FETAL HEART RATE ANALYSIS BY HIERARCHICAL DIRICHLET PROCESS MIXTURE MODELS

Kezi Yu,* J. Gerald Quirk,** and Petar M. Djurić*

*Department of Electrical and Computer Engineering **Department of Obstetrics/Gynecology, Stony Brook University Hospital Stony Brook, NY 11794, USA Email: {kezi.yu,petar.djuric}@stonybrook.edu, J.Gerald.Quirk@stonybrookmedicine.edu

ABSTRACT

In this paper, we propose to analyze fetal heart rate (FHR) signals by hierarchical Dirichlet process (HDP) mixture models. We investigate whether the clustering results of real-world FHR time series obtained by these models are informative in terms of determining the health status of a fetus. The FHR signals are divided into two groups, healthy and unhealthy, according to the umbilical arterial blood pH values of the fetuses. We computed the frequencies of clusters appearing in each of the groups, and applied the Mann-Whitney U test to compare the frequencies. The results showed that the frequencies of appearance of certain clusters are statistically significantly different across the two groups. This indicates that certain clusters may relate to pathological fetal heart rate patterns.

Index Terms— Fetal heart rate, Hierarchical Dirichlet process, mixture model

1. INTRODUCTION

Childbirth is a natural process with outcomes generally being good. However, neonatal morbidity and mortality still occur for various reasons and sometimes due to lack of proper treatment of the mother and the fetus. The standard Apgar score system evaluates the newborn baby on five criteria by physicians' observations. It has been in use for assessment of the neonatal health status and the need for medical management for more than 50 years. Determining the status of a fetus, however, is much more challenging. One approach to such assessments is by analyzing fetal heart rate (FHR) signals. It has been argued that antepartum and intrapartum FHR signals provide timely information of the fetal well-being. The National Institute of Child Health and Human Development (NICHD)

and International Federation of Gynecology and Obstetrics (FIGO) both published guidelines on FHR evaluation [1, 2].

Most of the time the evaluations of FHR signals are performed visually by physicians and consequently are subjective. With the goal of avoiding subjectivity, various automated methods for analysis and classification have been proposed. In [3], the authors employed signal processing techniques and neural networks to detect the FHR patterns defined in [1]. The results are promising since certain patterns are closely related to fetal hypoxia, which is the major threat to fetal health. In [4], generative models were used to classify FHR signals into two categories. Satisfactory specificity and sensitivity results based on pH values were obtained. Other state-of-the-art signal processing methods have also been proposed. Adaptive multiscale complexity analysis was performed on FHR data in [5]. There, the goal was to detect acidosis in fetuses based on characterization parameters. In [6], entropy-based estimators were applied to identify fetal distress. The results showed that the approximate entropy index was significantly distinguishable between suffering and normal fetuses.

The NICHD guidelines notwithstanding, the patterns in FHR signals are rather irregular and hard to interpret. In other words, the criteria in [1] are too strict and deterministic for the detection of unusual patterns. Therefore, a more flexible, and data-driven classification method is desirable. Dirichlet process mixture models (DPMMs) are well-suited for such classification tasks. DPMMs fall in the class of nonparametric Bayesian models. These models do not specify the number of clusters beforehand [7]. Instead, this number is inferred from the data. The number of clusters theoretically may grow to infinity as the number of data also increases to infinity. DPMMs allow for new classes of patterns to be discovered as more data are processed.

A more recent class of models are the hierarchical Dirichlet process (HDP) models [8]. They enable cluster sharing across datasets. This is particularly useful for topic modeling in document processing. Clearly, topics are often

This work has been supported by NIH under Award 1R21HD080025-01A1.

shared in different articles from a corpus of articles, and one should somehow benefit from this feature while performing classification. Similarly, the patterns in FHR signals are shared across different fetuses. This makes HDP mixture models a promising tool for analyzing FHR signals obtained from different fetuses jointly. In this paper, we propose to exploit the potential of HDP mixture models in FHR signal classification. By processing real-world FHR series, we show that these models hold a high promise for accurate classification.

The main contribution of this paper is twofold. We propose the use of HDPs for FHR classification and show on real data that the methodology has the potential to be accurate. We also propose a set of features that when used in conjunction with HDPs shows good classification performance.

