Embedded space, attractor,
PCA, kernel space,
fuzzy clustering

Tomasz PRZYBYŁA[1], Janusz WRÓBEL[1], Robert CZABAŃSKI[2],
Krzysztof HOROBA[2], Tomasz PANDER[1], Michał MOMOT[2]

SEGMENTATION OF BIOMEDICAL SIGNALS USING AN UNSUPERVISED APPROACH

The paper presents an unsupervised approach to biomedical signal segmentation. The proposed segmentation process consists of several stages. In the first step, a state-space of the signal is reconstructed. In the next step, the dimension of the reconstructed state-space is reduced by projection into principal axes. The final step involves fuzzy clustering method. The clustering process is applied in the kernel-feature space. In the experimental part, the fetal heart rate (FHR) signal is used. The FHR baseline and the acceleration or deceleration patterns are the main signal nonstationarities but also the most clinically important signal features determined and interpreted in computer-aided analysis.

1. INTRODUCTION

Pattern recognition methods play an important role in analysis of biomedical signals. Recognition process contains two main stages: a pre-processing stage and a classification stage [10, 12]. In the pre-processing stage, feature values of the recognized objects are estimated. Features are variables that carry information about processed objects. The estimated features create a feature-space. Features are placed in different regions in the feature-space for different objects. The second stage deals with classification of objects. In the classification process, the feature-space is transformed into a decision space. There are two approaches for the decision space creation. The first approach (supervised) requires a learning set. In the second approach (unsupervised) the decision space is created using only a recognition set. In most cases the unsupervised recognition involves a clustering method [1, 8]. In the presented work, at the first stage the state–space is reconstructed by applying the Taken's method [16] and then the state-space is projected onto its principal axes. This operation corresponds to the first stage of the general pattern recognition procedure, i.e. the feature-space is created. The feature-space contains typical aggregations of objects: "clouds" and "loops". In most cases, the "cloud" corresponds to the baseline of the analyzed signal, while the "loop" corresponds to these parts of the signal that significantly differ from the baseline [6]. Finally, the obtained feature–space is transformed into decision space by a fuzzy clustering method [9, 11].

The experimental part concerns a chosen group of biomedical signals registered during fetal monitoring. Fetal heart rate (FHR) signal is the main source of information on the fetal state in present-day perinatal medicine. The fetal heart rate is characterized by two main components: the basal fetal heart rate (baseline) and the variability of FHR. The waveform representing "a kind of a mean" of the fetal heart rate over time is referred to as the baseline. The FHR variability is associated mainly with short-lasting accelerations or decelerations of the fetal heart rate. The baseline and the accelerations or decelerations events are the main nonstationary features of the FHR signal. During routine fetal monitoring the FHR signal is usually analysed as one-hour recordings, and the nonstationarities are recognized to be the most clinically important features of FHR signal determined and interpreted both in classical visual and in computer-aided analysis. A correctly determined FHR baseline is a precondition for correct recognition of the acceleration and deceleration patterns. Even a small inaccuracy in the FHR baseline estimation may significantly distort the detection of accelerations or decelerations, which may subsequently lead to false interpretation of clinical symptoms. According to clinical guidelines [2], the baseline is the mean level of the FHR when this is stable, accelerations and decelerations being absent.

[1] Silesian University of Technology, Institute of Electronics, Akademicka St. 16, 44-100 Gliwice, Poland
[2] Institute of Medical Technology and Equipment, Biomedical Signal Processing Department, Roosevelta St. 118, 41-800 Zabrze, Poland

Acceleration is defined as a transient increase in heart rate of 15 bpm or more and lasting 15 seconds or more. Deceleration is a transient episode of heart rate slowing bellow the baseline level of more than 15 bpm and lasting 10 seconds or more. Considering that acceleration and deceleration patterns differ only in the direction of FHR changes and their recognition involves the same problems, the experimental analysis is limited to accelerations only

2. METHODS

2.1. STATE-SPACE RECONSTRUCTION METHOD

The applied technique is an outcome of the theory of nonlinear dynamical systems. In deterministic dynamical systems, the post-transient trajectory of the system is frequently confined to a set of points in the state-space, called an "attractor" [16]. The state-space can be reconstructed by the Taken's embedding operation. For a given signal *x*, the point in the reconstructed state–space is given by

$$x_n = \left[x(n), x(n+\tau), \cdots, x\left(n+(m-1)\tau\right)\right]^T \tag{1}$$

where *x(n)* is the processed signal, τ is the time lag and *m* is the embedding dimension. The product $(m-1)\tau$ is the embedding window. In many applications, the time lag $\tau=1$ is advantageous [6, 14]. So, henceforth this time lag value will be used in this study.

