Evaluation of Fetal Heart Rate Recordings Based on Clustering

Tereza Janíčková, Václav Chudáček, Jiří Spilka

Czech Technical University in Prague, Prague, Czech Republic

Abstract

Fetal heart rate (FHR) recorded within the cardiotocography (CTG) measurement is currently the main method to evaluate fetal health state during delivery. The CTG provides valuable information about fetal behavior as a reaction to stressful events (hypoxic episodes).

The presented paper proposes to use data driven analysis of FHR – the clustering analysis of features derived automatically from the signal using novel method of signal approximation called SAX. Even though the clustering is well grounded in signal processing tasks in the field of FHR research it is used sparingly due to high inter-individual variability of fetuses and difficulties to link different temporal events.

Data from the open access CTU-UHB database (552 CTG records) available at the Physionet are used and the 30 minutes segments at the end of the first stage of labor are analyzed.

The classification based on clustering achieved sensitivity of 58.7% and specificity of 69.4% for the two class classification – well comparable to other pH based studies. Sensitivity was improved to 71.4% for six cluster settings – thus suggesting different classes of FHR. This is in contrast with objective evaluation (two classes usually determined with pH threshold) and three classes (used by clinicians for evaluation). Nevertheless to properly describe the link between these clusters and clinical evaluation robust interpretation is still necessary.

1. Introduction

Cardiotocography (CTG) refers to the simultaneous recording of fetal heart rate (FHR) and uterine contractions (UC) using ultrasound Doppler transducers or scalp electrodes for the former and external belts or intrauterine catheters for the latter.

CTG succeeded intermittent auscultation as the main tool for providing information about fetal behavior and wellbeing with a hope to improve the quality of information compared to its predecessor. The sole purpose of CTG is to give hints to clinicians for timely intervention to prevent adverse long term consequences caused by intrapartum asphyxia.

However 40 years after the introduction of CTG into clinical practice the initial enthusiasm was replaced by skepticism and the CTG is now being blamed for an increased rate of cesarean sections [11].

Moreover, it turned out that the interpretation of CTG is far from being a trivial task, resulting in high inter and intra-observer variability among clinicians [1]. Despite all these controversial findings, CTG is still the prevalent method for intrapartum fetal surveillance [2] with its interpretation relying primarily on visual assessment of the CTG recording. The interpretation is based typically on the guidelines issued by the International Federation of Gynecology and Obstetrics (FIGO) guidelines [3] while several national updates and tweaks have also been released [4].

Many features were tested, e.g. time-domain features, frequency-domain features, time-frequency domain features and nonlinear ones [5, 12]. For the classification purposes, many machine learning paradigms have been tested starting from the simple ones such as k-nn to more complicated ones such as Support Vector Machines (SVMs) [7] artificial neural networks (ANNs) [6] and Hidden Markov Models (HMMs) [8] to name just a few.

2. Methods

2.1. Signal pre-processing

The Fetal Heart Rate signal (FHR), no matter the method that is used for its acquisition (ultrasound Doppler probe or a scalp electrode) is contaminated by a lot of artifacts – see Figure1a). As a result a preprocessing step is deemed necessary before any attempt to extract useful information in the form of a feature set.

For removing artefacts we used a method according to Bernardes [9]. First, we determined a stable segment, defined as a segment of 5 consecutive beats, wherein the total error is less than 10 bpm. If the difference is greater than 25 bpm, the spline interpolation between the last sample of the previous stable segment and the first sample following a stable segment is performed. Spline interpolation, was chosen because in this case complements the data better than linear interpolation.

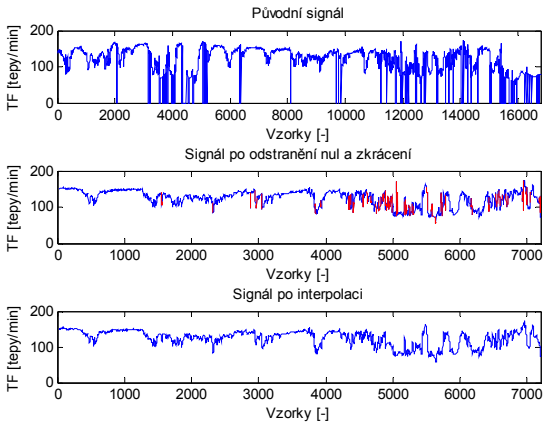

Figure 1. Filtration of the signal, interpolated segments in red.