The paper is organized as follows. In the next section, we provide a brief background on data acquisition and preprocessing as well as HDPs. In Section 3, we propose the details of feature extraction from FHR signals and clustering by HDP mixture models. The results of the clustering are shown in Section 4. Finally, in the last section we provide conclusions.

2. BACKGROUND

2.1. Data acquisition and preprocessing

In our work we used the database from the Czech Technical University (CTU) in Prague and the University Hospital in Brno (UHB) (put together by Václav Chudáček and his collaborators). The database is publicly available online. It contains 552 cardiotocography (CTG) recordings, each containing an FHR time series and uterine contraction (UC) signals, both sampled at 4 Hz. Additional information, including maternal data, delivery data, fetal data and fetal outcome data are also provided in the database. More details about it can be found in [9, 10].

It is inevitable that data acquired in realistic settings have artifacts and, thus, have to be preprocessed before being analyzed. The artifacts in FHR signals are generally caused by maternal and fetal movements or displacements of the transducer used in the acquisition. There are two types of artifacts, either the measured samples are incorrect or they are simply missing (e.g., they are equal to zero).

In practice, any successive five samples with differences smaller than 10 bpm among them are considered stable (provided that they are not zero). If these differences are between 10 and 25 bpm, they are not stable, and if they are greater than 25 bpm, they are considered artifacts. When the samples are artifacts or are simply missing and their duration is less than 15 seconds, they are substituted by linear interpolation; otherwise the bad samples are simply discarded without interpolation. For interpolation, we used



Fig. 1. Comparison between a raw FHR signal and the signal obtained after its preprocessing.

the piecewise cubic Hermite polynomial. Figure 1 shows a comparison of one FHR time series before and after preprocessing.

2.2. Methods

The HDP model was proposed in [11]. It was obtained as a recursive construction of models based on the Dirichlet process. In brief, one defines a global measure G_0 , which is distributed as a Dirichlet process, i.e.,

$$G_0|\gamma, H \sim DP(\gamma, H),$$
 (1)

where γ is a concentration parameter and H is a probability measure. Then, given G_0 , we have conditionally independent random measures G_i defined by

$$G_i | \alpha_0, G_0 \sim DP(\alpha_0, G_0), \tag{2}$$

where α_0 is also a concentration parameter. One can use metaphors to decribe the HDP model. One of them is the Chinese restaurant franchise (CRF), which we adopt here.

Consider a Chinese restaurant franchise with several restaurants, all of which share the same set of dishes. Each restaurant has infinite number of tables. We let θ_{ji} denote customer *i* in restaurant *j*, and ψ_{jt} represent the dish served at table *t* in restaurant *j*. We also let $\phi_1, \phi_2, \ldots, \phi_K$ denote *K* iid random variables distributed according to the base measure *H*. This measure is the source of the global dish menu.

Next, we need notation for counts of customers on each table, and tables served with each dish. Let n_{jtk} denote the number of customers in restaurant j at table t served with dish k, and m_{jk} be the number of tables in restaurant j serving

dish k. The conditional distribution of customer θ_{ji} given $\theta_{j1}, \ldots, \theta_{j,i-1}$, a scaling parameter α_0 , and a distribution G_0 is given by

$$\theta_{ji}|\theta_{j1},\dots,\theta_{j,i-1},\alpha_{0},G_{0} \\ \sim \sum_{t=1}^{m_{j\cdot}} \frac{n_{jt\cdot}}{i-1+\alpha_{0}} \delta_{\psi_{jt}} + \frac{\alpha_{0}}{i-1+\alpha_{0}} G_{0}, \quad (3)$$

where n_{jt} . represents the marginal count of the number of customers in restaurant j at table t, m_j . represents the number of tables in restaurant j, and $\delta_{\psi_{jt}}$ is a probability measure concentrated at ψ_{jt} . Recall that the distribution G_0 is obtained by (1). This model describes how the *i*th customer is seated in any restaurant, either by choosing an already occupied table with probability proportional to the number of customers at that table, or preferring a new table with probability $\alpha_0/(i - 1 + \alpha_0)$.

