unsupervised classification, fuzzy clustering,
principal component analysis, fetal monitoring

Tomasz PRZYBYŁA[*],
Janusz JEŻEWSKI[**], Dawid ROJ[**]

UNSUPERVISED CLUSTERING FOR FETAL STATE ASSESSMENT BASED ON SELECTED FEATURES OF THE CARDIOTOCOGRAPHIC SIGNALS

In modern obstetrics the cardiotocography is a routine method of fetal condition assessment based mainly on analysis of the fetal heart rate signals. The correct interpretation of recorded traces from a bedside monitor is very difficult even for experienced clinicians. Therefore, computerized fetal monitoring systems are used to yield the quantitative description of the signal. However, the effective techniques enabling automated conclusion generation based on cardiotocograms are still being searched. The paper presents an attempt to diagnose the fetal state basing on seventeen features describing the cardiotocographic records. The proposed method applies the unsupervised classification of signals. During our research we tried to classify the fetal state using the fuzzy c-means (FCM) clustering. We also tested how the efficiency of classification could be influenced by application of principal component analysis (PCA) algorithm. The obtained results showed that unsupervised classification cannot be considered as a support to fetal state assessment.

1. INTRODUCTION

Cardiotocography (CTG) is a widely used method of fetal monitoring, which enables evaluation of a fetal condition during pregnancy and in labour. It relies on simultaneous acquisition and analysis of three signals: fetal heart rate, maternal uterine contractions and fetal movement activity [14]. In traditional cardiotocography the signals are recorded and processed by a bedside fetal monitor. The visual evaluation of printed waveforms is subjective and considerably depends on the experience and knowledge of clinicians [3]. External computer-aided automated analysis allows for more accurate evaluation of signals, providing the obstetrician with a quantitative description of traces [5]. It considerably improves the objectivity and reproducibility of CTG records interpretation. However, so far computer-aided monitoring systems do not provide qualitative assessment of obtained numerical data [15]. In the proposed work we try to assess the fetal state as being normal or abnormal by means of clustering algorithm, basing on quantitative parameters describing the set of extracted features of cardiotocograms.

The clustering aims at assigning a set of objects to clusters in such a way that objects within the same cluster have a high degree of similarity, while objects belonging to different clusters are dissimilar [1, 9, 10]. The clustering methods can be considered as the classification methods that operate on the lowest amount of information about recognized object – its features. Clustering does not need the learning set for object classification. In [11] the partially supervised clustering method has been used for breast cancer classification. The breast cancer Wisconsin data set contains cases of two kind of cancer: cases of the malignant cancer and cases of the benign cancer. The data set has been clustered into two groups, where each group represents one of the cancer type. The obtained number of incorrect classified cases was only 3,4% of the total number of cases. Therefore, a similar approach is presented in this paper.

During our research we tried to classify the fetal state using the fuzzy c-means (FCM) clustering as the unsupervised classification method. Additionally, the principal component analysis (PCA) has been applied in the feature space as well as in the kernel space for the reduction of the features number.

2. METHODOLOGY

The research database used in our experiments contains the results of quantitative analysis of CTG traces recorded in hospital from bedside monitors. The trace report includes parameters describing the fetal heart rate (FHR) signal variability as well as the number of recognized uterine contractions and fetal movements [2, 7]. The original, raw research material included 749 records collected with the fetal surveillance system MONAKO [5]. As an input data set we used only these records which were registered after the 37^{th} week of gestation. Finally the data set included 307 traces of duration varying from 30 to 60 minutes.

The set of seventeen features describing CTG signals was used as an input variables. Fifteen parameters describe the fetal heart rate signal in time domain. Two of them (F1, F2) concern the basal fetal heart rate (average baseline value and fluctuation). The next describes the number of recognized accelerations (per hour) and the F4 the number of deceleration patterns. The largest group of parameters measure the long-term FHR variability (F5, F8, F9, F14, F15) and the short-term (beat-to-beat) variability (F6, F7, F10÷F13) of CTG signals. Finally the number of identified maternal uterine contractions (per hour) and the number of fetal movements were used (F16, F17). The detailed description of the CTG feature extraction procedures has been presented elsewhere [6, 7].

