CLASSIFICATION OF ECG SIGNALS USING WAVELET TRANSFORM AND HIDDEN MARKOV MODELS

D. Kičmerová, I. Provazník and J. Bardoňová

Faculty of Electrical Engineering and Communication, Department of Biomedical Engineering, Brno, Czech Republic

provazni@feec.vutbr.cz

Abstract: The automatic detection and classification of cardiac arrhythmias is important for diagnosis of cardiac abnormalities. This work is focused on classification of normal sinus rhythm and premature ventricular contractions. The proposed method uses wavelets for the feature selection and extraction (searching for a local maximum in the contour envelope successfully detects R-peaks) and Hidden Markov models for the classification. The ECG data is taken from standard MIT-BIH arrhythmia database. The classification accuracy is 95.6% for normal sinus rhythm beats and 93.3% for premature ventricular contraction beats.

Introduction

The medical domain is one of the areas in which automatic processing systems are most frequently applied. This is quite natural because modern medicine generates huge amounts of data, but at the same time there is often a lack of data understanding. New methods can help in dealing with this problem, they can simplify and usually speed up the processing of large volumes of data. New algorithms work in time-frequency domain and combine some advantageous characteristics known from classical methods – mainly they allow frequency analysis with time information about analyzed features.

The automated detection and classification of cardiac arrhythmias is important for diagnosis of cardiac abnormalities. Our previous work [1] was focused to detection of atrial premature contractions (APC) and premature ventricular contractions (PVC) among normal sinus rhythm (NSR). The used methods employed wavelets and a contour envelope computed from wavelet coefficients. Searching for a local maximum in the contour envelope detected R-waves in all the above mentioned types of heart cycles. The overall accuracy of the detection tested on 48 half-hour signals from MIT-BIH library was not less than 99.5 %. Using the same method, classification of APC and PVC was tested with overall accuracy of 94.6%.

The above methods were dependent on signal characteristics. In practice, unsupervised methods may be required to detect arrhythmias in changing environment. Hidden Markov models are often used for such tasks in speech recognition [3] but also in ECG processing. To obtain improved results, HMM can be combined with wavelet transform [2].

The ECG waveform and PVC's

Each individual heartbeat is comprised of a number of distinct cardiological stages, which in turn give rise to a set of distinct features in the ECG waveform. These features represent either depolarization (electrical discharging) or repolarization (electrical recharging) of the muscle cells in particular regions of the heart. Figure 1 shows a human ECG waveform and the associated features. The standard features of the ECG waveform are the P wave, the QRS complex and the T wave. Additionally a small U wave (following the T wave) is occasionally present.

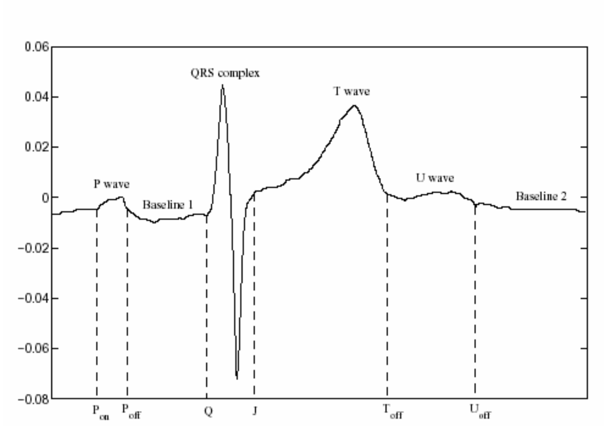

Figure 1: ECG waveform

The cardiac cycle begins with the P, which corresponds to the period of atrial depolarization in the heart. This is followed by the QRS complex, which is generally the most recognisable feature of an ECG waveform, and corresponds to the period of ventricular depolarization. The start and end points of the QRS complex are referred to as the Q and J points. The T wave follows the QRS complex and corresponds to the period of ventricular repolarization.

Premature ventricular contractions (PVC's) - these early depolarizations begin in the ventricle instead of the usual place, the sinus node. They are very common,

ISSN: 1727-1983

and are sometimes perceived as a palpitation. They often occur without the patient being aware of it at all. PVC's occure in Bigeminy, Trigeminy, Quadrigeminy, Ventricular tachycardia, Ventricular fibrillation, etc.

