Comparison of Methods for Premature Ventricular Beat Detection

Václav Chudáček, George Georgoulas, Chrysostomos Stylios, Martin Staviař, Martin Hanuliak, Lenka Lhotská, *Member, IEEE*

Abstract— **Holter ECG monitoring is used for long-term monitoring of patients with heart problems for diagnosis reasons. A lot of research work has been don in this field and many methods and procedures have been investigated. This paper discusses and compares a number of different approaches of clustering algorithms focusing on distinguishing premature ventricular complexes (V) from the normal (N) beats. Algorithms were tested on MIT-BIH database and results are computed for local and global training sets. Template matching technique using rule-based decision tree clustering algorithm for data reduction performed best with specificity of 96.63 % and sensitivity of 92.64 %.**

I. Introduction

HOLTER ECG monitoring [1] is used for surveillance patients with heart problems such as arrhythmias. Heart beats with unusual timing or unusual ECG morphology can be very helpful in erlry diagnosis of hearts with pathoelectrophysiology.

Many different methods have been proposed to solve the problem of discrimination between normal (N) and premature ventricular beats (V). Some are based on using beat shape description parameters [4, 5, 6] as well as beat representation by frequency-based features [7].

As, it is well known for any pattern recognition problem, the most crucial problem is the creation of the training set. Usually two approaches are used - local and global training sets. While local training set requires partial annotation of the signal before classification can be carried out – it brings usually better results both in sensitivity and specificity of the classifier. Global training set doesn't require additional annotation but also suffer of worse results in general.

Manuscript received June 30, 2006.

This work has been supported partially by the research program "Information Society" under grant No. 1ET201210527 "Knowledge-based support of diagnostics and prediction in cardiology" and by the joint research project "Intelligent System for Medical Data Analysis and Evaluation Based On Computational Intelligence Techniques" under the Greek-Czech bilateral cooperation agreement, funded by Greek GSRT and Czech MEYS.

V. Chudáček is a PhD. student of Czech Technical University in Prague, Czech Republic, dept. of Cybernetics, Gerstner laboratory. (Contact e-mail: chudacv@fel.cvut.cz).

G. Georgoulas is with Laboratory for Automation and Robotics, University of Patras, 26500 Patras, Greece (Email: georgoul@ee.upatras.gr)

C. Stylios is with the Dept. of Informatics and Communications Technology, TEI of Epirus, Artas, Greece (email: stylios@teiep.gr)

M. Staviař is a student of the Czech Technical University in Prague, Czech Republic, dept. of Cybernetics, Gerstner laboratory.

M. Hanuliak is with the BTL Medical Devices, Prague, Czech Republic

L. Lhotská is an assoc. professor of the Czech Technical University in Prague, Czech Republic, dept. of Cybernetics, Gerstner laboratory.

In this paper we present and compare a number of different methods (decision trees (DT), RBF network (RBF-net), template matching using rule-based decision tree clustering (TM-DTC) and Support Vector Machines (SVM) using local and global training sets as well as global rule-based approach for case of TM-DTC.

II. Data Set

Experimental data used to test the proposed approaches were taken from MIT-BIH Arrhythmia Database [8]. This database consists of 48 2-lead digitized records with annotations of the beat types. Each recording has a duration of 30 min and includes two leads – the modified limb lead II (in all cases but recordings 102 and 104), and one of the modified leads V1, V2, V4 or V5 [12]. The sampling frequency is 360 Hz. Two cardiologists have annotated all beats in the database. About 70% of the beats in the database are annotated as normal. Since we focused only on the discrimination between V and N beats, for classification only 36 recordings were selected - recordings with prevailing paced beats, or beats with permanent blocks were excluded from the set.

For computation reasons of the features, lead II was used on all recordings but recordings 102 and 104 where lead V5 was used.

III. Processing of the signal

Powerline interference, high-frequency electromyographic noise and low-frequency drift was filtered during the preprocessing phase. For this task, well known methods were applied [9].