[*] Silesian University of Technology, Institute of Electronics, ul. Akademicka 16, 44-100 Gliwice, Poland.
[**] Institute of Medical Technology and Equipment, Biomedical Informatics Department, ul. Roosevelta 118, 41-800 Zabrze, Poland

Table 1. Input vectors [F1, F2 … F17] for randomly chosen twelve cases from the set of 307 records.

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17
150,1	13,7	20,6	0	45,0	4,75	2,19	17,4	23,9	0,98	0,42	7,72	3,64	8,6	17,1	41,1	0
141,0	12,0	12,4	0	48,8	5,85	2,76	16,6	30,5	1,07	0,49	9,24	4,56	0	17,6	21,2	0
138,3	11,5	3,9	0	54,7	8,22	4,22	18,0	24,0	1,47	0,66	13,46	6,97	0	19,4	106,5	0
148,9	13,3	0	0	42,9	6,35	2,92	15,7	21,0	1,35	0,55	10,05	4,65	0	10,0	16,0	0
142,8	12,8	13,5	0	46,5	6,02	2,60	15,7	29,0	1,18	0,46	9,23	4,06	0	9,7	38,7	0
144,1	13,5	17,4	0	51,3	5,70	2,86	18,2	30,8	1,10	0,52	8,84	4,49	6,5	22,6	48,4	0
137,4	10,8	6,0	2,0	45,9	6,20	2,83	14,1	21,8	1,14	0,42	9,11	4,94	3,3	3,3	0	0
139,6	10,3	10,3	0	47,9	5,48	2,84	16,0	23,6	1,05	0,51	8,25	4,30	0	11,4	56,6	1,7
138,2	10,3	1,8	0	27,5	3,79	2,87	8,9	14,7	0,72	0,52	5,56	4,17	17,6	0	7,1	1,8
142,8	14,2	3,8	1,9	27,1	4,31	2,67	9,5	17,7	0,92	0,48	6,54	4,11	37,5	9,4	35,6	3,8
152,3	14,5	7,1	0	33,5	4,81	2,80	12,8	15,9	0,92	0,52	8,58	4,92	2,4	4,8	17,1	1,4
142,7	12,5	18,0	12,0	62,3	9,50	3,50	20,9	40,6	1,73	0,61	14,33	5,57	0	23,3	28,0	2,0

The output of our method was defined as two-state variable representing the so called fetal outcome – normal (0) or abnormal (1). In obstetrics, it is assumed that validation of the fetal state diagnosed during the pregnancy, i.e. when the CTG signals are being recorded, can be made only retrospectively, using the data forms describing the newborn state and the history of the labour progress. The CTG signals used as the research material are described by real fetal outcome. The set of 307 traces comprises 226 records corresponding to normal and 81 corresponding to abnormal fetal outcome. The records were assessed as abnormal if the value of at least one of four attributes describing the newborn (Apgar score – visually assessed newborn state, umbilical artery pH at birth, umbilical artery base excess at birth, centile of the fetal birth weight) was outside the physiological range. Some examples of the input data sets are presented in Table 1.

2.1. FUZZY CLUSTERING METHOD

The clustering methods can be divided into two main categories: hierarchical and partitional. In the hierarchical clustering a number of clusters need not to be specified a priori. The problems concerning an initialization and an occurrence of local minima are also irrelevant. However, it cannot incorporate a priori knowledge about the global shape or size of clusters since hierarchical methods consider only local neighbours in each step [4].

Prototype-based partitional clustering methods can be classified into two classes: hard (or crisp) methods and fuzzy methods. In the hard clustering methods every case belongs to only one cluster. In the fuzzy clustering methods every data point belongs to every cluster. Fuzzy clustering algorithms can deal with overlapping cluster boundaries.