An increased frequency of PVC's in patients with heart disease is statistically predictive of ventricular fibrillation and sudden death. In patients with some types of heart disease, PVC's or ventricular tachycardia do indicate an increased risk of serious arrhythmias. Therefore this work is focused on their detection.

Continuous wavelet transform

In the proposed method, input data were transformed by continuous wavelet transform (CWT):

$$CWT(a,b)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}\psi^{*}\left(\frac{t-b}{a}\right)f(t)dt \qquad (1)$$

where a is a scale and b is a time shift. Time–frequency spectrum enables to measure time–frequency changes in spectral components. Interpretation of a time-frequency resolution by CWT is following: CWT represents time-frequency decomposition realized by correlation of signal f(t) with basic functions derived from the mother wave ψ(t). Haar function was used as the mother wavelet ψ.

$$\psi_{H}(t)=\begin{cases}1 & 0\le t<\frac{1}{2}\\ -1 & \frac{1}{2}\le t<1\\ 0 & others\ ,\end{cases} \qquad (2)$$

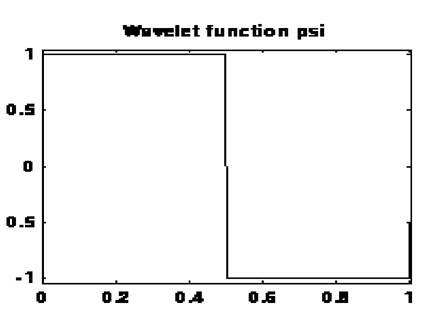

Figure 2: Haar function

Scales were chosen from an interval of <1; 40>, signal f was extracted as a period of 100 samples (f_s=360 Hz) around the R-wave of each QRS complex. Wavelet coefficients were squared and normalized.

MacQueen algorithm

The transformed input data were supplied to the discrete density HMMs (DD-HMM) in vector quantized form, where MacQueen algorithm (MQA) was used for the creation of the codebook from transformed data.

This iteration algorithm is provided in cycles in following steps:

1. step: Stochastic selection of L initial centroides $\mathbf{v}_1(1)$, $\mathbf{v}_2(1), \ldots, \mathbf{v}_L(1)$.

If we know prior information about solved problem, we can adapt our initial choice.

2. step: The division of all vectors of the training set T to L aggregates $T_1(k), \ldots, T_L(k)$ using relation:

$$\mathbf{x}\in T_j(k)\ \texttt{if}\ d(\mathbf{x},\mathbf{v}_j(k))\le d(\mathbf{x},\mathbf{v}_i(k)),\quad i,j=1,\ldots,L,\ i\ne j \qquad (3)$$

3. step: Computation of new centroides $\mathbf{v}_j(k+1)$ for all aggregates $T_1(k), \ldots, T_L(k)$ which minimalizes criteria:

$$J_j(k+1)=\sum_{\mathrm{x}\in \mathrm{Tj(k)}} d(\mathrm{x},\mathrm{v}_j(k+1)),\ j=1,\ldots,L, \qquad (4)$$

is determined by equation:

$$\mathrm{v}_j(k+1)=\frac{1}{n_j(k)}\sum_{\mathrm{x}\in \mathrm{Tj(k)}}\mathrm{x},\ j=1,\ldots,L, \qquad (5)$$

The condition of the end of algorithm: If the following conditions **a)** or **b)** are fulfilled, algorithm is finished:

a)

$$\mathrm{v}_j(k+1)=\mathrm{v}_j(k)\ \text{for all}\ j=1,\ldots,L. \qquad (6)$$

b) Decrease of total distortion $J(k)$ is in k-th iteration in relation with $J(k\text{-}1)$ under defined threshold, where

$$J(k)=\sum_{i=1}^{L}J_i(k). \qquad (7)$$

In another case it continues with the second step, with new distribution of trained set to aggregates.

$T_i(k)$ is the set of vectors from the i-th aggregate in the k-th step, $\mathbf{v}_i(k)$ is the centroide of the i-th aggregate in the k-th step, $J_i(k)$ is a value of criterion in the i-th aggregate in the k-th step, and $n_i(k)$ is the number of vectors $\boldsymbol{x}$ in aggregate T_i in k-th step [10].

