ECG signal,
fuzzy system,
bayesian learning

Michał MOMOT*, Alina MOMOT**

CHARACTERISTIC POINTS DETECTION IN ECG SIGNAL USING BAYESIAN LEARNING AND FUZZY SYSTEM

Characteristic points detection such as beginnings and ends of P-wave, T-wave or QRS complex is one of primary aims in automated analysis of ECG signal. The paper presents one possible approach based on Bayesian inference to design of kernel based classifier. The classification function is constructed using the probability distribution function of standard normal distribution and independent Gaussian random variables. The parameters of such variables are computed using iterative Expectation-Maximization algorithm. This approach is used to calculate parameters of classification function to modelling Takagi-Sugeno-Kang fuzzy systems. Numerical experiment of characteristic points detection in ECG signal using CTS database is also presented.

1. INTRODUCTION

The formulation and properties of an electrical impulse through the heart muscle result in time-varying potentials on the surface of the human body, which are known as the ECG signals. The signal represents various activities of the heart. A typical ECG signal is indicated in Figure 1. As can be seen from it, the dominant morphologies are the P, QRS and T waves. Occasionally a U-wave will be present immediately after the T-wave, the genesis of which is uncertain. The P-wave represents atria activation; the QRS complex represents ventricular activation or depolarisation. An initial downward deflection after the P-wave is termed as 'Q', the dominant upward deflection is 'R' and the terminal part is denoted as 'S'. The T-wave represents ventricular recovery or depolarisation. The ST segment, the T-wave and the U-wave together represent the total duration of ventricular recovery. The ST segment represents the greater part of ventricular depolarisation. The ST segment usually merges smoothly and imperceptibly with the T-wave [4].

The graph of ECG signal, known as the electrocardiogram, shows the results of nerve impulse stimuli by the heart, as the current is diffused around the surface of the body. The current at the body surface will be built on the voltage drop, which is a couple of μV to mV with impulse variations. This is very small amplitude of impulse that needs to be amplified to an amount, which is large enough for recording and displaying. Usually, an electrocardiograph requires an amplification of a couple of thousand times.

* Institute of Medical Technology and Equipment, 118 Roosevelt St., 41-800 Zabrze, Poland
** Silesian University of Technology, Institute of Computer Science, 16 Akademicka St., 44-101 Gliwice, Poland

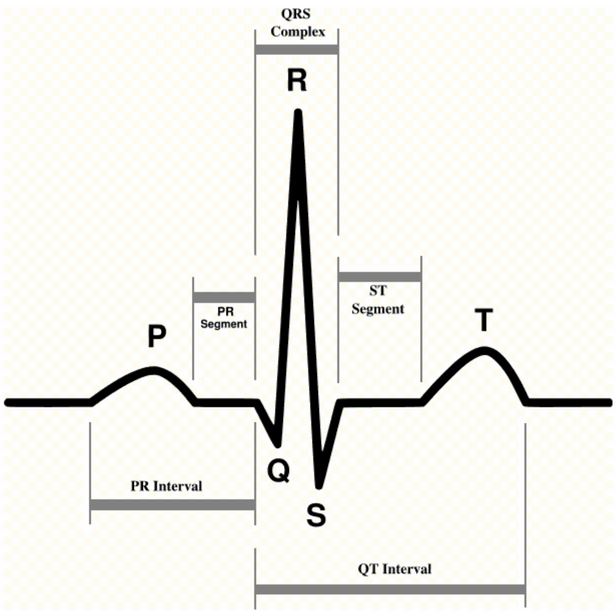

Fig.1. Schematic representation of normal ECG

The electrocardiogram is a useful, simple tool that may aid in the diagnosis of heart diseases. Many kinds of abnormalities can often be seen on an ECG. They include a previous myocardial infarction, an abnormal heart rhythm (arrhythmia), an inadequate supply of blood and oxygen to the heart (ischemia), and excessive thickening (hypertrophy) of the heart's muscular walls. Certain abnormalities on ECG can also suggest aneurysms that develop in weak areas of the heart's walls. Aneurysms may result from a heart attack. If the rhythm is abnormal (too fast, too slow, or irregular), the ECG may also indicate where in the heart the abnormal rhythm starts.

