On the Early Detection of Perinatal Hypoxia with Information-Theory based Methods

Ricardo Santiago-Mozos[1], Beatriz García-Vizuete[1], José María Lillo-Castellano[1], José Luis Rojo-Álvarez[1], Carlos Martín-Caballero[2]

[1] Dep. of Signal Theory and Communications, Universidad Rey Juan Carlos, Madrid, Spain
[2] Dep. of Obstetrics and Gynecology, Hospital Universitario Fundación Alcorcón, Madrid, Spain

Abstract

Perinatal hypoxia is a severe condition that may harm fetus organs permanently. When the fetus brain is partially deprived from oxygen, the control of the fetal heart rate (FHR) is affected. We hypothesized that advanced processing of the FHR can reveal whether the fetus is under perinatal hypoxia.

We analyzed FHR morphology with normalized compression distance (NCD) that compares two arbitrary sequences and outputs their dissimilarity. This parameter-free measure exploits linear and non-linear relations in the data and allows the comparison of sequences of different sizes. It was applied to raw FHR sequences and to a set of statistics computed from them (e.g. moments on 5 minutes signal windows). We classified the cases from the NCD dissimilarity matrix by using a simple nearest neighbor classifier and leave-one-out cross-validation.

Best results in a database with 26 FHR recordings (13 controls and 13 cases) were provided by the central moment of order 3 calculated over sliding windows of 5 minutes on the interval from 4 to 3 hours to delivery. The resulting accuracy was 0.88 with sensitivity 0.92 and specificity 0.85.

1. Introduction

Perinatal hypoxia is caused by the lack of oxygenation at tissues and might cause serious sequels, such as brain or adrenal hemorrhage, necrotizing enterocolitis, delayed neurological development, mental handicap, seizures (West syndrome), or cerebral palsy [1]. The severity of the hypoxia is commonly quantified using the Apgar Score [2]. An score lower than 7 at five minutes after delivery is considered pathological, and gas analysis of umbilical cord is performed for confirmation, where a low pH gives evidence of metabolic acidosis.

Continuous electronic fetal monitoring, or Cardiotocography (CTG) consists in simultaneous evaluation of the Fetal Heart Rate (FHR) and the uterine activity [3, 4]. After CTG generalization, the gynecologists value basal FHR, its accelerations and decelerations in relation to uterine contractions, and beat-to-beat FHR variability [5]. The following signal types are considered clearly pathological: late decelerations, whose minimum has a delay of at least 30 seconds with respect to the peak of contractions; sustained bradycardia; low variability (less than 5 beats); and a "sine"-rhythm, which is characterized by a long-term variability but almost no variability in the short-term. However, visual interpretation of CTG has not shown advantages over intermittent auscultation in low-risk pregnancies [6]. In addition, visual interpretation of CTG has low specificity, and requires confirmation by invasive pH determination of scalp blood of the fetus, which is not always feasible [7]. Scalp blood pH values ≤ 7.20 are considered as risk of perinatal hypoxia. Therefore, gynecologists, by balancing the risk of hypoxia, indicate Cesareans, forceps and vacuum extraction more often than necessary [7].

Automatic analysis of CTG have also been proposed. For instance, it has been shown that automatic ST analysis combined with CTG increases the ability of obstetricians to identify hypoxia [8]. A system-identification approach to model FHR and uterine activity as an input-output system reported around 50% sensitivity, with 7.5% of false positives, 1h and 40 minutes before delivery [9]. Other non-invasive approaches have been proposed to complement CTG, such as Doppler velocimetry and pulse oximetry [10].

In this paper we analyze the readily available FHR to determine whether the fetus is suffering hypoxia. We decide using a nearest neighbor (NN) classifier using as *distance* a general information theory measure, the normalized compression distance (NCD) [11], related to Kolmogorov Complexity and mutual information [12]. This technique was successfully used for clustering the fetuses of a multicentre study with the aim of identifying the abnormal ones [13]. This paper builds up on that work and, as

NCD is only an approximation of the Kolmogorov Complexity, we compare the performance of NCD applied to raw series with its performance when applied to series of moments obtained from 5-minute sliding-windows.

2. Methods

This section reviews the methodology followed in this paper. Namely, we detail NCD, a generic dissimilarity measure for sequences, the statistical moments, and the classification procedure.

2.1. Normalized compression distance

In FHR classification, a typical approach is to extract several features from the time series and use them as input to the classifier [14]. We follow a different approach in this paper, and instead of extracting features, we use the dissimilarities of each time series with the ones in the training set as the input to our classifier. This type of learning is termed *dissimilarity learning* [15]. We select a dissimilarity measure that is based on the mutual information among the sequences, which (theoretically) includes any kind of relations among the sequences.