2.2. PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is a technique for extracting a structure from high–dimensional data sets [13, 15]. PCA is an orthogonal transformation of the coordinate systems in which the data are described. The new coordinate system (known as the principal coordinates) is obtained by the projection onto the so-called principal axes of the data. Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ be a dataset, where $\mathbf{x}_i \in \Re^m$. Each element from the dataset is described by *m* features associated with the time stamp *(i)*. Determination of the principal axes begins with centering the data samples and then computing the sample covariance matrix, i.e.

$$\mathbf{C} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T , \tag{2}$$

where $\bar{\mathbf{x}}$ is the sample mean,

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i , \tag{3}$$

and $N = |\mathbf{X}|$ is the cardinal number of the dataset.

The principal axes γ_i are equal to the eigenvectors that correspond to the largest eigenvalues of the covariance matrix ***C***. By definition, the following dependence holds true

$$\mathbf{C}\gamma_i = \lambda_i\gamma_i , \tag{4}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l$ are eigenvalues of ***C***.

The projection onto *l*–dimensional principal space is a linear transformation of $\mathbf{x}_i$ according to the following equation:

$$\mathbf{y}_i = \Gamma^T(\mathbf{x}_i - \bar{\mathbf{x}}), \tag{5}$$

where $\mathbf{y}_i \in \Re^l$ is the *l*–dimensional representation of $\mathbf{x}_i$, and $\Gamma = [\gamma_1, \cdots \gamma_l]^T$.

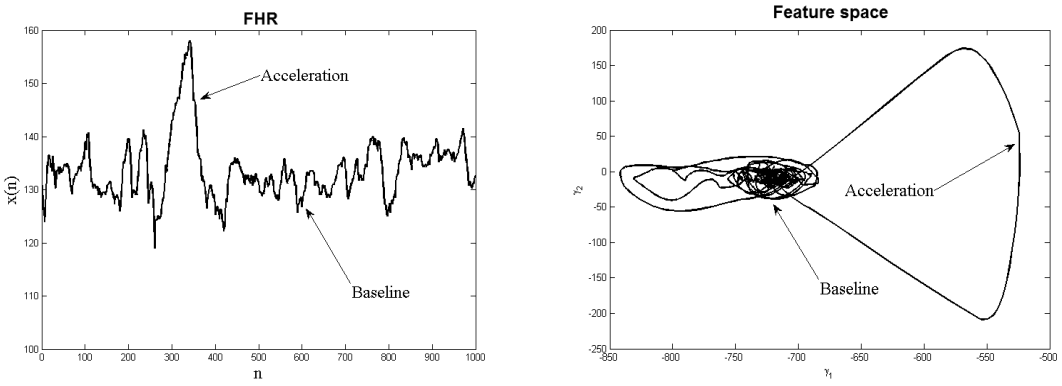

Fig. 1. An example of the fetal heart rate signal with two marked components: baseline and acceleration. Right part presents the reconstructed state-space trajectory onto first two principal directions. The trajectory was obtained for the embedded dimension m=30, and the time lag τ=1

Figure 1 shows an example of the proposed approach. It includes the original fetal heart rate signal as well as the projection of the original signal onto two principal axes. In the original signal two characteristic parts can be distinguished. The first part of the signal contains the baseline with relatively small fluctuations. This part corresponds to the "cloud" in the principal space. A size of the agglomeration is related to the amplitude of fluctuations around the baseline. The second part of the signal contains an acceleration episode. The acceleration episode is an increase of the fetal heart rate, and it is represented by the "loop" in the principal projection.