2.2. Data used

No All the experiments were carried out using the CTU-UHB database [10]. The database consists of 552 records acquired using STAN and Avalon devices between years 2009 and 2012 at the obstetrics ward of the University Hospital in Brno, Czech Republic. The majority of the recordings were acquired using an ultrasound probe (412 records), and all recordings were regularly sampled at 4Hz. A detailed description of the CTU-UHB is provided in [10].

2.3. SAX

The most common method of approximating time series is PAA. After dividing the segments in each segment, we calculate the mean or median. The number will characterize the entire segment. Then it will be added to any interval, according to previously established lines. Intervals are assigned the integers from 1 to K, where K is the number of predefined intervals. The entire database is then expressed by a single matrix of integers. Column specifies the number of signals, rows are individual features. This method is very simple to calculate and nicely captures the changes in the signal. The approximation also suppresses artifacts that could have not been removed by filtration.

SAX (Symbolic Approximation) [13,14] is the main method of symbolic expression time series and currently is often used in time series analysis. It meets the requirements for sufficient dimension reduction in the time-series and enables the possibility to calculate the distance between the time-series. Representation by sax significantly accelerates the process of acquiring knowledge from data, while preserving the accuracy of the results. SAX is based on the PAA, but instead of numbers representing each segment, symbols are used. To obtain a symbolic representation of time series, first perform the transformation by PAA. Each segment can be expressed as a real number, it is therefore necessary to discretize these levels. From the histogram of the time series boundary intervals can be obtained.

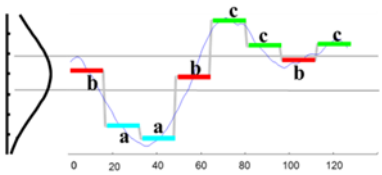

Figure 2. Principle of SAX.

One of the most important features of SAX is that it allows to lower-limit distances measures. This means that the distance between the approximated lines is always less than the distance between rows in the original shape. This property is useful to determine the error and the acceleration calculation [15].

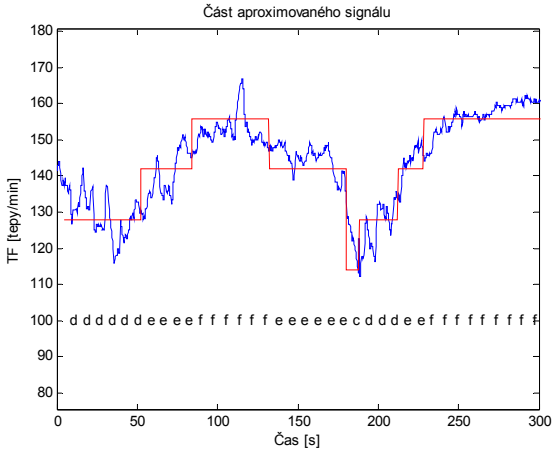

Figure 3. CTG signal represented by SAX.

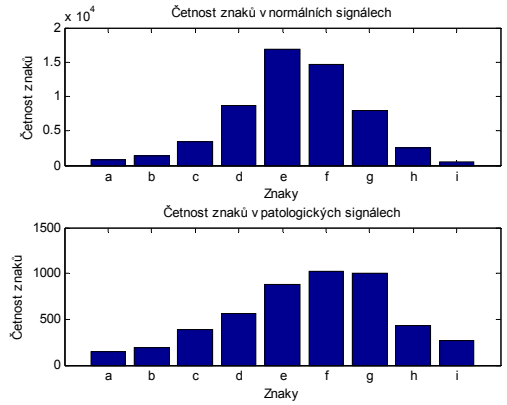

Figure 4. Disribution of characters in SAX. Normal class above, pathological class below.