The Kolmogorov complexity, $K(\mathbf{s})$, of a sequence $\mathbf{s}$ is the length of the shortest binary program that produces $\mathbf{s}$ on an universal Turing machine [11]. $K(\mathbf{s})$ can be seen as the information of the sequence (or the information needed to generate it) [16]; $K(\mathbf{s}|\mathbf{t})$ is the length of the shortest program to produce $\mathbf{s}$ if $\mathbf{t}$ is given as an input; and $K(\mathbf{s}, \mathbf{t})$ is the length of the shortest program that generates $\mathbf{s}$, $\mathbf{t}$, and allows to separate them. Unfortunately, Kolmogorov Complexity is not computable.

NCD is a practical dissimilarity measure for sequences [11]. Given two sequences $\mathbf{s}_i, \mathbf{s}_j$, the $\text{NCD}(\mathbf{s}_i, \mathbf{s}_j)$ is defined as:

$$\text{NCD}(\mathbf{s}_i, \mathbf{s}_j) = \frac{C(\mathbf{s}_i, \mathbf{s}_j) - \min\{C(\mathbf{s}_i), C(\mathbf{s}_j)\}}{\max\{C(\mathbf{s}_i), C(\mathbf{s}_j)\}} , \quad (1)$$

where $C(\cdot)$ is the compression length in bits given by the selected compressor C ($C(\mathbf{s}_i)$ and $C(\mathbf{s}_i, \mathbf{s}_j)$ are, respectively, the number of bits needed to compress $\mathbf{s}_i$ and the concatenation of $\mathbf{s}_i$ and $\mathbf{s}_j$). C provides a computable approximation of the Kolmogorov Complexity. This normalized measure has a simple interpretation: the lower its value, the more similar the sequences (their high mutual information simplifies compressor task, which requires fewer bits). The normalization term in the denominator of (1) enables the comparison between sequences of different sizes.

2.2. Statistical moments

Statistical moments have a handful of advantages for time series. Moments are easy to compute and robust to signal loses, as we can just ignore unknown parts (no interpolation is needed as in frequency-related statistics). In addition, moments have low computational burden.

The raw moment of order n is defined as

$$M_n(\mathbf{s}) = \frac{1}{L} \sum_{k=1}^{L} s[k]^n , \quad (2)$$

where $s[k]$ is the k-th element of the sequence $\mathbf{s}$. Central moments are defined as:

$$\mu_n(\mathbf{s}) = \frac{1}{L} \sum_{k=1}^{L} (s[k] - M_1(\mathbf{s}))^n . \quad (3)$$

2.3. Classification and evaluation

Cross-validation is a common approach to estimate the expected accuracy when the number of instances available is low. In this paper we follow a leave-one-out cross-validation approach to evaluate performance of the different alternatives. The classification is done by a NN classifier which decides the label of the test series as the one that has closer NCD distance to it.

NCD dissimilarity is not necessarily symmetric. Therefore, in order to use it with the NN classifier, we tried two flavors of NCD dissimilarity measure: type *min*, which uses the minimum compression size of the concatenation of the sequences, i.e., $\min\{\text{NCD}(\mathbf{s}_i, \mathbf{s}_j), \text{NCD}(\mathbf{s}_j, \mathbf{s}_i)\}$; or type *mean*, which uses their mean, i.e., $0.5(\text{NCD}(\mathbf{s}_i, \mathbf{s}_j) + \text{NCD}(\mathbf{s}_j, \mathbf{s}_i))$.

3. Experiments

We first describe the database used in the experiments. Then, we examine the performance of NCD + NN on the raw FHR time series. Finally, we examine the performance of NCD + NN on the series that result of computing the moments on 5-minute sliding windows extracted from FHR series.

3.1. Data description

FHR records[1] were acquired with a Philips cardiotocograph for a total of 32 recordings, 15 controls and 17 cases (only 13 controls and 13 cases had signal in the $4 \leftrightarrow 3$ hours to delivery interval) . A case was declared whether: 1) the PH of the umbilical artery was $\leq$ 7.05; or 2) the

[1]Data are available from the website: http://sites.google.com/site/hufahypoxia.

APGAR score 5 minutes after delivery was ≤ 7 and a reanimation type III or greater was required.

Records have considerable variability both in start/ending times and pauses as labor duration vary. In addition, the cardiotocograph may be disconnected at any time for a number of reasons. Also, the signal is lost sometimes as the fetus and mother move. The cardiotocograph provides three signal qualities (lost, medium and high). We decide to consider the window between 4 to 1 hours before birth for our analysis, even though not all patients have signal along all this window. This window selection establishes a trade-off among number of patients and early detection of hypoxia.