3. CLUSTERING METHOD

3.1. FUZZY C–MEANS

The fuzzy c–means clustering method [1] is the prototype–based method, where the objective function (clustering performance index) is defined as follows:

$$J(\mathbf{U},\mathbf{V}) = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^{p} \left\|\mathbf{x}_k - \mathbf{v}_i\right\|^2, \tag{6}$$

where: **U** is the partition matrix, **V** is the set of centroids, *c* is the number of groups, and *N* is the number of objects. The FCM seeks the centroids by minimizing the objective function (6) with respect to the probabilistic constraint, i.e.

$$\sum_{i=1}^{c} u_{ik} = 1, \quad \forall 1 \le k \le N \tag{7}$$

The parameter *p* controls the fuzziness of the membership degrees, usually *p=2*. By applying the Lagrange multipliers method, the optimal values of the partition matrix are given by

$$u_{ik} = \left(\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^2 \right)^{-1/p-1}, \quad \forall 1 \le i \le c, \text{ and } \forall 1 \le k \le N \tag{8}$$

where $d_{ik}^2 = \left\|\mathbf{x}_k - \mathbf{v}_i\right\|^2$. The optimal values of the centroids are given by

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} u_{ik}^p \mathbf{x}_k}{\sum_{k=1}^{N} u_{ik}^p} \quad \forall 1 \le i \le c\,. \tag{9}$$

The membership degrees and the centroids can be updated by the expectation–maximization (EM) algorithm. In the E–step, the partition matrix is updated applying (8), while in the M–step the centroids are computed based on (9). The algorithm recursively proceeds until the termination criteria is satisfied.

3.2. KERNEL FCM

Let Φ be an arbitrary non–linear mapping function from original feature space to a space of higher–dimensionality (kernel space). The mapping procedure allows to apply a linear classifier in the kernel space, while in the feature space the original problem could be non–linear and not separable linearly [13, 15]. In the kernel space, the dot product can be expressed by a Mercer kernel K given by $K(\mathbf{x},\mathbf{y}) = \Phi(\mathbf{x})^T \Phi(\mathbf{y})$, where $\boldsymbol{x}$, $\boldsymbol{y}$ belongs to the feature space. Hence, the distance in the kernel space can be replaced by a Mercer kernel function (kernel trick) [7, 13]. The mapping function Φ need not be known. Table 1 lists some of the most widely used kernel functions [5].

Table 1. List of common kernel functions

Gaussian	$\exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{\sigma^2}\right)$	$\sigma^2 > 0$
Polynomial	$\left(\mathbf{x}^T\mathbf{y}+\theta\right)^d$	$\theta > 0, d \in N$
Sigmoidal	$\tanh\left(\mathbf{x}^T\mathbf{y}+\theta\right)$	$\theta > 0$

In the kernel FCM (KFCM) algorithm, the centroids are located in the kernel space. The distance between data sample $\boldsymbol{x}_k$ and the centroid $\boldsymbol{v}_i$ in the kernel space with a mapping function Φ is given by:

$$d_{ik}^2 = \left\|\Phi(\mathbf{x}_k) - \mathbf{v}_i\right\|^2. \tag{10}$$

The KFCM method minimizes the objective function [2] given by:

$$J = \sum_{k=1}^{N}\sum_{i=1}^{c} u_{ik}^p \left\|\Phi(\mathbf{x}_k) - \mathbf{v}_i\right\|^2 = \sum_{k=1}^{N}\sum_{i=1}^{c} u_{ik}^p d_{ik}^2\,. \tag{11}$$

For the Euclidean distance, the optimal values of centroids located in the kernel space are defined as follows:

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} u_{ik}^p \Phi(\mathbf{x}_k)}{\sum_{k=1}^{N} u_{ik}^p}, \quad \forall i = 1,\cdots,c\,. \tag{12}$$