IV. Feature extraction

Feature extraction stage plays crucial role in any classification task. For this research work, we decided to describe each beat by set of 13 parameters. Feature extraction described in more detail in [9] and it was based on signal basic analysis. which computes not only intervals and amplitudes on each beat, but also description of wave morphology. Extracted parameters that describe the basic shape of the beat are:

- ampR — amplitude of R-peak
- ampS — amplitude of S-peak
- ampQ — amplitude of Q-peak
- ampTp — amplitude of positive peak of T-wave

- ampTn — amplitude of negative peak of T-wave
- ratRT — ratio of amplitudes R-wave : T-wave
- ratRS — ratio of amplitudes R-wave : S-wave
- ratQR — ratio of amplitudes Q-wave : R-wave

The features extracted contain also well-known parameters for distinction of normal beats from the pathologic ones such us:

- intQRS — width of interval
- intQTc — width of QT interval corrected to the heart rate by Fridericia [2] equation:

$$QT_c = \frac{QT}{RR^{\frac{1}{3}}} \tag{1}$$

Features for describing the visual look of the P, QRS, and T complexes are:

- morphP – morphology of the P-wave
- morphQRS – morphology of the QRS-wave
- morphT – morphology of the T-wave

These features describe the morphology of P-wave and T-wave the deflections crossing of the isoline their values:

- One extreme of positive or negative value
- Flat shape of the wave
- Alternating shape with two opposite extremes, negative-positive or positive-negative

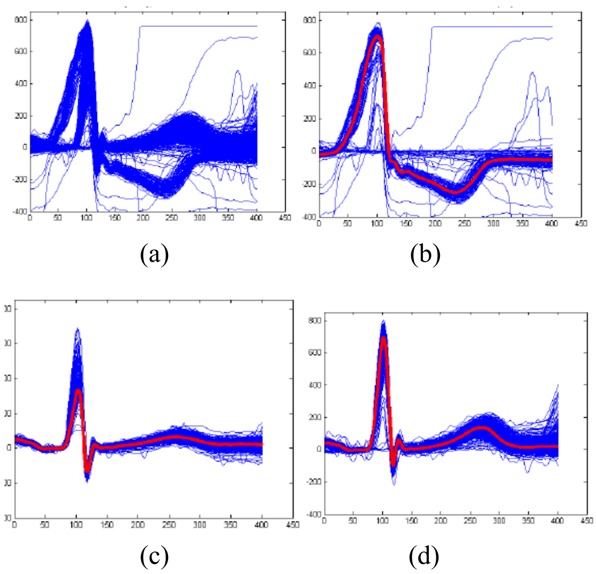

Fig. 1. Example of the clustering algorithm on record 116
a) Original full recording b)c)d) sub-clusters labeled V, N, N
Red line represents the median of the cluster used for future computations

The templates were then compared to the medians of the final clusters using correlation coefficient [14].

The final cluster was labeled either as N or V according to the larger similarity to the appropriate template.

V. Brief Description of Used Methods

A. Rule-based Decision Tree Clustering

Rule-based Decision Tree clustering method is described in detail in [9]. In principle the strength of this method is to decrease the amount of beats taking into consideration in the classification process. The whole 24-hour holter monitoring can have more than 100.000 beats. This method makes unrealizable any further computation involving beat to beat comparison.

Therefore rule-based decision tree was made based on general information about the characteristics of the normal and pathological beats. Using this method, we were able to cluster beats of the 30-min recording to up to 30 classes (with median of 10 classes). Typical results of the decision can be seen on Figure 1.

Then the cluster is represented by the median of the cluster and this median might be used for further classification/diagnosis.

B. Template Matching

Template matching method was used as a second - classification - step after the implementation of the rule-based decision tree clustering. Templates for comparison were computed as a median out of randomly selected ten N and V beats from each signal.

C. Decision Tree using J48 algorithm

Decision tree learning is a common method, mainly applied in the area of data mining. Here it is used the decision tree induction algorithm called J48 and it is an improved version of the C4.5 algorithm [11], which is implemented in the WEKA machine learning platform [10]. J48 constructs a decision tree, which can then be used to classify any unseen data. It generates decision trees consisted of nodes that evaluate the existence or significance of individual features. Following the path from the root to the leaves of the tree, a sequence of such tests is performed, resulting in a decision about the appropriate class of the gastric cell. The decision trees are constructed in a top-down fashion, by choosing the most appropriate attribute each time. The attributes are evaluated according to an information-theoretic criterion, which provides an indication of the "classification power" of each attribute. This criterion is based on information entropy

$$Entropy(S) = \sum_{j=1}^{n} -p_j \log_2 p_j \tag{2}$$

to measure the "quantity" of information required to describe the classification of the items in a dataset *S*, into one of *n* classes, distributed according to probability *pj*. This criterion is called *information gain* and measures the reduction of the entropy, when splitting the dataset according to a characteristic attribute *A*:

$$Gain(S,A) = Entropy(S) - \sum_{u \in Values(A)} \frac{|S_u|}{|S|} Entropy(S_u) \quad (3)$$

where *Values*(*A*) is the set of values that *A* can take and *Su* is the set of vectors for which *A* takes the value *u* Thus, the attribute providing the highest information gain is selected as the "best" discriminator and added to the decision tree. Once an attribute is chosen, the training data are divided into subsets, corresponding to different values of the selected attribute. This process is repeated for each subset, until a large proportion of the instances in each subset belong in a single class.

