FEATURE EXTRACTION AND SELECTION, BASED ON HYBRID, MIXED DOMAIN FEATURES IN HRV SIGNALS NEURAL CLASSIFIER

P.S. Kostka* **, E.J. Tkacz* **

* Silesian University of Technology, Institute of Electronics, Division of Microelectronics and Biotechnology, Gliwice, Poland
** Medical University of Silesia, Faculty of Pharmacy and Laboratory Medicine, Department of Bionics, Sosnowiec, Poland.

pkostka@slam.katowice.pl , etkacz@slam.katowice.pl

Abstract: Neural classifier system with preliminary feature extraction and selection process using time-frequency representation of heart rate variability (HRV) signal is presented. The crucial point of described method is hybrid multi-domain feature set creation, combining different type parameters as well as feature selection based on the measure of class separability property, computed for each extracted feature. Regarding specific properties of non-stationary HRV signal, wavelet transform was chosen as time-frequency representation tool. Obtained results are connected both with optimal feature extraction and selection of HRV signals from patient with coronary artery disease as well as classifier performance verification, where the comparison of supervised multilayer perceptron with unsupervised learnt self organizing maps (SOMs) was presented.

Introduction

Common way used to improve the classifier performance, where the original input signal described in *N*-element space $x_1 \in X \subseteq \Re^N$ is mapped to output classifier vector space with *K*-class labels $y \in Y = \{y_1, y_2, ...y_K\}$ (N>>K) is the reduction of too high input feature vector size in intermediate feature extraction and selection stage. The basic goal of this preliminary stage is to reveal only the most discriminate features for given task and discard remain, reducing also the classifier complexity. Proposed feature extraction tools almost always must depend on the specificity of classification task to be sensitive to features, which will be able to distinguish between health and pathology cases. Classifier presented in this paper was designed for the problem of coronary artery disease detection based on heart rate variability (HRV) signal analysis. This signal reflects interaction between cardiovascular and autonomous nervous system (ANS), which controls the hemodynamics and the heart work [2]. Presented in literature experimental results of synchronize HRV recordings before, during and after percutaneous transluminal coronary angioplasty [3] allowed to state, that when a coronary artery is blocked, the control is usually affected due to blood flow restrictions and to pressure changes induced mechanically in the affected area of the artery [4][5]. Such experiments confirm the modulation of HRV signal by ANS mainly in frequency ranges: low-frequency band (LF,0-0.07 Hz), mid-frequency band (MF,0.07-0.15 Hz), and high-frequency band (HF, 0.15-0.45 Hz) [6].

Taking into consideration well known facts, that HRV features are included both in time and frequency domain the crucial point of feature extraction part proposed in this work is creation of hybrid – multitype feature vector combining time (T), frequency (F) and time-frequency (T-F) signal representation parameters. Assuming, that important HRV based features are characterized by local information in the duals domains of time and frequency and treating HRV as non-stationary signal from its nature, wavelet transform was chosen as T-F signal representation tool [4][7][8].

In next feature selection stage the most representative feature set is created based on feature ranking competition algorithm described in next sections. Supervised learnt neural network structure based on multilayer perceptron (MLP) as well as for comparison the unsupervised Kohonen Self Oraganizating Maps (SOMs) fulfill the role of nonlinear classifier of extracted signal representation parameters in proposed method of screening examinations of coronary artery disease [9][10].

Because of the specificity of HRV signal in proposed classifier, hybrid ,mixed domain, a new feature vector is created from time (T), frequency(F) and time-frequency (T-F) analysis parameters.

Materials and Methods

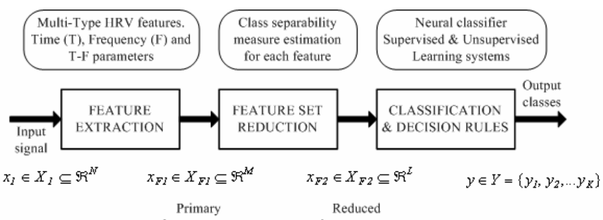

Figure 1. General structure of feature extraction and classification stages of proposed method.

Proposed classifier for screening examinations of patient with coronary artery disease, based on their heart rate variability (HRV) signals consists of following stages (Fig.1):

I. HRV signal preprocessing consisting of:
   - Continous representation of unevenly spaced original HRV signal by means of Derative Cubic Spline Interpolation (DCSI)
   - Resampling with sampling frequency f_S=5 [Hz];

II. New feature vector extraction, using hybrid, multi-type features from three groups:
   - Time domain features: statistical parameters of original HRV signal and its breath related component Respiratory Sinus Arrhythmia (RSA) and NRSA reflecting influence of remain factors on HR modulation
   - Frequency domain features: spectral parameters computed for analyzed HRV signal.
   - Mixed domain features: T-F HRV representation, based on wavelet transform, which is suitable for non-stationary signals like HRV.