3.2. Raw data analysis

We analyzed the raw sequences to investigate the attainable accuracy without any preprocessing step and without using prior knowledge about FHR signals and useful parameters. Our aim was to evaluate the ability of NCD to extract information on the series. We considered four time intervals: 1) 4 to 3 hours to delivery (13 cases and 13 controls in the database); 2) 3 to 2 hours to delivery (13 cases and 14 controls); 3) 2 to 1 hours to delivery (15 cases and 16 controls); and 4) 4 to 1 hours to delivery (15 cases and 16 controls). In addition, we analyzed three types of sequences: a) only high quality signal; b) high and medium qualities signal; c) including all signal qualities. The final sequences were the raw sequences with the non-considered qualities removed. NCD was computed using the software of their authors [17]. We tried three different compressor architectures (zip, bzip2 and lzma).

Best results are summarized in Table 1, where we see that high quality signal and the interval from 4 to 3 hours before delivery were the best for predictions. In addition, we see for the same time interval that prediction using all the signal is better than using only high and medium qualities, which shows that taking into account lost signals, which may occur when the fetus moves, might increase prediction accuracy.

3.3. Moments on sliding windows

In this experiment we first classified the sequences using statistical moments. We considered raw and central moments of orders $n = \{1, 2, \ldots, 10\}$. We standardized the moments applying the n-th root and making them zero-mean, unit-variance. We tried NN and Support Vector Machines (SVM) as classifiers. We trained classifiers with all moments as input and using backward selection for selecting best moments. Parameter tuning and feature selection were done inside the cross-validation loop. Best result was provided by a SVM with radial basis function kernel and backward selection, which gave an accuracy of 0.69 (18 out of 26) with high quality signal in the $4 \leftrightarrow 3$ hours to delivery interval.

Table 1. NCD and NN classifier best results for raw signals. Interval expresses the signal interval in hours to delivery. Acc, Sen and Spe stand for accuracy, sensitivity and specificity, respectively. Matrix shows the matrix type used in the computations.

Interval	Acc.	Sen.	Spe.	Compressor	Matrix
Only high signal quality					
$4 \leftrightarrow 3$	**0.73**	0.69	0.77	zip	min
$3 \leftrightarrow 2$	0.63	0.57	0.69	bzip2	min
$2 \leftrightarrow 1$	0.58	0.75	0.40	lzma	sum
$4 \leftrightarrow 1$	0.66	0.82	0.47	zip	min
High and medium signal qualities					
$4 \leftrightarrow 3$	**0.58**	0.62	0.54	zip	min
$3 \leftrightarrow 2$	0.56	0.79	0.31	bzip2	min
$2 \leftrightarrow 1$	0.55	1.0	0.07	lzma	sum
$4 \leftrightarrow 1$	0.56	0.59	0.53	lzma	min
All signal qualities					
$4 \leftrightarrow 3$	**0.66**	0.77	0.54	zip	min
$3 \leftrightarrow 2$	0.56	0.14	1.0	lzma	min
$2 \leftrightarrow 1$	0.53	0.76	0.27	lzma	min
$4 \leftrightarrow 1$	0.59	0.59	0.60	zip	min

As the combinations of statistical moments for FHR classification described above reported worse accuracies than NCD on the raw signals, we tried NCD to empower simple moments on sliding windows, e.g., instead of using the mean of the whole signal as a descriptor, we created a sequence with the means obtained in the sliding windows. We computed all the sequences for raw and central moments of orders $n = \{1, 2, \ldots, 10\}$. The sequences were later applied the transformation $\bar{\mathbf{s}}_i = \sqrt[n]{\mathbf{s}_i}/A_n$, where A_n was the maximum value of the sequences in all patients for each moment. Then, NCD pairwise distances were obtained for $\bar{\mathbf{s}}_i$ and accuracies were estimated using leave-one-out cross-validation with a nearest neighbor classifier.

Best results were obtained with high and medium signal qualities and 5 minutes windows with 2 minutes-overlap. These results are summarized in Table 2. The best predictive interval was the 4 to 3 hours to delivery. The best accuracy for individual moments gave an accuracy of 0.88 (23 out of 26), a sensitivity of 0.92 (12 out of 13) and a specificity of 0.85 (11 out of 13).

4. Conclusion

The NCD analysis of the readily available FHR traces might help obstetricians when deciding about fetuses hypoxia. We showed how to apply NCD to raw signals and

Table 2. NCD and NN classifier best results for moments in 5-minute sliding-windows signals. M indicates the statistical moment used.

