4 reports the results with two hyperellipsoid (m=2). For example, the use of two HB produce a light improvement in the accuracy, and the visual inspection of the example reported in Fig. 4 shows clearly the ability and the geometrical interpretability of the classifier.

4. Discussion and conclusions

Hyperbox classifiers have been investigated for the detection and classification of different types of heartbeats in the ECG, which is of major importance in the diagnosis of cardiac dysfunctions. In particular, the learning capacity and the classification ability for normal beats (N) and premature ventricular contractions (PVC) have been tested, with particular interest in the aspect of the interpretability of the results. The MIT-BIH arrhythmia database has been used for testing and validating the proposed method. The results showed that a limited number of hyperboxes or hiperellipsoids increased the geometrical interpretability without a significant reduction of the accuracy.

References

[1] Christov I, Jekova I, Bortolan G, Premature ventricular cantraction classification by the Kth nearest-neighbours rule, Physiological Measurement, 2005; 26:123-130.

[2] Laguna P, Jane R, Caminal P, Adaptive feature extraction for QRS classification and ectopic beat detection. In: Murray A, Arzbaecher R Eds.Computers in Cardiology, 1991; 613-616.

[3] Ham FM, Ham S, Classification of cardiac arrhythmias using fuzzy ARTMAP, IEEE Trans. Biomedical Engineering, 1996; 43:425-430.

[4] Al-Nashash H, Cardiac arrhythmia classification using neural networks, Technol Health Care, 2000; 8:363-72.

[5] Mark R, Moody G, MIT-BIH Arrhythmia data base directory, Massachusetts Institute of Technology, Cambridge, 1988.

[6] Christov I, Bortolan G, Ranking of pattern recognition parameters for premature ventricular cantractions classification by neural networks, Physiological Measurement, 2004; 24:1281-1290.

[7] Jaklic A, Leonardis A, Solina F, Segmentation and recovery of superquadrics, Kluwer Academic Publishers, London, 2000.

[8] Bortolan G, Pedrycz W, Hyperbox Classifier for arrhythmia classification, Kybernetes, 2007; 36:531-547.

Address for correspondence

Giovanni Bortolan

Institute of Biomedical Engineering ISIB - CNR

Corso Stai Uniti, 4

35127 Padova, Italy

E-mail: bortolan@isib.cnr.it