III. Feature Selection, based on class separability index computed for every extracted feature. For each subspace wavelet coefficients were squared and normalised to obtain the energy probability distribution.

IV. Classification – Multilayer perceptron with non-linear activation function and unsupervised learning SOM structure.

V. Decision rules, which assign the neural network outputs to pathological or physiological groups.

Feature extraction.

A generalized feature extraction method can be expressed as a map f: $X_1 \rightarrow X_{F1}$, such that $X_{F1} \in \Re^M$ is the M-dimensional feature space , where M<<N.

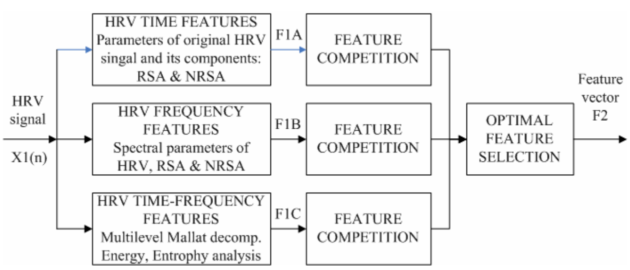

Figure 2. Structure of hybrid multi-type HRV feature extraction and selection algorithm.

As presented in Fig.2. three feature sets based on HRV signal were computed.

I. The RSA and NRSA components were obtained from HRV and then set of statistical parameters (mean, std deviation, range) characterizing these signals were calculated to create the $F1_A$ feature vector:

$F1_A = [HRV_M , HRV_{STD} , HRV_{RNG} , RSA_M , RSA_{STD} , RSA_{RNG} , NRSA_M , NRSA_{STD} , NRSA_{RNG}]$

II. The energy of LF and HF component of HRV spectrum as well as their ratio were included in $F1_B$ feature vector:
$F1_B = [HRV_{LF\_EN}, HRV_{HF\_EN}, HRV_{LF\_HF\_R}]$

III. The most complex feature set F1C was created based on time-frequency (T-F) HRV analysis.
The specificity of HRV signal, which as it is well known has important features included both in time and frequency domain conditioned the area of appropriate feature extraction methods searching to the field T-F signal representation.

Considering the possibility of using several T-F methods including: Short Term Fourier Transform, Wigner Distribution [11][12] and their modification - Smoothed Wigner Distribution as well as Choi Williams Distribution their main limitation considering our classification task is that, they require the analyzed signals to be full or quasi – stationary. In case of HRV signal, which is non-stationary from its nature, mentioned above methods could not be able to reveal all important features for further classification. That's why we decided to choose as feature extraction tool verified in many applications wavelet transform, which is suitable to deal with non-stationary signals.

Wavelet transform as a feature extractor.

As a feature extraction tool the wavelet transform based on multilevel Mallat signal decomposition [13] was used. Taking into consideration specific features of the HRV signal, especially that its significant frequency components are included in the range: $f_{HRV} \in <0;0.5>$ *[Hz]* , the grid of discrete wavelet scale – a values was created, corresponding to Mallat signal decomposition levels (for sampling frequency f_S = 5[Hz], six levels corresponding to scale values: a^i , i=3..8 were taken into cosideration). Multilevel Mallat decomposition on every level corresponds to two-channel filtering using low and high pass filters to extract the detail and approximation signal component respectively (2),(3):

$$c_j^G(k) = \frac{1}{2^j}\left\langle f(x), \phi\left(\frac{x-k}{2^j}\right)\right\rangle \qquad (2)$$

$$c_j^H(k) = \frac{1}{2^j}\left\langle f(x), \psi\left(\frac{x-k}{2^j}\right)\right\rangle \qquad (3)$$

where: c_j – wavelet coefficient on j^{th} decomposition level and k^{th} translation, $\phi(x)$ - scaling function, $\psi(x)$ – wavelet function.