D. Support Vector Machines (SVM)

Support Vector Machines are learning systems that are trained using an algorithm from optimization theory. The aim of a support vector classifier is to "construct" a good separating hyperplane in a high dimensional feature space. More specifically, given a data set $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ of labeled examples $y_i \in \{-1, 1\}$ and a kernel function *K*, through an optimization process for each $\mathbf{x}_i$ a coefficient a_i such as to maximize the margin between the hyperplane and the closest instances to it is found. Every new pattern **x** is classified to either one of the two categories through

$$f(\mathbf{x}) = sign\left(\sum_{i=1}^{n} y_i a_i K(\mathbf{x}, \mathbf{x}_i) + b\right) \quad (4)$$

where *b* is a threshold parameter. The coefficients a_i are found by solving the following quadratic programming problem of maximizing:

$$\sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i,j=1}^{n} a_i a_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (5)$$

subject to $0 \le a_i \le C$, $i = 1, ..., n$ and $\sum_{i=1}^{n} a_i y_i = 0$

In the above formulation (L1 Soft-Margin SVM) *C* is the margin parameter that determines the trade-off between the minimization of the classification error and the maximization of the margin. Depending on the choice of the kernel function different hyperplanes and different classifiers are constructed. Among the most popular kernels for classification tasks is the Gaussian Radial Basis Function kernel given by:

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right) \quad (6)$$

Even though SVMs were primarily designed for binary classification problems, they can also be used to deal with multi-class classification problems. The most common approaches to create *M*-class classifiers, are the "one versus the rest" and the "pairwise classification" [15]. Taking into account that those methods can give comparable results and generally no multi-class approach outperforms the others, we only considered the pairwise classification (or "one to one") approach.

According to this approach, a classifier is trained for every possible pair of classes. That is, for a problem with *M* classes, it results in (*M*-1)*M*/2 binary classifiers. After the training of the classifiers in order to classify a test pattern we evaluate the output of each one of the classifiers and the pattern is classified to the class that gets the highest number of "votes".

Other researchers have also applied successfully SVM in the same application area [16].

To perform the experimental tests, we used the software package LIBSVM [17] to design and apply SVMs.

E. RBF Network

Radial Basis Functions are powerful techniques for interpolation in multidimensional space. A RBF Neural Network is built into a distance criterion with respect to a centre. RBF networks have 2 layers of processing: In the first, input is mapped onto each RBF in the 'hidden' layer. Usually a Gaussian RBF is chosen. Especially for classification problems the output layer is typically a sigmoid function of a linear combination of hidden layer values, representing a posterior probability [18].

RBF networks have the advantage of not suffering from local minima in the same way as multilayer perceptrons. This is because the only parameters that are adjusted in the learning process are the linear mapping from hidden layer to output layer. Linearity ensures that the error surface is quadratic and therefore has a single easily found minimum.

VI. Training and Testing set

A. Local training set

Local training was built out of 200 beats from the begging of each record. Basically this approach means algorithm is trained on the beginning of the signal and is tested on the rest. Although this approach seems to be quite arbitrary; but it supposed annotation for a part of the signal by cardiologist.

Such an approach may prove to be useful for personalized medicine – where brief annotation in the very beginning can be made (while patients is still in the clinician office) and the accuracy of the classifier can be significantly enhanced.

B. Global training set

Global training set consist of up to 20 N and V beats for each recording with total of 1141 beats in the global training set. This global training set is then used for training of the classifier and then tested on all recordings.

VII. RESULTS

Features were computed first in Matlab and then converted into format suitable for WEKA software [10] where decision tree experiments were conducted as well as RBF-network computations. TM-RBDT clustering and template matching algorithm was implemented in computed in Matlab. For the case of SVM the LIBSVM [17] was used.

The specificity representing the accuracy of the classification of normal QRS complexes, and the sensitivity that represents the accuracy of detecting the V beats. The results for methods described in section V are shown in Table 1.