The discrete density Hidden Markov model (DD-HMM)

The Hidden Markov model is a finite state machine having a set of states Q, each of which is associated with probability distribution, an output alphabet O, transition probabilities A, output probabilities B, and initial state probabilities Π. The current state is not observable. Instead, each state produces an output with a certain probability B. The DD-HMM stage is proceeded by the pre-processing steps (CWT and MacQueen algorithm) called parameter extraction. Thus, the input to the DD-HMM is a discrete time sequence of parameter vectors quantized with codebook from transformed raw time data series.

DD-HMM combined with CWT was designed to classify PVC and NSR cycles in long-term ECG recordings. The used DD-HMM had left-to-right topology. Number of states was 10. The first state is designated as the initial state and the last state as the output state.

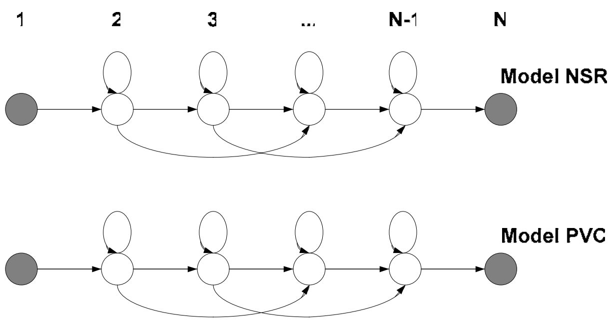

Figure 3: Structure of the used hidden Markov model.

Three problems are considered fundamental in hidden Markov modelling applications:

1) Estimation of hidden Markov model parameters from a set of representative training data (parameters include state transition probabilities, output probabilities).

2) Efficient calculation of P (O|λ) - the probability that a given observation sequence was generated by a particular hidden Markov model λ.

3) Determination of X*, the most likely underlying state sequence corresponding to a given observation sequence 0 such that P (0| X*,λ) is maximum.

The importance of solving the first problem is obvious; model parameters must first be estimated before the models can be used for classification purposes.

Baum-Welch algorithm [3],[10] was used as a training method to find hidden Markov model parameters A, B, and Π with the maximum likelihood of generating the given symbol sequence in the observation vector.

To determine the parameters of a DD-HMM it is first necessary to make a rough guess at what they might be.

Initial parameters were:

$$B=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0 \\ 0 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0 \\ \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad (8)$$

$$A=\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.33 & 0.33 & 0.33 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.33 & 0.33 & 0.33 & 0 & 0 & 0 & 0 & 0 \\ \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (9)$$

$$\Pi=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (10)$$

The probability of state occupation must be calculated. This is done efficiently using the so-called Forward-Backward algorithm. The number of iterations was 10 and the number of stages was 12.

The solution to the second problem is often used as the basis for a classification system. By computing P (O|λ) for each of *i* possible models and choosing the most likely model, classification can be inferred.

An alternative classification approach uses the solution of the third problem to find the single best state sequence which maximizes P (0| X*,λ). Classification can then be inferred by choosing the model with the most likely best state sequence, which requires less computation than determining the most likely model.

Logarithm Viterbi algorithm was used for the recognizing. It determines the most probable route to the next state, and remembers how to get there. This is done by considering all products of transition probabilities with the maximal probabilities already derived for the preceding step. The largest such is remembered, together with what provoked it.

Scaling the computation of Viterbi algorithm to avoid underflow is non-trivial. However, by simply computing of the logarithm it is possible to avoid any numerical problems.

Results

The ECG data for experiments were taken from standard MIT-BIH arrhythmia database, lead MLII. Training and testing data was prepared automatically as a sequence of 100 samples around R wave, which was detected using CWT and the method of contour envelope. Transformed data (Figure 4) after vector quantization were processed by DD-HMMs. The difference in time-frequency domain between PVC's and NSR is documented on the picture below.

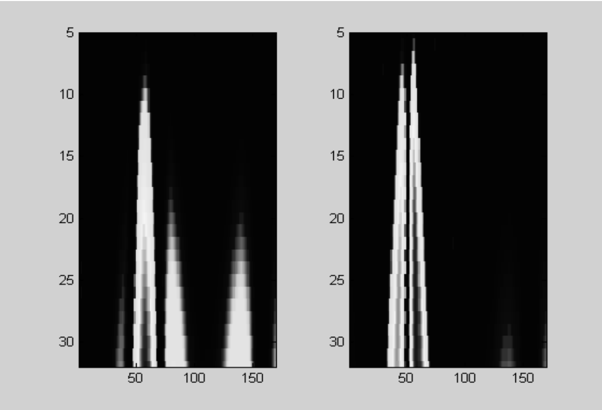

Figure 4: Continuous wavelet transform of the raw time data series (left panel – premature ventricular contraction, right panel – normal sinus rhythm).