Applying (12), the distance in the kernel space (10) can be expressed as follows

$$\begin{aligned} d_{ik}^2 &= \left\|\Phi(\mathbf{x}_k) - \mathbf{v}_i\right\|^2 = \left(\Phi(\mathbf{x}_k) - \mathbf{v}_i\right)^T \left(\Phi(\mathbf{x}_k) - \mathbf{v}_i\right) \\ &= \Phi(\mathbf{x}_k)^T \Phi(\mathbf{x}_k) - 2\Phi(\mathbf{x}_k)^T \mathbf{v}_i + \mathbf{v}_i^T \mathbf{v}_i \end{aligned}. \tag{13}$$

Using the expression for the centroids (12) in the formula (13) results in:

$$d_{ik}^2 = \left\|\Phi(\mathbf{x}_k) - \mathbf{v}_i\right\|^2 = K(\mathbf{x}_k, \mathbf{x}_k) - 2\frac{\sum_{l=1}^{N} u_{il}^p K(\mathbf{x}_k, \mathbf{x}_l)}{\sum_{l=1}^{N} u_{il}^p} + \frac{\sum_{l=1}^{N}\sum_{j=1}^{N} u_{il}^p u_{ij}^p K(\mathbf{x}_l, \mathbf{x}_j)}{\left(\sum_{l=1}^{N} u_{il}^p\right)^2} \tag{14}$$

Table 2. Main steps of the proposed segmentation algorithm

1)	Initialize the embedded dimension *m* and the time lag τ
2)	Find the reconstruction of the state—space applying the Taken's theory on the analyzed signal
3)	Compute the principal directions Γ and transform the state—space into the principal—space
4)	Fix the number of clusters *c*
5)	Find the partition matrix applying the KFCM method

Applying the Lagrange multipliers with respect to the probabilistic constraint, the optimal values of the partition matrix can be calculated as follows:

$$u_{ik} = \left(\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^2\right)^{-1/p-1} \quad \forall\, i = 1,\cdots,c \text{ and } k = 1,\cdots,N, \tag{15}$$

where d_{ik}^2 is given by (14).

The clustering procedure in the kernel space can be described as follows: update the partition matrix **U** using (12), until the termination criteria is satisfied [3, 4, 17]. The table 2 lists the main stages of the proposed algorithm.

4. EXPERIMENTS

In this section, we present some experiments with a chosen group of biomedical signals by means of the proposed algorithm. The termination criteria in the clustering procedure is defined as follows:

$$\sum_{i=1}^{c}\sum_{k=1}^{N}\left|u_{ik}^{(o)} - u_{ik}^{(o-1)}\right| < 10^{-5} \tag{16}$$

where $u_{ik}^{(o)}$ denotes the value of the partition matrix in the *o-th* iteration.

The maximum number of iteration is fixed at *100*. In most cases, the termination criteria was satisfied before maximum number of iterations was reached. In our experiments, the time lag is fixed τ*=1*. An example of the FHR signal (with duration of 200 s) is presented in Fig. 2. As it was mentioned, the presented fragment contains an acceleration pattern. For the segmentation the following parameters were set: the embedding dimension *m=60*, the number of groups *c=2* (the baseline and the acceleration pattern), as the kernel the polynomial kernel with θ*=10* and *d=4* was chosen. The sum of the first two eigenvalues of the covariance matrix is equal to *79*% of the sum of all eigenvalues. Therefore, only the first two principal directions are taken into account. The obtained results are also presented in the Fig. 2