2.4. Hierarchical clustering

The result of this method is called dendrogram, which has the form of a binary tree. It is up to the user to interpret the dendrogram. The algorithm progresses from bottom-up or top-down. Either at the beginning of each element represents a cluster based on similarity to gradually merges with other clusters (agglomerative clustering), or all the data in one cluster, which is divided (divisive clustering). The procedure is repeated until the desired texture is reached [16].

2.5. K-means

K-means is a very popular iterative algorithm. The goal of k-means is to find clusters, where the minimized error, defined as the sum of the distances of elements from the center of your cluster. At the beginning of the algorithm to randomly select the center of clusters, their number is determined by the value of K. Based on the smallest distance is assigned to elements of the clusters. The next step is recalculated center cluster. The procedure is repeated until there are changes. The advantages of this algorithm are its simplicity and efficiency, but the user must know the number of clusters and largely depends on the initial start [17].

K-means algorithm is very efficient method of clustering. It can be used for multi-dimensional data, such as time series. Each signal sample is considered a variable corresponding element. This is an iterative algorithm with a relatively high computational cost. K-means is usually chosen randomly initial centers of clusters, clusters may therefore be created from the same data differ. Better results can be achieved properly pre-selected centers, or we can repeat the algorithm several times and choose the best result. We chose the second option, the calculation always repeated several times. K-means can also use various kinds of distances, with us as the most successful appear to be correlation and Euclidean distance.

Cluster can be either approximation signals (adjusted to the same length), frequency elements in the dictionary LZ77 compression, or the frequency of groups of characters derived from frequency analysis.

The goal of clustering is to divide the signals into two groups, normal and pathological. The data structure can be formed by more than just groups of two, so we tried the method with different numbers of clusters. Pathological and normal signals may comprise more than one cluster. In each cluster formed calculate frequency components of normal and pathological. Clusters with the highest frequency normal signal then mark as normal and the other pathological.

Since each signal is represented by multiple symptoms, k-means works in a multidimensional space. In order to render the results, we used principal components (PCA), which reduces symptoms of us. When two components, the individual signals can be plotted as points.

3. Results

The calculation of k-means was performed with different settings. As k-means determines the initial centers of clusters randomly we repeated the calculation of a total of 100 times and chose the result with the smallest total distance within clusters. This value will be considered as an objective function that determines the quality of the clusters.

Results of aggregation depends on the parameters. The key is to properly determine the length of the window and the overlay boundaries of intervals and the number and not the least measure of distance. We first clustered signals approximated by PAA.

Pathological signals are indicated as a positive result, normal signals are therefore negative finding. From the generated clusters, we determined the numbers of positive and negative elements, and compared them with the results of classification by pH values. Procedure with Euclidean distances repeated for different numbers of clusters. It turned out that the best number of clusters is 6, wherein the normal signal comprises 3 clusters, and clusters 3 pathological forms.

The best overall results were reported from the division to 6 clusters as shown in Table 1 and Figure 5. The sensitivity and specificity achieved were 63.6 % and 71% respectively.

Table 1. Results using SAX features and k-NN

	ClassN	ClassP
Normal	28	16
Pathological	147	361

4. Conclusion

In this work we have shown, that the clustering approach to the CTG analysis is a viable approach. Additionally using the features acquired by SAX which are robust, enables us to create classifier without input from clinicians.

Acknowledgements

This work was partially supported by the research project by Czech Grant Agency project number 14-28462P entitled Statistical methods of intrapartum CTG signal processing in the context of clinical information and by Grant No. NT11124-6/2010 funded by Ministry of Healthcare of the Czech Republic.