Set of training sequence consisted of 180 beats (PVCs from signals 106, 119, 203, and NSRs from signals 100, 106, 115). Testing set consisted of different signals with frequent PVCs (200, 201, 208, 214) and signals with frequent NSRs (101, 112, 114, 116, 117, 119, 201, 208). Total number of tested beats was 15330. Algorithms were designed in Matlab environment using statistical toolbox. Results are summarized in the table below.

Table 1: Detection of premature ventricular contraction and normal sinus rhythm beats.

	Number of beats	Accuracy(%)
PVC	1583	93.30
NSR	13747	95.65

Accuracy of classification (%) is defined as:

$$\frac{\text{number of correctly classified beats}_{NSR,PVC}}{\text{total number of beats}_{NSR,PVC}} * 1(\qquad (11)$$

Discussion and Conclusions

An algorithm employing unsupervised way of classification was proposed. This work is focused on classification of normal sinus rhythm and premature ventricular contractions. There is demonstrated that wavelet methods can be used to generate an encoding of the ECG which is tuned to the unique spectral characteristics of the ECG waveform features. With this pre-processing step the performance of the models is significantly better than models trained on the raw time series data.

Left-to-right 10-state DD-HMM and pre-processing of data with CWT using Haar wavelet was used. The algorithm was tested to distinguish between NSR and PVC. In the testing phase, ECG signals are classified using the trained models. The classification accuracy is 95.6% for NSR beat, 93.3% for PVC beat. Classification of PVC's suggests the algorithm could exceed results of recent systems. The popularity of hidden Markov modelling in speech recognition applications has led to the development of special purpose hardware implementations for Viterbi algorithm computation [6]-[7]. Using such hardware, an arrhythmia analysis system based on hidden Markov modelling could be operated 60 times faster than the recording speed.

Acknowledgement

The research was supported by the grants 102/04/0472 and 305/04/1385 from GACR, 1596/2004 from FRVS, and Research Programme of Brno University of Technology MSM 0021630513.

References

[1] KIČMEROVÁ D. (2004): 'Detection and classification of ECG signals in time-frequency domain', Diploma Thesis, Brno University of Technology.

[2] BARDOŇOVÁ J., PROVAZNÍK I., NOVÁKOVÁ M., NOVÁKOVÁ Z. (2000): 'Hidden Markov model in wavelet analysis of myocardial ischemia in rabbit', *Computers in Cardiology*, p. 419–422, Boston.

[3] RABINER L R. (1989): 'A tutorial on hidden Markov models and selected applications in speech recognition', *Proceedings of the IEEE*, 77, pp. 257–286.

[4] HUGHES N. P., TARASSENKO L., ROBERTS S. J. (2003), 'Markov Models for Automated ECG Interval Analysis', *Proceedings of the NIPS*, Oxford.

[5] COAST A D., STERN R. M. (1990): 'An Approach to Cardiac Arrhythmia Analysis Using Hidden Markov Models', *IEEE Transactions on Biomed. Eng.,* 37, pp. 826-836.

[6] GLINSKI S. C., LALUMIA T. M., CASSIDAY D. R., KOH T., GERVESHI C., WILSON G. A., KUMAR J. (1987): The graph search machine (GSM): A VLSI architecture for connected speech recognition and other applications. *Proceedings of the IEEE,* 75, pp. 1172-1184.

[7] BISIANI R., ANANTHARAMAN T., BUTCHER L. (1989): 'BEAM: An accelerator for speech recognition', Proceedings International Conf. Acoustics, Speech. and Signal Process., Glasgow, Scotland, pp. 217-220.

[8] LOU H.-L. (1995): 'Implementing the Viterbi Algorithm', *IEEE Signal Processing Magazine*, pp. 42-52.

[9] CHENG W. T., CHAN K. L. (1998): 'Classification electrocardiogram using Hidden Markov models', Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 143-146.

[10] PSUTKA, J. (1995): '*Komunikace s počítačem mluvenou řečí.* Academia, Praha.