As a next step, to create the new features vector, for every signal component obtained on each decomposition level a set of parameters was computed. For each subspace wavelet coefficients were squared and normalised to obtain the energy probability distribution (4):

ISSN: 1727-1983 © 2005 IFMBE

$$p_i = \frac{c_i^2}{\sum_{k=1}^{n} c_k^2} \tag{4}$$

For each wavelet scale, the sorted series may be considered as an inverse empirical cumulative energy distribution function (ECDF). Base on this parameters the Shannon entropy E (5) of energy distribution p_i (4) were calculated as a measure of energy unpredictability in each wavelet decomposition subspace.

$$E = \sum_i p_i \log_2(p_i) \tag{5}$$

This procedure allowed to reveal a group of new features based on energy and entropy measures.

The whole set of new feature vectors $\overline{F1_{C1}}$ .. $\overline{F1_{C5}}$, created as a result of multilevel Mallat signal decomposition, which is put to the input of classifier structure includes the following groups of parameters as a series for every of ith decomposition level:

I. Mean values of wavelet coefficients in each subband (frequency distribution information) - $\overline{F1_{C1}}$
II. Standard deviations of wavelet coefficients (level of change of frequency distribution information) - $\overline{F1_{C2}}$
III. Energy of ith component - $\overline{F1_{C3}}$
IV. Shannon entropy of wavelet component (distribution of the amount of information included in every subband) - $\overline{F1_{C4}}$
V. Shannon entropy E of energy distribution p_i (5) - $\overline{F1_{C5}}$

2.2 Feature selection

Feature set may be considered near to optimum if it minimizes chosen error based criterion function. There are two approaches to feature selection problem:

- Feature subset selection approach used in this work.
- Feature projection, which tries to find optimal original feature combination (projection) into smaller set of new features. Principle component analysis (PCA) [14] or projection pursuit [15] are often used feature projection methods.

In presented HRV classifier structure, dimension reduction method based on selecting the best feature subset according to assumed criteria were used. Because the evaluation of optimal cost function using probability of misclassification is too complex [16], in presented approach simpler criterion based on class separability (CS) were applied.

Considering a two class problem with an original M-dimensional feature set space X_{F1} a feature selection algorithm used in presented work is following:

I. Create two feature matrices M_{F1}^{p}, M_{F1}^{q}, representing two classes: p and q consisting of patterns (vectors) $x_{F1}^{(p,m)}$, $x_{F1}^{(q,m)}$ from learning data set (6), (7):

$$M_{F1}^{p} = \left[x_{F1}^{(p,1)}, x_{F1}^{(p,2)} \ \ldots \ x_{F1}^{(p,P)} \right] \tag{6}$$

for class p, and

$$M_{F1}^{q} = \left[x_{F1}^{(q,1)}, x_{F1}^{(q,2)} \ \ldots \ x_{F1}^{(q,P)} \right] \tag{7}$$

for class q, where:

$$x_{F1}^{(p,m)} = \left[x_{F1\_1}^{(p,m)}, x_{F1\_2}^{(p,m)} \ \ldots \ x_{F1\_M}^{(p,m)} \right]^T \tag{8}$$

is a mth pattern in class p.

II. Assuming, that we are trying to evaluate the "discriminant power" of each single feature separately (not e.g. feature combination), according to class separability criteria the discriminability of ith feature is represented by ith row in feature matrices M_{F1}^{p} or M_{F1}^{q} (depending on class type).

III. Define $DM(p_i, q_i)$ as a discriminate measure for the ith feature, which expresses what is the value of separability weight of this given feature in classification process.

Different type of $DM(p_i, q_i)$ can be considered and several was tested [17]:

a) Fisher's class separability index:

$$DM(p_i, q_i) = \frac{\left(mean(p_i) - mean(q_i)\right)^2}{\mathrm{var}(p_i) + \mathrm{var}(q_i)} \tag{9}$$

,where mean(·) and var(·) are computed across ith matrice row.

b) Relative entropy:

$$DM(p_i, q_i) = p_i \log \frac{p_i}{q_i} \tag{10}$$

c) Symetric relative entropy:

$$DM(p_i, q_i) = p_i \log \frac{p_i}{q_i} + q_i \log \frac{q_i}{p_i} \tag{11}$$

d) Euclidean distance:

$$DM(p_i, q_i) = \left\| p_i - q_i \right\| \tag{12}$$

Calculate $DM(p_i, q_i)$; $i = 1..M$ for every of M primary features.

IV. Sort obtained in III $DM(p_i, q_i)$ values to create a feature rank as a results of feature competition.

V. Choose the most discriminant L features to create a new feature vector X_{F2}

Results

Proposed structures were tested using the set of clinically characterized heart rate variability (HRV) signals of 62 patients, as cases with a coronary artery disease of different level. Additionally similar control group of healthy patients was analyzed. Whole database was divided into learning and verifying set.