Int.	Acc.	Sen.	Spe.	M	Comp.	Matrix
$4 \leftrightarrow 3$	**0.88**	0.92	0.85	μ_3	lzma	min
$3 \leftrightarrow 2$	0.70	0.64	0.77	μ_2	bzip2	min
$2 \leftrightarrow 1$	0.77	0.81	0.73	M_4	zip	sum
$4 \leftrightarrow 1$	0.81	0.82	0.80	M_4	lzma	sum

how to use it to analyze sequences derived from them, for example, sequences of statistical moments computed on sliding windows. 88% accuracy 3 hours before delivery is a promising result because we are identifying stressed fetuses which were not considered suspicious of hypoxia at that labor stage. In addition, the methods proposed are simple to understand, simple to apply and generally applicable to other time series classification problems. As future work, a further study with more patients should be performed to open the application of these techniques to the industry.

Acknowledgements

This work has been partly supported by the Spanish Government under the Research Project TEC2010-19263. Author R. Santiago-Mozos is supported by the Juan de la Cierva Program of the Spanish Ministry of Science and Innovation (Ref: JCI-2011-11150).

References

[1] Leuthner SR, Das UG. Low Apgar scores and the definition of birth asphyxia. Pediatric clinics of North America June 2004;51(3):737–45.

[2] Apgar V. A Proposal for a new method of evaluation of the newborn infant. Anesthesia Analgesia 1953;32:260–267.

[3] Hon EH. The Electronic Evaluation of the Fetal Heart Rate: Preliminary Report. Obstetrical Gynecological Survey 1958;13:654–656.

[4] Hammacher K, Huter K, Bokelmann J, Werners P. Fetal heart frequency and perinatal condition of the fetus and newborn. Gynecologica 1968;166:349–360.

[5] Ayres-de Campos D, Bernardes Ja. Twenty-five years after the FIGO guidelines for the use of fetal monitoring: time for a simplified approach? International journal of gynaecology and obstetrics the official organ of the International Federation of Gynaecology and Obstetrics July 2010; 110(1):1–6.

[6] Devane D, Lalor JG, Daly S, McGuire W, Smith V. Cardiotocography versus intermittent auscultation of fetal heart on admission to labour ward for assessment of fetal wellbeing. Cochrane database of systematic reviews Online January 2012;2:CD005122.

[7] Tasnim N, Mahmud G, Akram S. Predictive accuracy of intrapartum cardiotocography in terms of fetal acid base status at birth. Journal of the College of Physicians and Surgeons Pakistan JCPSP October 2009;19(10):632–5.

[8] Amer-Wåhlin I, Hellsten C, Noren H, Hagberg H, Herbst A, Kjellmer I, Lilja H, Lindoff C, Mansson M, Martensson L. Cardiotocography only versus cardiotocography plus ST analysis of fetal electrocardiogram for intrapartum fetal monitoring: a Swedish randomised controlled trial. The Lancet August 2001;358(9281):534–538.

[9] Warrick PA, Hamilton EF, Precup D, Kearney RE. Classification of normal and hypoxic fetuses from systems modeling of intrapartum cardiotocography. IEEE transactions on bio medical engineering April 2010;57(4):771–9.

[10] Siristatidis C, Salamalekis E, Kassanos D, Loghis C, Creatsas G. Evaluation of fetal intrapartum hypoxia by middle cerebral and umbilical artery Doppler velocimetry with simultaneous cardiotocography and pulse oximetry. Archives of gynecology and obstetrics December 2004;270(4):265–70.

[11] Li M, Chen X, Li X, Ma B, Vitanyi PMB. The Similarity Metric. IEEE Transactions on Information Theory December 2004;50(12):3250–3264.

[12] Cover TM, Thomas JA. Elements of Information Theory. 2 edition. John Wiley & Sons, 2006. ISBN 0-471-24195-4.

[13] Costa Santos C, Bernardes J, Vitanyi PMB, Antunes L. Clustering Fetal Heart Rate Tracings by Compression. In Computer-Based Medical Systems, 2006. CBMS 2006. 19th IEEE International Symposium on. 2006; 685–690.

[14] Signorini MG, Magenes G, Cerutti S, Arduini D. Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings. IEEE transactions on bio medical engineering March 2003;50(3):365–374.

[15] Pekalska E, Duin RP. Dissimilarity representations allow for building good classifiers. Pattern Recognition Letters June 2002;23(8):943–956.

[16] Kolmogorov AN. Three approaches to the definition of information. Problemy Peredachi Informatsii 1965;1(1):3–11.

[17] Cilibrasi R, Cruz AL, de Rooij S, Keijzer M. CompLearn. `http://www.complearn.org`, 2008.

Address for correspondence:

Ricardo Santiago-Mozos
Universidad Rey Juan Carlos, 28943 MADRID, SPAIN
ricardo.santiago.mozos@urjc.es