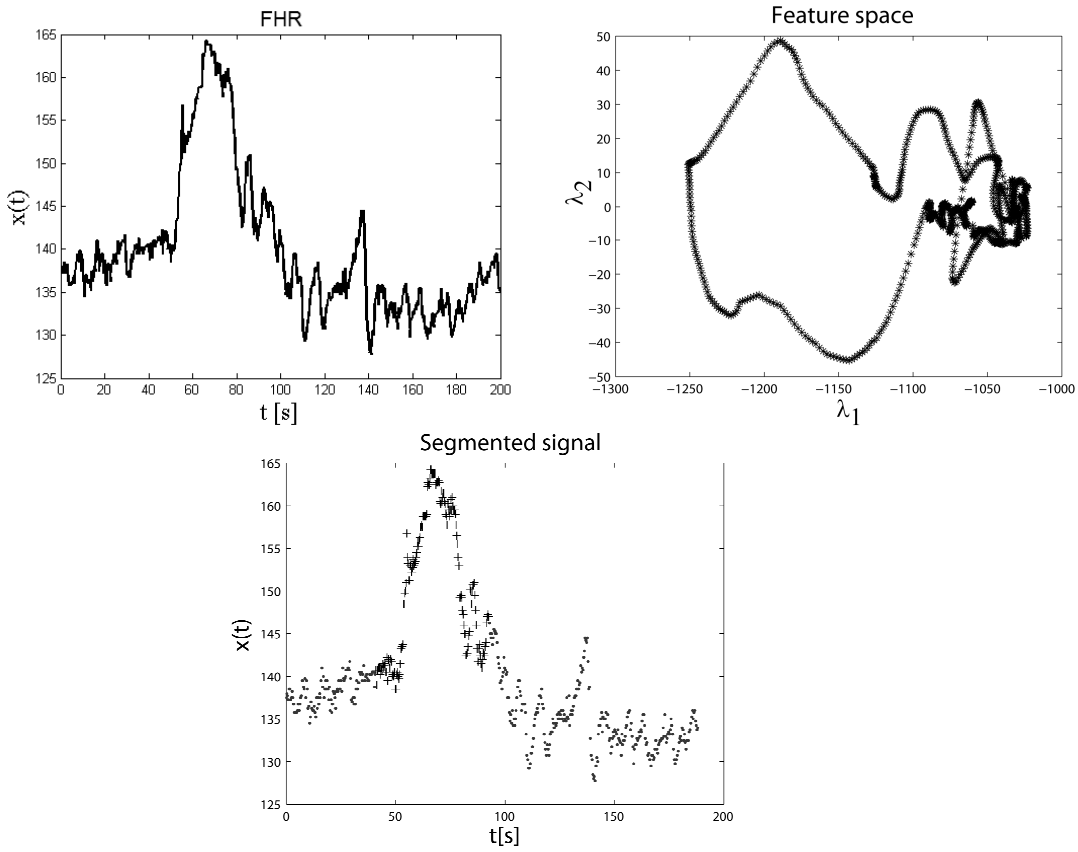

Fig. 2. Segmentation of the FHR signal. The original signal is presented on the top-left and its representation in the feature space, for the embedding dimension *m=60*, on the top-right. At the bottom the segmented signal is presented with the recognized acceleration pattern depicted by crosses and the baseline by red dots

The second FHR signal in Fig. 3 contains two acceleration patterns. Similarly as in the previous experiment, our goal was to find them. In this case, the number of clusters was *c=3* due to "the valley" between the two accelerations.

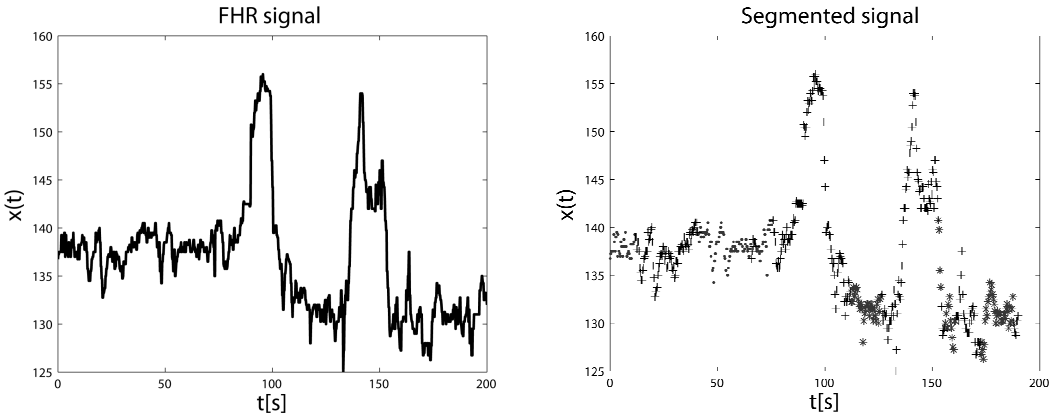

Fig. 3. Segmentation of the FHR signal. The left part presents the original signal and the right the segmented signal. The recognized accelerations patterns are depicted by crosses, the valley between them by stars and the baseline by dots. For this case the embedded dimension is *m=50*