Typical ECG findings include diffuse concave-upward ST-segment elevation and, occasionally, PR-segment depression. In order to perform these measurements it is necessary to locate characteristic points such as beginnings and ends of P-wave, T-wave or QRS complex on the timeline. However, this is difficult task because of the presence of noise.

Usually in case of electrocardiographic (ECG) signal, two principal sources of noise can be distinguished: the ‘technical’ caused by the physical parameters of the recording equipment and the ‘physiological’ representing the bioelectrical activity of living cells not belonging to the area of diagnostic interest (also called background activity). Both sources produce noise of random occurrence, overlapping the ECG signal in both time and frequency domains [1].

Moreover detection of characteristic points such as beginnings and ends of P-wave, T-wave or QRS complex is difficult because the ECG waveform is a non-stationary signal, even when observed in a perfectly healthy normal subject. These non-stationarities are severe in case of abnormal subjects due to the association of transient phenomena.

The paper presents the theoretical approach to the problem of characteristic points detection which incorporates Bayesian inference to design of kernel based classifier. The classification function is constructed using the probability distribution function of standard normal distribution and independent Gaussian random variables. The parameters of such variables are computed using iterative Expectation-Maximization algorithm. This approach is used to calculate parameters of classification function to modelling Takagi-Sugeno-Kang fuzzy systems[5], [6]. The paper also presents numerical experiment of characteristic points detection in ECG signal using CTS database [8].

2. DETECTION ALGORITHM

2.1. CLASSIFIER DESIGN METHOD

The classification task aims at inferring a functional relation $f : \mathrm{X} \to \mathrm{Y}$ between numerical input data and categorical output values. The design of classifier is based on finite training set $T = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), ..., (\boldsymbol{x}_N, y_N)\}$. Usually the inputs are *d*-dimensional real vectors, $\boldsymbol{x} \in \Re^d$ and outputs might be integer values, representing class labels. Usually the function *f* is assumed to have a fixed structure and to depend on a vector of parameters $\boldsymbol{\beta}$. In this case the goal becomes to estimate the parameters from the training data.

In this paper the two-class case will be taken into account, hence $\mathrm{Y} = \{0,1\}$, and the classification is based on function of form

$$f_\beta(x) = \Phi\left(\beta_0 + \sum_{i=1}^{N} \beta_i h_i(\boldsymbol{x})\right), \tag{1}$$

where Φ is the cumulative distribution function of standard Gaussian distribution $N(0,1)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{t^2}{2}) dt \tag{2}$$

and $\boldsymbol{h}(\boldsymbol{x}) = (h_1(\boldsymbol{x}), h_2(\boldsymbol{x}), \ldots, h_N(\boldsymbol{x}))^T$ is vector of fixed base functions. This is called generalized linear model, as in [2], [3].

The main goal of the classifier design procedure is to achieve high generalization ability [7], as to avoid over-fitting to training data, but still to be able to capture main behaviour of input-output relationship. This may be obtained by controlling complexity of learned function, using the variety of tools. In the latter part of paper the Bayesian approach to classifier design will be presented, in order to find sparse solutions (having only a few non-zero coefficients), which lead to good generalization ability. As the base functions, the values of kernel functions will be used:

$$\boldsymbol{h}(\boldsymbol{x}) = (1, K_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{x}_1), K_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{x}_2), \ldots, K_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{x}_N))^T. \tag{3}$$