The partition of an input data set can be described by $c \times N$ matrix $\mathbf{U}$, called the partition matrix, in the following form [1]:

$$\mathbf{U} = \begin{bmatrix} u_{11} & \cdots & u_{1k} & \cdots & u_{1N} \\ u_{21} & \cdots & u_{2k} & \cdots & u_{2N} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ u_{c1} & \cdots & u_{ck} & \cdots & u_{cN} \end{bmatrix} = [\mathbf{u}_1 \cdots \mathbf{u}_k \cdots \mathbf{u}_N]. \tag{1}$$

For the fuzzy clustering methods, the fuzzy partition matrix is defined in the following way:

$$M_{fc} = \left\{ \mathbf{U} \in \Re^{c \times N} \,|\, \forall_{i,j} u_{ik} \in [0,1]; \forall_{1 \le k \le N} \sum_{i=1}^{c} u_{ik} = 1; \forall_{1 \le i \le c} \sum_{k=1}^{N} u_{ik} < N \right\}, \tag{2}$$

where: N is the number of objects, and c is the number of clusters.

The most familiar fuzzy clustering method is the fuzzy c-means clustering method proposed in [1]. The FCM method is the prototype-based method, where the objective function has been defined as follows:

$$J_m(\mathbf{U}, \mathbf{V}) = \sum_{k=1}^{N} \sum_{i=1}^{c} u_{ik}^m \left\| \mathbf{x}_k - \mathbf{v}_i \right\|^2, \tag{3}$$

where: **U** is the fuzzy partition matrix, **V={v₁, v₂, … v_c}** is the set of prototype vectors and $\underset{1\le i\le c}{\forall} v_i \in \Re^p$, $\mathbf{x}_k$ is the feature vector $\underset{1\le k\le N}{\forall} x_k \in \Re^p$, p is the number of features describing the clustering objects, and m is the fuzzyfying exponent called the fuzzyfier.

The optimization of objective function (3) is completed with respect to partition matrix **U** and prototypes of the clusters **V**. The optimal values of the partition matrix can be computed as follows:

$$\underset{1\le i\le c}{\forall}\underset{1\le k\le N}{\forall} u_{ik} = \begin{cases} \left[\sum_{j=1}^{c}\left(\frac{\|\mathbf{x}_k-\mathbf{v}_i\|}{\|\mathbf{x}_k-\mathbf{v}_j\|}\right)^{2/(m-1)}\right]^{-1} & if \quad \Im_k=\varnothing \\ 0 & if \quad \underset{i\in\tilde{\Im}_k}{\forall}, \\ 1 & if \quad \Im_k\neq\varnothing \end{cases} \tag{4}$$

where the sets $\Im_k$ and $\tilde{\Im}_k$ are defined as follows:

$$\begin{aligned} &\underset{1\le k\le N}{\forall}\Im_k = \left\{i \,|\, 1\le i\le c; \|\mathbf{x}_k-\mathbf{v}_i\|^2=0\right\} \\ &\tilde{\Im} = \{1,2,\ldots,c\}-\Im_k \end{aligned} \tag{5}$$

The optimal values of the cluster prototypes can be computed using the formula:

$$\underset{1\le i\le c}{\forall}\mathbf{v}_i = \frac{\sum_{k=1}^{N} u_{ik}^m \mathbf{x}_k}{\sum_{k=1}^{N} u_{ik}^m}. \tag{6}$$

The FCM method can be described as follows:

1° For given data set $\mathbf{X}=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N\}$, where $\mathbf{x}_k \in \Re^p$, fix the number of clusters $c \in \{2, 3, \ldots, N\text{-}1\}$, the fuzzyfing exponent $m \in [1, \mu)$ and assume the tolerance limit ε. Initialize randomly partition matrix **U**, fix iteration counter k=1,

2° Calculate prototype values **V** based on (6),

3° Update the values of the partition matrix using (4),

4° If the partition matrices in two successive steps are similar enough, i.e. $\left\|\mathbf{U}^{(k)}-\mathbf{U}^{(k-1)}\right\|<\varepsilon$ then STOP the clustering algorithm, otherwise $k=k+1$ and go to (2°).