5. CONCLUSIONS

The objective of the presented investigation was to show an unsupervised approach to signal segmentation. Kernel fuzzy c-means is applied in our proposed method as the clustering method. Hence,

the problem of signal segmentation can be described by natural words such as high, low or baseline. For the segmentation process, only the number of desired segments (if exist) is required. As one of the segmentation results, the partition matrix is obtained. The partition matrix can be used in more sophisticated recognition systems. The obtained preliminary results show advantages of the proposed approach. Our current work concentrates on improvements of the efficiency (mainly in the clustering stage) by applying the CUDA technology from NVIDIA.

ACKNOWLEDGEMENT

This work was financed in part by the Polish National Science Center.

BIBLIOGRAPHY

[1] BEZDEK J.C., Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, New York, 1981.
[2] DE CAMPOS D., BERNARDES J., Twenty-five years after the FIGO guidelines for the use of fetal monitoring: Time for a simplified approach?, Int. J. Gynecol. Obstet., 2010, Vol. 110, pp. 1–6.
[3] FILIPPONE M., CAMASTRA F., MASULLI F., ROVETTA S., A survey of kernel spectral methods for clustering, 2008, Patt. Rec. 41, pp. 176–190.
[4] FILIPPONE M., MASULLI F., ROVETTA S., Applying the possibilistic c–means algorithm in kernel induced spaces, IEEE Trans. Fuzzy Sys. 2010, 18, pp. 572–584.
[5] GRAVES D., PEDRYCZ W., Kernel–based fuzzy clustering and fuzzy clustering. Comparative experimental study. Fuzzy Sets and Sys, 2010, 161, pp. 522–543.
[6] KOTAS M., Projective filtering of time–aligned ECG beats, IEEE Trans. Biomed. Eng, 2004, 51, pp. 1129–1139.
[7] MULLER K.R., MIKA S., RATSCH G., TSUDA K., SCHOLKOPF B., An introduction to kernel–based learning algorithms, IEEE Trans. Neural Net, 2001, 12, pp. 181–201.
[8] PEDRYCZ W. Knowledge–based clustering, Wiley–Interscience, 2005.
[9] PRZYBYLA T., PANDER T., HOROBA K., et al., A new Approach to Unsupervised Classification, Journal of Medical Informatics and Technologies, 2011, Vol. 17, pp. 105–111.
[10] PRZYBYLA T., JEZEWSKI J., HOROBA K., et al., Hybrid Fuzzy Clustering Using LP Norms, in Intelligent Information and Database Systems, Eds. NGUYEN N., KIM C., JANIAK A., LNAI 6591/Lecture Notes in Computer Science, 2011, Springer Verlag, pp. 187–196.
[11] PRZYBYLA T., ROJ D., JEZEWSKI J., et al., Generalized Fuzzy Clustering Metod, Journal of Medical Informatics and Technologies, 2010, Vol. 16, pp. 69–76.
[12] SAYED–MOUCHAWEH M., MESSAI N., A clustering–based approach for the identification of a class of temporally switched linear systems, 2012, Patt. Rec. Lett. 33, pp. 144–151.
[13] SCHOLKOPF B. SMOLA A., Learning with kernels, MIT Press 2002.
[14] SCHREIBER T., KAPLAN D., Nonlinear noise reduction for electrocardiograms, 1996, Chaos 6, pp. 87–92.
[15] SHAWE–TAYLOR J., CRISTIANINI N., Kernels methods for pattern analysis, Cambridge Univ. Press 2004.
[16] SMALL M., Applied nonlinear time series analysis. Applications in physics, physiology and finance. World Scientific 2005.
[17] TSAI D.M., LIN C.C., Fuzzy c–means based clustering for linearly and nonlinearly separable data, 2011, Patt. Rec. 44, pp. 1750–1760.