As in [3], the classification rule is defined as:

$$P(y = 1 \mid x) = \Phi\left(\beta_0 + \sum_{i=1}^{N} \beta_i K_\theta(\boldsymbol{x}, \boldsymbol{x}_i)\right) \tag{4}$$

and $P(y=0\,|\,x)=1-P(y=1\,|\,x)$. The consequence of classification function form is the need of setting values of parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. The estimates of β_i are evaluated using Bayesian inference. The Laplace distribution with parameter λ is taken as a common prior distribution of β_i and the parameters are assumed to be independent. This leads to learning procedure, where the posterior probability of correct classification is maximized. The vector $\boldsymbol{z}=(z_1,\ldots,z_N)$ of hidden variables is introduced:

$$z_j = w_j + \beta_0 + \sum_{i=1}^{N} \beta_i K_\theta(\boldsymbol{x}_j, \boldsymbol{x}_i) \tag{5}$$

where w_j are independent Gaussian variables $N(0,\sigma_j)$. This is generalization of model described in [2], where all variables w_j have common standard deviation. The estimation of parameters β_i is performed by Expectation-Maximization procedure. In the E-step the expected value of $\boldsymbol{\beta}$ is computed:

$$Q(\boldsymbol{\beta}\,|\,\hat{\boldsymbol{\beta}}^t) = \int p(\boldsymbol{z}\,|\,\boldsymbol{y},\hat{\boldsymbol{\beta}}^t)\log p(\boldsymbol{\beta}\,|\,\boldsymbol{z},\boldsymbol{y})\,dz \tag{6}$$

(where the upper index denotes successive iteration number) and next, in the M-step the value of $Q(\boldsymbol{\beta}\,|\,\hat{\boldsymbol{\beta}}^t)$ is maximized with respect to $\hat{\boldsymbol{\beta}}^t$:

$$\hat{\boldsymbol{\beta}}^{t+1} = \arg\max_{\beta} Q(\boldsymbol{\beta}\,|\,\hat{\boldsymbol{\beta}}^t)\,. \tag{7}$$

This leads to following iterative procedure:

$$\boldsymbol{v}^t = E(\boldsymbol{z}\,|\,\hat{\boldsymbol{\beta}}^t, \boldsymbol{y})\,, \tag{8}$$

$$\hat{\boldsymbol{\beta}}^{t+1} = (\boldsymbol{H}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{H}+\boldsymbol{A}^{-1})^{-1}\boldsymbol{H}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v}^t\,, \tag{9}$$

where $\boldsymbol{H}=(h(x_1),h(x_2),\ldots,h(x_N))$, $\Sigma = diag(\sigma_1^2,\sigma_2^2,\ldots,\sigma_N^2)$, and $\boldsymbol{A}$ is (diagonal) covariance matrix of prior distribution of $\boldsymbol{\beta}$. The iteration stops as

$$\frac{\left\|\hat{\boldsymbol{\beta}}^{t+1}-\hat{\boldsymbol{\beta}}^t\right\|}{\left\|\hat{\boldsymbol{\beta}}^t\right\|} < \varepsilon \tag{10}$$

for arbitrarily chosen value of ε.

2.2. MODELLING TAKAGI-SUGENO-KANG SYSTEM

The procedure described in the previous section may be applied to modelling Takagi-Sugeno-Kang (TSK) fuzzy system [5], [6]. In this case, the set of input variables $\boldsymbol{x} \in \Re^d$ is clustered using fuzzy *c*-mean clustering assuming that each of *c* clusters corresponds to a fuzzy if-then rule in the TSK system. For each cluster the classifier is designed using input and output data and the overall output of TSK system is computed as aggregation of outputs of individual classifiers. The values of σ_j for each classifier can be established by following formula:

$$\sigma_j^2 = \left(A^{(i)}(x_j)\right)^{-p} \qquad \forall j \in \{1,2,...,N\} \;\; \forall i \in \{1,2,...,c\}, \tag{11}$$

where $A^{(i)}(x_j)$ is the membership of input value x_j in ith cluster and $p \in (0,+\infty)$ is the parameter determining the influence of this membership on uncertainty about the single input data. The output of TSK system is still interpreted as a posterior probability of $\boldsymbol{x}$ belonging to class 1.

2.3. APPLICATION TO CHARACTERISTIC POINTS DETECTION

The algorithm described above can be applied to detection of characteristic points such as beginnings and ends of P-wave, T-wave or QRS complex on the timeline of ECG signal. The main idea of using this classification procedure is to decide whether or not, does the investigated sample appear to be the specific characteristic point. In this case the input vectors are formed from the finite time window around this sample:

$$\boldsymbol{x}_n = (u(n-M),\ldots,u(n),\ldots,u(n+M)), \tag{12}$$

where $u(n)$ is time series representing digital ECG signal in single channel and M determines radius of time window.