2.2. PRINCIPAL COMPONENT ANALYSIS

The principal component analysis (PCA) is extensively used for a dimension reduction and a feature extraction. The method applies eigenvector decomposition on the covariance matrix to decorrelate features and hence to extract the uncorrelated features [8]. The PCA method is an orthogonal transformation of the coordinate system in which the data are described. The new coordinate system is obtained by projection on so-called principal axes of the data. Data may in fact lie in a lower-dimensional subspace even if no individual feature is constant. This corresponds to the subspace not being aligned with any of the axes. The principal component analysis in nonetheless able to detect such a subspace [4].

In the primal principal component analysis method, for the given data set in p-dimensional space, the l largest eigenvectors of the p×p covariance matrix are computed. Hence, the linear transformation is applied to the input data samples as follows

$$\mathbf{Y}=\Gamma'\mathbf{X}, \tag{7}$$

where: **Y** is the derived data set, **Γ'** is the linear transformation matrix which columns are the eigenvectors, and **X** is the input data set.

The primal principal component analysis method can be described as follows [13]:

1° For given data set $\mathbf{X}=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N\}$, where $\mathbf{x}_k \in \Re^p$, fix dimension l,

2° Compute covariance matrix $\mathbf{C}=\frac{1}{N}\sum_{k=1}^{N}(\mathbf{x}_k-\boldsymbol{\mu})(\mathbf{x}_k-\boldsymbol{\mu})^T$, where $\boldsymbol{\mu}$ is the mean of the data, i.e. $\boldsymbol{\mu}=\frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_k$,

3° Compute the eigenvectors of the covariance matrix $[\boldsymbol{\Gamma},\boldsymbol{\Lambda}]=eig(\mathbf{C})$,

4° Compute the derived data set as $\mathbf{Y}=\boldsymbol{\Gamma}'\mathbf{X}$.

The PCA method (primal PCA) is a linear feature technique. The nonlinear feature extraction, which is directly related to PCA is called kernel PCA. The idea of kernel PCA is first to map input data into some new feature space F typically via nonlinear function Φ and then perform a linear PCA in the mapped space. However, the F space often has a very high dimension. To avoid computing mapping Φ explicitly, kernel PCA employs only Mercer kernels which can be decomposed into a dot product [12]

$$K(x,y)=\Phi(x)\Phi(y)\,. \tag{8}$$

As an example of the Mercer kernel, the polynomial of d^{th}-order or the Gaussian kernel can be mentioned [12, 13].

Let $\mathbf{X}$ be the data set with zero mean, and $\Phi(\mathbf{X})$ be the mapping function into F feature space. The prime PCA method in the F space solves the eigenvectors of the correlation matrix $\Phi(\mathbf{X})\,\Phi(\mathbf{X})^T$, which is also called the kernel matrix $K(\mathbf{X}, \mathbf{X})$. In the kernel PCA, the first l eigenvectors of $K(\mathbf{X}, \mathbf{X})$ are obtained to define a transformation matrix $\boldsymbol{\Gamma}$'.

The kernel principal component analysis can be described as follows [12]:

1° For given data set $\mathbf{X}=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N\}$, where $\mathbf{x}_k \in \mathcal{R}^p$, fix dimension l, and fix kernel function K,

2° Compute the kernel matrix as $\mathbf{K}_{ij}=K(\mathbf{x}_i,\mathbf{x}_j)$, and next centre the data in the kernel space $\mathbf{K}\leftarrow\mathbf{K}-\frac{1}{N}\mathbf{j}\mathbf{j}^T\mathbf{K}-\frac{1}{N}\mathbf{K}\mathbf{j}\mathbf{j}^T+\frac{1}{N^2}\left(\mathbf{j}^T\mathbf{K}\mathbf{j}\right)\mathbf{j}\mathbf{j}^T$, where $\mathbf{j}=[1,\ldots,1]^T$,

3° Compute the eigenvectors of the covariance matrix $[\boldsymbol{\Gamma},\boldsymbol{\Lambda}]=eig(\mathbf{K})$,

4° Compute the derived data set as $\mathbf{Y}=\boldsymbol{\Gamma}'\mathbf{K}$.