Classification task was defined as the trial of two group (healthy and pathology cases) distinguish, based on new feature subsets obtained in feature extraction and selection stages of whole procedure.

First group of results is connected with searching for the optimal feature subset for given classification task. To find the most discriminant parameters obtained from input HRV signal analysis, for every feature included in time (T) domain HRV feature group - $F1_A$, frequency (F) domain feature set - $F1_B$ and T-F HRV representation parameters: $F1_{C1}$-$F1_{C5}$ a discriminate measure $DM(p_i, q_i)$ was computed (Table 1). Table 1 presents the computed feature separability measure $DM(p_i, q_i)$ distribution among all HRV based parameters group used in described classifier system.

Apart from feature subset included the L features with maximum class separability properties $DM(p,q)$, additional subset consisting of seven features – one, the best representative from each of HRV parameters group was created and this vector seemed to be optimal feature set (OFS) for HRV signal classification.

Table 1: Results of normalised discriminate measure $DM(p_i, q_i)$ computed for each feature from assumed HRV based feature group.

$F1_A$	$F1_B$	$F1_{C1}$	$F1_{C2}$	$F1_{C3}$	$F1_{C4}$	$F1_{C5}$
0,22	0,1	0,46	0,33	0,82	0,68	0,91
0,19	0,55	0,39	0,21	0,74	0,6	1
0,15	0,22	0,16	0,44	0,39	0,32	0,45
0,58		0,08	0,29	0,22	0,33	0,26
0,65		0,05	0,2	0,15	0,21	0,16
0,4						
0,25						
0,11						
0,29						

Fig.3 presents influence of the most discriminative features number included in new feature vector on final learning phase error of 2-layer perceptron, used in classifier system part.

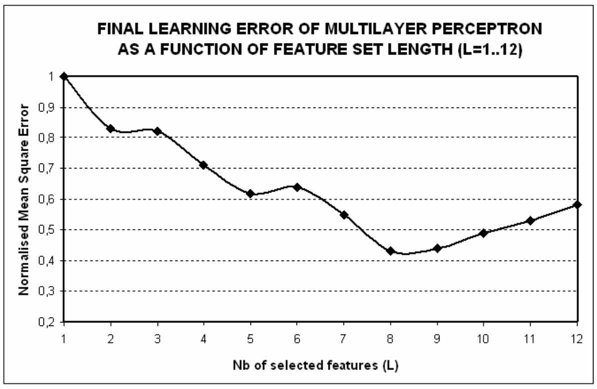

Figure 3. Neural network classifier training error as a function of number of chosen features in new feature vector.

Finally whole presented classifier system were verified using test set of HRV data from patient with coronary artery disease. Common used sensitivity and specificity classifier performance measures obtained for different new feature vector length (L) are presented Table 2 for two types of neural classifier structures, supervised multilayer perceptron and unsupervised learnt Kohonen self organizating maps (SOMs).

Table 2: MLP and SOM neural classifier performance with(+FE&FS) and without(no FE&FS) feature extraction (FE) and feature selection (FS) stages.

Classifier type	Sensitivity[%]	Specificity[%]
MLP no FE&FS	66	69
SOM no FE&FS	58	61
MLP + FE&FS	86	89
SOM + FE&FS	84	88

Discussion

New heart rate variability (HRV) signal representation belonging to the reduced $X_{F2} \subseteq \Re^L$ space, obtained as a result of feature selection process from hybrid multi-domain: time (T), frequency (F) and T-F domain HRV parameters was presented. Evaluation of all extracted HRV features based on the measure of its class separability property presented in Table 1 showed, that the most discriminant features for given classification task are the parameters in $F1_{C3}$, $F1_{C4}$ and $F1_{C5}$ feature vectors. These feature vectors include energy, entropy and Shannon entropy E of energy distribution parameters respectively of Mallat HRV signal decomposition components. The most significant features from these vectors are assigned to *d3* and *d4* level of discrete wavelet analysis (frequency subbands: *0.3125 ÷ 0.6250* and *0.1563 ÷ 0.3125* [Hz]). It corresponds to rather high frequency (HF) components of HRV PSD function.

Results of optimal feature selection based on learning phase error optimisation were not until the end confirmed in final whole system verification step using training set of data. Assuming the measure of classifier sensitivity and specificity as system performance indicators as presented in Table 2, the best results were obtained for new feature vector created by taken the best one feature from each feature vectors: - $F1_A$, $F1_B$ and $F1_{C1}$-$F1_{C5}$ (based on indicators included in Table 1). Results obtained for two different tested neural classifier part structures: MLP and SOMs were comparable.