3. NUMERICAL EXPERIMENTS

The experiment was performed using data from CTS database of electrocardiographic signals [8]. Only signal from first channel was used and interfered Gaussian noise with zero-mean and signal to noise ratio equal 2.5. The radius M of time window was set to 25 and the polynomial kernel function was chosen:

$$K_\theta(x,x_i) = \left(1 + \sum_{k=1}^{d} \theta_k x_i^{(k)} x^{(k)}\right)^r, \tag{13}$$

with r=1 and $\theta_1 = \theta_2 = \ldots = \theta_d = 1$. The parameters λ (Laplace distribution parameter) and p (the parameter determining the influence of the membership on uncertainty about the single input data) was determined during the learning phase using cross-validation method.

The experiment was performed individually for signals ANE20000, ANE20001, ANE20002 from CTS database. In each case the learning set and test set contained the same signals disturbed by white Gaussian noise independent from each other. In the learning phase of the experiment classifier was trained on single ECG signal from database disturbed by noise. Such constructed classifier was used to locate characteristic points on the same single ECG signal disturbed by noise independent from the one in the learning phase. For the single ECG signal the experiment was repeated 200 times and the results was averaged. Performance of the algorithm was empirically estimated by averaged durations of the P-wave and QRS complex respectively, based on beginnings and endings determined on test set. Table 1 presents results of experiment.

Table 1. Results of experiment

		P-wave duration [ms]	QRS complex duration [ms]
ANE20000	reference value	126.0	94.0
	computed value	107.8	98.4
ANE20001	reference value	142.0	94.0
	computed value	117.4	98.2
ANE20002	reference value	102.0	94.0
	computed value	97.2	100.8

4. CONCLUSIONS

The paper presents the theoretical approach to the problem of characteristic points detection which incorporates Bayesian inference to design of kernel based classifier for modelling Takagi-Sugeno-Kang fuzzy systems. As can be seen in Table 1, the results of numerical experiments show usefulness of the presented method for the characteristic points detection in ECG signal, because the computed values are similar to the reference value.

It seems desirable performing experiments aiming to compare performance of the algorithm using other kernel functions as well as to investigate the influence of the M parameter, which determines dimension of input vector. The interesting question also is the behaviour of the method in the presence of different type of noise, including impulse disturbances. It is worth noting that this method only uses information from single channel of ECG signal. That is why the generalized version of this approach which incorporates all channels will be developed in future.

This research was supported by the Polish Ministry of Education and Science as the research project „Wiarygodne metody wyznaczania punktów charakterystycznych sygnału EKG w trakcie elektrokardiograficznej próby wysiłkowej" (1603/T11/2005/29).

BIBLIOGRAPHY

[1] AUGUSTYNIAK P., Time–frequency modelling and discrimination of noise in the electrocardiogram, Physiological Measurement, 24, pp. 1–15, 2003.
[2] FIGUEIREDO M., Adaptive Sparseness for Supervised Learning, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 25, No. 9, pp. 1150-1159, 2003.
[3] FIGUEIREDO M., JAIN A.K., Bayesian Learning of Sparse Classifiers, IEEE Computer Society Conference on Computer Vision and Pattern Recognition - CVPR'2001, Hawaii, 2001.
[4] SCHAMROTH L., An introduction to electrocardiography. Oxford Press, 1986.
[5] SUGENO M., KANG G.T., Structure identification of fuzzy model, Fuzzy Sets and Systems, Vol. 28, pp. 15-33, 1988.
[6] TAKAGI T., SUGENO M., Fuzzy identification of systems and its application to modelling and control, IEEE Trans. on System, Man and Cybernetics, Vol. 15, No. 1, pp. 116-132, 1985.
[7] VAPNIK V.N., The nature of statistical learning theory, Springer, New York, 1995.
[8] International Electrotechnical Commission Standard 60601-3-2, 1999.