The unsupervised classification can be presented in the following way:

1° For given data set $\mathbf{X}$ in the p-dimensional space, fix dimensionality l of derived data set $\mathbf{Y}$,
2° Choose the PCA method – primal or kernel,
3° Compute transformation matrix $\boldsymbol{\Gamma}$', and derived data set $\mathbf{Y}$,
4° Perform classification via clustering method on the $\mathbf{Y}$ data set.

3. RESULTS

In the numerical experiment two types of kernel were used. The first type was the polynomial kernel of d^{th} degree, and the second type of kernel was the Gaussian kernel. The polynomial kernel of d^{th} degree was defined as follows:

$$K(\mathbf{x},\mathbf{y})=\langle\mathbf{x},\mathbf{y}\rangle^d\,, \tag{9}$$

and the Gaussian kernel was defined in the following way:

$$K(\mathbf{x},\mathbf{y})=\exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right). \tag{10}$$

For the clustering method, the following parameters: number of clusters $c=2$, fuzzyfier exponent $m=2$ have been fixed. The specificity index and the sensitivity index have been used for description of the performance quality of the proposed approach. The sensitivity (SE) measures the proportion of actual positive cases (abnormal fetal state) which are correctly identified, and the specificity (SP) measures the proportion of negative cases (normal fetal state) which are correctly identified. Since the partition matrix is randomly initialized, the clustering process has been repeated five times. Hence, the presented index values are the mean values. In the first experiments whole data set, i.e. the set without the reduction of dimension has been grouped into two classes: the first class comprised 165 records and the second class comprised 142 records. We made an assumption that the more numerous class corresponds to the normal fetal outcome.

The results of classification carried out using the proposed method were compared to real fetal outcome, assessed after the delivery. On that basis two parameters describing the performance of classification were calculated. The results were as follows:

SE = 51,85%,
SP = 55,75%.

The influence of feature number l on the performance indices has been tested in the second numerical experiment. The primal PCA was used for the features extraction and the number of extracted features was changed from l=2 up to l=10. The obtained results are presented in Table 2.

Table 2. Values of performance indices SE and SP obtained for primal PCA and various number of features *l*

No. of features	Sensitivity and Specificity [%]
$l = 2$	**SE** = 44,44 **SP** = 58,41
$l = 3$	**SE** = 48,15 **SP** = 57,08
$l = 5$	**SE** = 50,62 **SP** = 55,75
$l = 10$	**SE** = 51,85 **SP** = 55,75

As in the second experiment, the influence of feature number l on the performance indexes was tested, but this time the kernel PCA was used for the feature extraction. For the polynomial kernel the polynomial degree *d* was changed, and for the Gaussian kernel value the σ parameter was changed. The obtained results are presented in Table 3.

Table 3. The performance of indices SE and SP obtained for kernel PCA and various number of features *l*

No. of features	Sensitivity and Specificity [%]					
	Polynomial kernel			Gaussian kernel		
	d=2	d=3	d=5	σ=0.1	σ=1,0	σ=5,0
$l = 2$	**SE**=35,80 **SP**=57,08	**SE**=28,40 **SP**=70,35	**SE**=20,99 **SP**=78,32	**SE**=37,04 **SP**=68,14	**SE**=50,62 **SP**=57,52	**SE**=50,62 **SP**=56,64
$l = 3$	**SE**=35,80 **SP**=57,08	**SE**=29,63 **SP**=70,35	**SE**=20,99 **SP**=78,32	**SE**=38,27 **SP**=66,81	**SE**=50,62 **SP**=56,64	**SE**=50,62 **SP**=56,19
$l = 5$	**SE**=35,80 **SP**=57,08	**SE**=29,63 **SP**=69,47	**SE**=20,99 **SP**=77,43	**SE**=38,27 **SP**=64,60	**SE**=51,85 **SP**=55,31	**SE**=51,85 **SP**=55,75
$l = 10$	**SE**=35,80 **SP**=57,08	**SE**=29,63 **SP**=69,47	**SE**=20,99 **SP**=76,99	**SE**=40,74 **SP**=57,96	**SE**=51,85 **SP**=55,31	**SE**=51,85 **SP**=55,75

First of all, the obtained results show that the unsupervised classification cannot be used in real medical diagnostic procedures. The low values of commonly used performance indices SE and SP, prove that the classification method should use experts knowledge as the training set to ensure robustness.