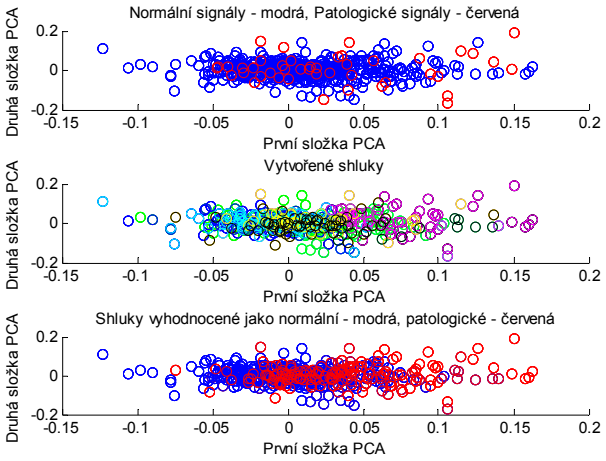

Figure 5. Results based on 6 clusters using SAX features.

References

[1] Dash S, Quirk JG, Djuric PM. Learning dependencies among fetal heart rate features using Bayesian networks. Engineering in Medicine and Biology Society (EMBC). 2012: 6204 -6207.

[2] Keith RD, Beckley S, Garibaldi JM, Westgate JA, Ifeachor AC, Greene KR. A multicentre comparative study of 17 experts and an intelligent computer system for managing labour using the cardiotocogram. Br J Obstet Gynaecol 1995; 102: 688-700.

[3] Ocak H. A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being. J Med Syst 2013; 37: 9913.

[4] Lazarsfeld PF. The Logical and Mathematical Foundations of Latent Structure Analysis, Measurement and Prediction, ed. Samuel A. Stouffer. NY: John Wiley & Sons pp. 362–412, 1950.

[5] Spilka, J, Georgoulas G, Karvelis P, Chudáček V, Stylios CD, & Lhotská L. Discriminating normal from "abnormal" pregnancy cases using an automated fhr evaluation method. Proc of 8th Hellenic Conference on AI. SETN 2014, Ioannina, Greece, 2014:521–531.

[6] Spilka, J, Georgoulas G, Karvelis P, Chudáček V, Stylios C D, Lhotská L. A three class treatment of the FHR classification problem using latent class analysis labeling. Proc.of 36th Annula International Conference of the IEEE Engineenig in Medicine and Biology Socierty, Chicago 2014

[7] Domingos, P. A few useful things to know about machine learning. Communications of the ACM, 55(10), 78-87.

[8] deHaan, J, van Bemmel J, Versteeg B, Veth A, Stolte L, Janssens J, Eskes T: Quantitative evaluation of fetal heart rate patterns. I. Processing methods. European Journal of Obstetrics and Gynecology and Reproductive Biology 1971; 1: 95–102.

[9] Yeh SY, Forsythe A, Hon EH. Quantification of fetal heart beat-to-beat interval differences. Obstet Gynecol 1973; 41: 355–363.

[10] Pardey J, Moulden M, Redman CWG. A computer system for the numerical analysis of nonstress tests. Am J Obstet Gynecol 2002;186: 1095–1103.

[11] Magenes G, Signorini MG, Arduini D. Classification of cardiotocographic records by neural networks. Proc IEEE-INNS-ENNS International Joint Conference on Neural Networks IJCNN 2000;.3:637-641.

[12] Task-Force. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Eur Heart J 1996;17: 354–381.

[13] Lempel A, Ziv J. On the complexity of finite sequences. IEEE Transactions on Information Theory. IT 1976;22: 75–81.

[14] de Sa JM. Pattern recognition: concepts, methods, and applications. Springer. 2001.

[15] Liu H, Motoda H. Computational methods of feature selection. CRC Press 2007.

[16] Guyon I, Elisseeff A. An introduction to variable and feature selection. The Journal of Machine Learning Research 2003; 3: 1157-1182.

[17] Guyon I, Gunn S, Nikravesh M, Zadeh L. Feature extraction. Foundations and applications of heart rate monitoring: nomenclature, interpretation, and general management principles. Obstet Gynecol 114: 192-202.

[18] Bernardes J, Moura C, de Sa JM, Pereira-Leite L. The Porto system for automated cardiotocographic signal analysis. J Perinat Med 1991; 19: 61–65.

Address for correspondence.

Václav Chudáček.
Technická 2, Praha 6, Czech Republic.