Now we consider the distribution of ψ_{jt} which is drawn from the global menu. The distribution is given by [11]

$$\psi_{jt}|\psi_{11},\psi_{12},\ldots,\psi_{jt-1},\gamma,H$$

$$\sim \sum_{k=1}^{K} \frac{m_{\cdot k}}{m_{\cdot \cdot}+\gamma} \delta_{\theta_k} + \frac{\gamma}{m_{\cdot \cdot}+\gamma}H, \quad ^{(4)}$$

where the dot notation is defined similarly as in the previous equation. As before, the probability of each dish at a table is determined by the number of tables already serving the dish. This completes the Chinese restaurant franchise metaphor of HDP. Note that the dishes are shared across all the restaurants, which is the key property of HDPs.

Finally, we briefly describe HDP mixture models (HDPMMs). Once the dishes ψ_{jt} are drawn for each table, we need to draw for each customer the actual plate served to the customer. It is drawn from yet another distribution, and we have

$$x_{jit}|\psi_{jt} \sim F(\psi_{jt}) \tag{5}$$

where x_{jit} is the plate of the *i*th customer in the *j*th restaurant sitting at table *t*, $F(\psi_{jt})$ is a distribution of the plate given the dish ψ_{jt} . In signal processing, the plates x_{ji} represent observations.

In our work, we have predefined segments of data. Each segment belongs to a certain cluster of similar segments. The cluster of segments are all "customers seated" on a same table. The number of different clusters is not predetermined. The objective of our work is to determine if we can find types of clusters that provide information about the well-being of fetuses.

3. EXPERIMENTAL SETTINGS

In this section, we describe how the FHR signals are processed after preprocessing and how they are classified by the HDPMMs. In our description, we continue to use the CRF metaphor. Consider each FHR time series to be a restaurant. We divide each FHR series into small segments, where each segment represents a customer. According to the HDP mixture models, the segments form a mixture of clusters (equivalent to dishes in the CRF metaphor), and the clusters are shared among all the FHR series.

Before we proceed, we need to represent each segment with a small set of features. These features are then used for clustering. We reiterate that the number of clusters across all the FHR series is not predetermined. First we explain the extraction of features and then how we implement the HDPMMs.

3.1. Feature Extraction

We set the length of one segment to be 40 samples (i.e., the segments were 10 seconds long). After preprocessing, we computed the baseline of the FHR signal by a 5-min median filter. Upon estimating the baseline, we subtracted it from the FHR series. Therefore, the remaining signal reflects the variation of the heart rate around the baseline. In our work we did not use the shape of the baseline as part of our classification procedure.

We experimented with different types of features, including the mean, the variance, and the parameters of an autoregressive exogenous (ARX) model of the segments. Since we intended to model the FHR patterns defined in [1] (acceleration, deceleration, etc), we chose the order of the time component in the ARX model to be 2. Finally, we worked with four parameters of the ARX model and the mean value of the segment.

The ARX model of the segment had the form

$$x_t = ax_{t-1} + c_2t^2 + c_1t + c_0, (6)$$

where t is the time index of the observed sample x_t in the segment, and a, c_0 , c_1 , and c_2 are unknown parameters of the model, which are idiosyncratic for the segment. We estimated these parameters by using MMSE estimators. The MMSE estimates and the mean values were computed from the residual signal (after the baseline was subtracted) instead of the original FHR signal.

3.2. Clustering by HDPMM

In the experiments, we used the last 30 minutes of data. Typically, the time series long before delivery are stationary, and the data near delivery are more informative in predicting fetal health. We selected 60 FHR signals, 30 of which were from healthy and 30 from unhealthy fetuses. The classification of the fetuses was based on the umbilical arterial blood pH. After feature extraction, we had 60 sets of data, each set having 180 feature vectors of dimension five. We set the concentration parameters γ to 10 and α_0 to 5.



Fig. 2. The classification result of one FHR time series. Each color represents a cluster.

The mixture was assumed to be distributed according to a multivariate Gaussian distribution, with unknown mean vector and unknown covariance matrix. In order to simplify the computations, the base measure H was assumed to be the normal-inverse Wishart distribution, which is the conjugate prior of the multivariate normal distribution. This allowed us to implement a collapsed Gibbs sampling scheme for inference, as proposed in Section 5.1 in [11]. With this scheme, one does not need to draw samples of the unknown means and covariances of the multivariate Gaussians that represent the clusters. This approach altogether reduced the variance of the obtained results.