For the primal PCA method, the performance of classification increases for the higher number of selected features. This is caused by the linear feature extraction in the primal PCA. It can be noticed, that for the kernel PCA, the performance of the classification is nearly constant (for the whole percents). The highest performance of the proposed method has been achieved for the kernel PCA with Gaussian kernel, and for the σ parameter σ=0,1 and for number of features *l*=5.

4. CONCLUSIONS

In the presented work, the unsupervised classification of the fetal state has been shown. The principal component analysis has been used for the feature number reduction and two approaches has been presented for the feature reduction methods: the primal (linear) PCA and the kernel (nonlinear) PCA. It is too bad that obtained results showed that unsupervised classification cannot be considered as a support to fetal state assessment. Our future works concern on the improvement the performance of the unsupervised approach via semi-supervised clustering and discriminant analysis in the kernel space.

ACKNOWLEDGEMT

This work was supported in part by the Ministry of Science and Higher Education resources in 2007-2010 under Research Project R13 028 02.

BIBLIOGRAPHY

[1] BEZDEK J.C., Pattern Recognition With Fuzzy Objective Function Algorithms. Plenum, New York, 1981.

[2] CZABAŃSKI R., JEŻEWSKI M., WRÓBEL J. et al., The Prediction of The Low Fetal Birth Weight Based on Quantitative Description Of Cardiotocographic Signals, Journal of Medical Informatics and Technologies, Vol. 12, pp. 97-102, 2008.

[3] DAWES G.S., REDMAN C.W.G., Patterns of the normal human fetal heart rate, British Journal of Obstetrics and Gynaecology., 89, pp. 276-284, 1982.
[4] DUDA R.O., HART P.E., STORK D.G., Pattern Classification. Wiley-Interscience, New Jersey, 2000.
[5] JEŻEWSKI J., WRÓBEL J., HOROBA K. et al., Centralised Fetal Monitoring System With Hardware-Based Data Flow Control, Proc. of III Int. Conf. MEDSIP, pp. 51-54, Glasgow, 2006.
[6] JEŻEWSKI M., CZABAŃSKI R., HOROBA K. et al., Prediction of newborn sex with neural networks approach to fetal cardiotocograms classification. In: Pietka E., Kawa J. (eds) Advances in Soft Computing Series, Vol. 47, pp. 299-306, 2008.
[7] JEŻEWSKI M., WRÓBEL J., HOROBA K. et al., The prediction of fetal outcome by applying neural network for evaluation of CTG records. In: Kurzyński M., Puchała E., et. al. (eds) Advances in Soft Computing Series, Vol. 45, pp. 532-541, Springer Verlag, 2007.
[8] JIANG X. Asymmetric Principal Component Analysis and Discriminant Analyses for Pattern Classification, IEEE Trans Pattern Analysis and Machine Intelligence, Vol. 31, pp. 931-937, 2009.
[9] KAUFMAN L., ROUSSEEUW P. Finding Groups In Data. Wiley-Interscience, New Jersey, 1990.
[10] PEDRYCZ W. Knowledge-Based Clustering. Wiley-Interscience, New Jersey, 2005.
[11] PRZYBYŁA T., Breast Cancer Diagnosis via Fuzzy Clustering With Partial Supervision, Task Quarterly 8, pp. 193-198, 2004.
[12] SCHOELKOPF B., SMOLA A.J. Learning with Kernels. The MIT Press, 2002.
[13] SHAWE-TAYLOR J., CRISTIOANINI N. Kernel Methods for Pattern Analysis. Cambridge University Press, 2004.
[14] SŁOMKO Z., Biofizyczne Monitorowanie w Medycynie Perinatalnej, PWN, Warszawa, 1991.
[15] van GEIJN H.P., Fetal monitoring _ present and future: the evaluation of fetal heart rate patterns, European Journal of Obstetrics Gynecology and Reproductive Biology, 24, pp. 117-119, 1987.