ECG signal, Bayesian classifier design,
Expectation-Maximization, Takagi-Sugeno-Kang system

Michał MOMOT [*], Alina MOMOT [**]

BAYESIAN APPROACH TO CLASSIFIER DESIGN WITH APPLICATION TO QRS DETECTION IN ECG SIGNAL

The paper presents application of Bayesian approach 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. The paper presents also application of algorithm of computation parameters of classification function to modeling Takagi-Sugeno-Kang fuzzy systems. Finaly the application to detection of QRS cycles in ECG signal is presented, with the results of numerical experiment using AHA databese.

1. INTRODUCTION

The classfication 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 $T = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_N, y_N)\}$. Usually the inputs are *d*-dimensional real vectors, $\mathbf{x} \in \Re^d$ 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

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

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

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

[*] 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

and $\mathbf{h}(\mathbf{x})=(h_1(\mathbf{x}),h_2(\mathbf{x}),\ldots,h_N(\mathbf{x}))^T$ is vector of fixed base functions. This is called generalized linear model, as in [1], [2]. The main goal of the classifier design procedure is to achieve high generalization ability [5], 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 desing 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 funcions will be used:

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

2. CLASSIFIER DESIGN METHOD

As in [2], the classification rule is defined as:

$$P(y=1\,|\,x)=\Phi\left(\beta_0+\sum_{i=1}^{N}\beta_i K\ (\mathbf{x},\mathbf{x}_i)\right), \tag{4}$$

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 $\mathbf{z}=(z_1,\ldots,z_N)$ of hidden variables is introduced:

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

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:

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

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

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

This leads to following iterative procedure:

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

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

where $\mathbf{H} = (h(x_1), h(x_2), \ldots, h(x_N))$, $\Sigma = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$, 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.

3. MODELING TAKAGI-SUGENO-KANG SYSTEM

The procedure described in the previous section may be applied to modeling Takagi-Sugeno-Kang (TSK) fuzzy system [3], [4]. In this case, the set of input variables $\mathbf{x} \in \Re^d$ 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:

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

where $A^{(i)}(x_n)$ is the membership of input value x_n in *i*th cluster and *p* 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 $\mathbf{x}$ belonging to class 1.

4. APPLICATION TO DETECTION QRS CYCLES

The algorithm described above was aplied to detection of QRS cycles in ECG signal. The main idea of using this classification procedure is to decide whether or not, does the investigated sample appear to be the center (fiducial point) of some QRS cycle. In this case the input vectors are formed from the finite time window around this sample:

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

where $u(n)$ is time series representing digital ECG signal in single channel. The experiment was performed using data from AHA (American Heart Association) database of electrocardiographic signals. The database contains 80 two-dimensional time series representing two-channel ECG signals with sampling rate 250Hz. Only signal from first channel was used and the width M of time window was set to 12. The polynomial kernel function was chosen:

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

with r=1 and $\theta_1 = \theta_2 = \ldots = \theta_d = 1$. The parameters λ and p was determined during the learning phase using cross-validation method. The experiment was performed individually for each of 80 signals in AHA database. In each case the learning set consisted of samples in first 30 QRS cycles. The rest of samples was the test set. Table 1 presents results of experiment in form of sensitivity ($\frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ \ false\ negatives}$) and positive predictability ($\frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ \ false\ positives}$).

Signal index	Sensitivity (%)	Positive predictability (%)	Signal index	Sensitivity (%)	Positive predictability (%)
1201	99.13	98.59	5201	98.79	97.90
1202	99.07	98.37	5202	98.67	98.30
1203	99.00	98.46	5203	99.01	98.01
1204	99.06	98.33	5204	99.04	98.21
1205	99.01	98.90	5205	99.13	97.99
1206	99.07	98.34	5206	98.90	98.20
1207	99.34	98.15	5207	98.78	98.04
1208	99.37	98.41	5208	99.20	98.30
1209	99.35	98.83	5209	99.01	98.26
1210	99.02	98.66	5210	98.89	98.17
2201	98.79	98.48	6201	98.98	97.67
2202	99.01	98.45	6202	98.79	97.48
2203	99.05	98.46	6203	98.86	97.94
2204	98.89	98.09	6204	98.84	97.69
2205	98.99	98.55	6205	98.66	97.88
2206	99.04	98.47	6206	98.88	98.03
2207	99.54	98.22	6207	98.67	97.57
2208	99.17	98.47	6208	98.57	97.56
2209	99.43	98.40	6209	98.56	97.37

2210	99.12	98.34	6210	98.46	97.50
3201	99.17	98.26	7201	98.15	97.03
3202	99.21	98.44	7202	98.59	97.33
3203	99.45	98.55	7203	98.36	97.20
3204	98.67	98.33	7204	98.56	97.40
3205	98.97	98.29	7205	98.36	97.54
3206	99.14	98.49	7206	98.35	97.12
3207	99.15	98.70	7207	98.55	97.03
3208	99.23	98.50	7208	98.12	97.14
3209	99.42	98.39	7209	98.14	97.32
3210	99.47	98.69	7210	98.56	97.31
4201	98.87	98.40	8201	97.99	97.00
4202	99.24	98.50	8202	98.45	97.05
4203	99.34	98.66	8203	98.88	97.12
420	99.56	98.36	8204	98.45	97.02
4205	99.34	98.29	8205	98.33	97.40
4206	99.34	98.40	8206	98.67	97.23
4207	99.45	98.35	8207	98.37	97.22
4208	99.56	98.50	8208	97.82	97.03
4209	99.45	98.20	8209	98.01	97.09
4210	99.15	98.31	8210	97.50	97.03

Tab. 1 Results of QRS detection experiment on AHA database

BIBLIOGRAPHY

[1] FIGUEIREDO M., Adaptive Sparseness for Supervised Learning, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 25, No. 9, pp. 1150-1159, 2003.
[2] FIGUEIREDO M., JAIN A.K., Bayesian Learning of Sparse Classifiers, IEEE Computer Society Conference on Computer Vision and Pattern Recognition - CVPR'2001, Hawaii, 2001.
[3] SUGENO M., KANG G.T., Structure identification of fuzzy model, Fuzzy Sets and Systems, Vol. 28, pp. 15-33, 1988.
[4] TAKAGI T., SUGENO M., Fuzzy identification of systems and its application to modeling and control, IEEE Trans. on System, Man and Cybernetics, Vol. 15, No. 1, pp. 116-132, 1985.
[5] VAPNIK V.N., The nature of statistical learning theory, Springer, New York, 1995.