4. RESULTS

We ran the Gibbs sampler and after 50 iterations, four clusters emerged from the data set. One example of clustering is shown in Fig. 2. Each color in the graph represents a different cluster. It is difficult to interpret each cluster visually, and in particular, somehow connect it to a diagnosis.

In order to determine if the clusters have information about the well-being of the fetus, we counted the number of segments in each cluster of each FHR series. We wanted to investigate whether the frequency of occurrence of each cluster is different between healthy and unhealthy fetuses. We computed the mean and 95% confidence interval of the numbers of segments in each cluster of the healthy and unhealthy group, respectively. The results are shown in Fig. 3. We also ran a Mann-Whitney U test to determine whether each cluster is significantly different across the two groups. The results are shown in Table 1. The significance level was set to p = 0.05.

The statistical analysis of the clustering results showed that three out of the four clusters had significant differences. Cluster 2 appeared more often in the FHR signals of healthy fetuses, whereas clusters 3 and 4 were seen more frequently in unhealthy fetuses.



Fig. 3. Mean and 95% confidence intervals of 4 clusters. Clusters marked with star are statistically significantly different.

Table 1. Results of Mann-Whitney U test.

Cluster index	1	2	3	4
p-value	0.054	0.007	0.018	0.024
Significant	No	Yes	Yes	Yes

5. CONCLUSION

In this paper, we proposed the use of hierarchical Dirichlet process mixture models for classification of FHR signals. After preprocessing, baseline estimation and its subtraction from the FHR signal, we applied ARX modeling of segments of the signals of duration of 10 seconds. The parameters of these signals were modeled as multivariate Gaussians with unknown means and covariances. The hierarchical Dirichlet process mixture model of the feature set produced four different clusters. According to the statistical results of these clusters, the numbers of segments in three of the four clusters are significantly different between healthy and unhealthy fetuses (with p values less than 0.05). The results suggest that this approach to classification of fetal heart series has strong potential and that it deserves further investigation.

6. REFERENCES

- G. A. Macones et al., "The 2008 National Institute of Child Health and Human Development Workshop report on electronic fetal monitoring: Update on definitions, interpretation, and research guidelines," *Journal of Obstetric, Gynecologic, & Neonatal Nursing* vol. 37(5), pp. 510-515, 2008.
- [2] G. A. Rooth, A. Huch, and R. Huch. "FIGO news: Guidelines for the use of fetal monitoring," *International Journal of Gynecology & Obstetrics* vol. 25, pp. 159-67, 1987.
- [3] P. Warrick, E. Hamilton, and M. Macieszczak, "Neural network based detection of fetal heart rate patterns," *Proceedings of the International Joint Conference of Neural Networks*, Montreal, Canada, 2005.
- [4] S. Dash, J. G. Quirk, and P. M. Djurić, "Fetal heart rate classification using generative models," *IEEE Transactions on Biomedical Engineering*, vol. 61(11), pp. 2796-2805, 2014.
- [5] H. Helgason, P. Abry, P. Gonçalves, C. Gharib, P. Gaucherand, and M. Doret, "Adaptive multiscale complexity analysis of fetal heart rate," *Biomedical Engineering*, vol. 58(8) pp. 2186-2193, 2011.
- [6] M. Ferrario, M. G. Signorini, G. Magenes, and S. Cerutti, "Comparison of entropy-based regularity estimators: Application to the fetal heart rate signal for the identification of fetal distress," *IEEE Transactions on Biomedical Engineering*, vol. 53(1), pp. 119-125, 2006.
- [7] S. J. Gershman and D. M. Blei, "A tutorial on Bayesian nonparametric models," *Journal of Mathematical Psychology* vol. 56(1), pp. 1-12, 2012.
- [8] Y. W. Teh and M. I. Jordan, "Hierarchical Bayesian nonparametric models with applications," *Bayesian Nonparametrics*, vol. 1, 2010.
- [9] V. Chudáĉek, J. Spilka, M. Burša, P. Janku, L. Hruban, M. Huptych, and L. Lhotská, "Open access intrapartum CTG database," *BMC Pregnancy and Childbirth*, vol. 14(1), 2014.
- [10] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101(23), pp. e215-e220, 2000.
- [11] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *Journal of the American Statistical Association*, vol. 101(476), 2006.