TABLE I

Results for algorithms Specificity and Sensitivity of decision tree(DT) with local training set, DT with global training set, Rule-Based DT, RBF network.

Type of algorithm	SPECIFICITY [%]	Sensitivity [%]
TM -RBDT	96.63	92.64
DT global training	94.00	91.23
DT local training	99.83	74.08
RBF global training	94.70	84.14
SVM	94.68	90.72

VIII. DISCUSION

We expected our results of decision tree method to be comparable to those of Bortolan et.al [13], which showed significantly worse specificity 75.4%-88.5 % and sensitivity 80.7%-85.8% with global training as compared to the same methods using local training set with specificity 94.4%-98.7% and 91.3%-97% specificity.

On the contrary our decision tree performed well with the global training set and had significantly worse sensitivity with the local set. The reason can be found observing the recordings with worst sensitivity results – 5 recordings had less than 1% of the beats V beats in its local training set.

SVM is shown to be an interesting and promising method and the achieved results are comparable to [17].

The best performance was shown by template matching method using rule-based clustering algorithm.

IX. CONCLUSION

Template matching method using rule-based clustering had yielded best results. There is still place for further development especially with the templates creation and feature selection tasks.

For acquiring better results for the locally trained decision trees there is need for usage of techniques for adjusting the numbers of V and N beats in the training set.

REFERENCES

[1] Holter, N.J. (1961): ‘New methods for heart studies’, Science 134 pp.1214.

[2] L. S. Fridericia: Die Systolendauer im Elektrokardiogramm bei normalen Menschen und bei Herzkranken. Acta Medica Scandinavica, Stockholm, 1920, 57: 469-486.

[3] Daskalov, I.K., Dotsinsky, I.A.; Christov, I.I., Developments in ECG acquisition, preprocessing, parameter measurement, and recording, IEEE Eng. Med. Biol. 17 (1998) 50_/58.

[4] de Chazal, P., O’Dwyer, O., Reilly, R.B. (2004): ‘Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features’, IEEE Transaction on Biomedical Engineering, 51, No. 7

[5] Cuesta-Frau, D., Perez-Cortes, J.C., Andreu-Garcıa, G. (2003): ‘Clustering of electrocardiograph signals in computer-aided Holter analysis’, Computer methods and programs in Biomedicine 72, pp. 179-196

[6] Lagerholm, M., Peterson, C., Braccini, G., Edenbrandt, L., Sornmo, L. (2000): ‘Clustering ECG complexes using hermite functions and self-organizing maps’ IEEE Trans. Biomed. Eng., vol. 47, pp. 838–848

[7] Senhadji L, Carrault G, Bellanger JJ, Passariello G. Comparing wavelet transforms for recognizing cardiac patterns. IEEE Eng Med Bio Mag 1995;14:167–173.

[8] Goldberger AL, Amaral L, Glass L, Hausdorf JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220

[9] Chudáček, V. - Hanuliak, M. - Lhotská, L. Clustering of Heartbeats for Automated ECG Holter Analysis In: The 3rd European Medical and Biological Engineering Conference - EMBEC´05 [CD-ROM]. Praha: ČLS JEP, 2005, vol. 11, ISSN 1727-1983.

[10] Quinlan, J. R.: C4.5: Programs for machine learning, Morgan – Kaufmann 1993.

[11] Witten I., Frank E., Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, 1999.

[12] Mark R, Moody G, MIT-BIH Arrhytmia data base directory. Cambridge: MIT, 1988

[13] Bortolan G, Jekova I , Christov I Comparison of Four Methods for Premature Ventricular Contraction and Normal Beat Clustering Computers in Cardiology 2005;32:921−924.

[14] Chiu C., Lin T. Liau B. Using Correlation Coefficient in ECG waveform for arrhythmia detection, Biomedical Engineering applications, Basis & Communication, Taiwan

[15] Scholkopf B, Smola A,. Learning With Kernels. Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, England pp. 211-214, 2002.

[16] Chang C, Lin C,. “LIBSVM: A Library For Support Vector Machines,” 2003.

[17] Tsipouras M, Voglis C, Lagaris I, Fotiadis D,. Cardiac Arrhuthmia Classification using Support Vector Machines. In: The 3rd European Medical and Biological Engineering Conference EMBEC’05, [CD-ROM]. Prague: ČLS JEP, 2005, vol. 11, ISSN 1727-1983.

[18] Buhmann M, Ablowitz M,. Radial Basis Functions : Theory and Implementations. Cambridge University 2003