Described method of feature selection was limited to the case, that every analysed feature was taken as single feature (not e.g. several feature combination) what could affect the process of discriminaty measure (DM(pi,qi) computation. That's why to create optimal feature vector simply taking the feature with maximal value of DM is not enough. It was shown during our tests, where the best classification results were obtained for the feature combination consisting of the best feature representative from each group: $F1_A$, $F1_B$, $F1_{C1}$-$F1_{C5}$ but not for the simply combination of feature with the highest value of class separability parameter DM(pi,qi) computed for every feature in feature selection stage.

Conclusions

To conclude, obtained results showed, that before pattern classifier can be properly designed and effectively used, it is necessary to consider the feature extraction and data reduction problems. Feature extraction should consists in choosing those features, which are most effective for preserving the class separability.

Presented classification procedure gave satisfactory results, considering described classification algorithm as a contribution to coronary artery disease detection on preliminary screen examination stage.

References

[1] DUDA R.O., HART P.E. (1973): 'Pattern classification and scene analysis', John Wiley & Sons, New York.

[2] MARCIANO F., BONADUCE D., PETRETTA M., VALVA G., MIGAUX M.L. (1995): 'Spectral behaviour of heart rate variability in acute ischemic episodes' in *Computers in Cardiology.* Los Alamitos, CA: IEEE Comput. Soc. Press, pp. 111-114.

[3] CLARIÁ F. VALLVERDÚ M. CAMINAL P. (2002): 'The effects of coronary occlusion location on the RR signal'. IEEE Engineering in Medicine and Biology. July/August. pp.59-64.

[4] THAKOR N.V., SHERMAN D.L. (1994): 'Biomedical problems in time-frequency-scale analysis – new challenges', Proceedings of the IEEE-SP pp. 536–539.

[5] SAUL J.P. (1995): 'Heart Rate Variability and Sudden Death: What's the Connection?' Amsterdam, The Netherlands: IOS Press.

[6] LÁNG E., CAMINAL P., HORVÁTH G., JANÉ R., VALLVERDÚ M., SLEZSÁK I., BAYÉS DE LUNA A. (1998): 'Spectral analysis of heart period variance (HPV)—A tool to stratify risk following myocardial infarction' *J. Med. Eng. Techonol.,* vol. 22, no. 6, pp. 248-256.

[7] AKAY M. (1995): 'Time-Frequency and wavelet analysis', IEEE EMB Magazine 14(2).

[8] AKAY Y.M., AKAY M., WELKOWITZ M. KOSTIS J. (1994): 'Noninvasive detection of coronary artery disease', IEEE Engineering in medicine and biology, 11/12, pp. 761-764.

[9] TKACZ E., KOSTKA P. (2000): 'An Application of Wavelet Neural Systems for Classification Patients with Coronary Artery Disease Based on HRV Analysis', World C. of Med. Physics and Biom. Eng. - Chicago.

[10] KOSTKA P., TKACZ E. (2002): 'The comparison of the wavelet-neural systems and radial neural networks in classifier structure',16th Intern. Eurasip Conf. Biosignal2002, pp.46-48,June 2002, Brno,Czech Republic.

[11] HLAWATSCH F., BOUDREAUX-BARTELS G.F. (1992): 'Linear and quadratic time-frequency signal representations', IEEE Signal Processing Mag., vol. 9, pp. 21-67.

[12] COHEN L. (1995): 'Time-Frequency Analysis', Englewood Cliffs, N.J.: Prentice Hall Signal Processing Series.

[13] MALLAT S. (1989): 'A theory for multiresolution signal decomposition: the wavelet representation', IEEE Transaction on Pattern Analysis, 7(11):674-693, 1989.

[14] KARHUNEN, JOUTSENSALO J. (1995): 'Generalizations of principle component analysis, optimization problems, and neural networks', Neural Networks, vol. 8, No. 4, pp. 549-562.

[15] FRIEDMAN W.J., TUKEY, J.W. (1974): 'A projection pursuit algorithm for exploratory data analysis' IEEE Transactions on Computer, Vol. 23, pp. 881-889.

[16] FUKUNAGA K. (1990): 'Introduction to statistical pattern recognition'. 2nd edition. Academic Press, San Diego, CA.

[17] KULLBACK S. LEIBLER R.A. (1951): 'On information and sufficiency', Annals of Mathematics and Statistics, 22, pp. 79-86.

ISSN: 1727-1983 © 2005 IFMBE