Hierarchical Dirichlet Process Hidden Markov Model for unsupervised learning from bioacoustic data

Bartcus Marius BARTCUS@UNIV-TLN.FR
Aix Marseille Université, CNRS, ENSAM, LSIS, UMR 7296, 13397 Marseille, France
Université de Toulon, CNRS, LSIS, UMR 7296, 83957 La Garde, France

Faicel Chamroukhi CHAMROUKHI@UNIV-TLN.FR
Aix Marseille Université, CNRS, ENSAM, LSIS, UMR 7296, 13397 Marseille, France
Université de Toulon, CNRS, LSIS, UMR 7296, 83957 La Garde, France

Abstract

Hidden Markov Models (HMM) are one of the most used models in statistics and machine learning for modeling sequential data. One of the main issues in HMM is the one of selecting the number of hidden states required for the expectation-maximization (EM) learning scheme. The infinite Hidden Markov Model (IHMM) is a Bayesian non-parametric alternative for standard parametric HMMs that offers a principled way to tackle this challenging problem by relying on a Hierarchical Dirichlet Process (HDP) prior. In this paper, we present an application of the the Hierarchical Dirichlet Process for Hidden Markov Model (HDP-HMM) to a challenging problem of unsupervised learning from bioacoustic data. We investigate two different approaches, the first one uses Gibbs sampling and the second one uses Beam sampling. The problem for humpback whale song decomposition consists in simultaneously finding the structure of possible hidden whale song units and automatically inferring the unknown number of the hidden units from the data. The considered data are Mel Frequency Cepstral Coefficients (MFCC) of recording of bioacoustic signals. The experimental results show the good performance of the proposed Bayesian non-parametric approach.

1. Introduction

Hidden Markov Model (HMM) (Rabiner, 1989) is one of the most used models in statistics and machine learning for the sequential and time series data. However, one main issue in standard HMM is the one of selecting the number of hidden states. This is the problem of model selection; the number of states being needed to be known before learning the model. The Bayesian Non-Parametric (BNP) approach (Robert, 1994; Hjort et al., 2010) gives a good alternative for model selection. The infinite Hidden Markov Model (IHMM) introduced by (Beal et al., 2002) is a Bayesian non-parametric extension of the standard finite state HMM by providing a principled way to inference the number of states from the data in an automatic way as the learning proceeds. They rely on Hierarchical Dirichlet Process (HDP) to define a prior over the transition matrix as developed by (Teh et al., 2006). This model is known as the hierarchical Dirichlet process for the Hidden Markov Model (HDP-HMM) (Teh et al., 2006). HDP-HMM infers the posterior distribution over the number of states. However the basic HDP-HMM has the limitation of an inadequate modeling of the temporal persistence of states (Fox et al., 2008). This problem has been addressed in Fox et al. (2008). Another solution for the inference of the hidden Markov model in the infinite scenario was proposed by (Van Gael et al., 2008) where the Beam sampling is used. The Beam sampling was seen to have better performance and more robust than Gibbs sampling.

In this work, we rely on this Bayesian Non-Parametric formulation for the Hidden Markov Model and present an application of the HDP-HMM to a challenging problem of unsupervised learning from bioacoustic data. We investigate two different approaches, the first one uses Gibbs sampling and the second one uses Beam sampling. The paper is organized as follows: First, we give a brief introduction of the finite Hidden Markov Model in Section 2, then Section 3 briefly discusses previous work on infinite Hidden Markov Models, the background for the Hierarchical Dirichlet Process, the model and the inference algorithm are described in this section. Then, section 4 presents experimental results

Proceedings of the 31^{st} International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32.

for the humpback whale song data, where we have been treated the two approaches for learning the infinite HMM by using the Hierarchical Dirichlet Process: the Gibbs sampling and the Beam sampling.

2. Hidden Markov Model

The finite Hidden Markov Model (HMM) (Rabiner, 1989; Derrode & Pieczynski, 2006) is well adapted to sequential data. It assumes that the observed sequence $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ where $\mathbf{x}_t \in \mathbb{R}^d$ is the multidimensional observation at time t, is governed by a hidden state sequence $\mathbf{z} = (z_1, \ldots, z_T)$, where z_t takes its values in a finite set $\{1, \ldots, K\}$. The HMM is described by the initial state distribution $\pi_{1i} = p(z_1 = i)$, the transition matrix which can be denoted by $\boldsymbol{\pi}$ with elements $(\pi_{ij} = p(z_t = j|z_{t-1} = i))$ and the parameters $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1 \ldots \boldsymbol{\theta}_K\}$ of the parametric conditional probability densities of the observed data $p(\mathbf{x}_t|z_t = k; \boldsymbol{\theta}_k)$ (the emission densities).

Given the parameters of the HMM $\{\boldsymbol{\pi}, \boldsymbol{\Theta}\}$, the joint distribution of the hidden states $\mathbf{z}$ and observations $\mathbf{X}$ can be written:

$$p(\mathbf{z}, \mathbf{X}|\pi_1, \boldsymbol{\pi}, \boldsymbol{\Theta}) = p(z_1)p(\mathbf{x}_1|z_1) \prod_{t=2}^{T} p(z_t|z_{t-1})p(\mathbf{x}_t|z_t) \tag{1}$$

In the equation (1) the distribution of state z_{t-1} conditioned on z_t denotes the transition probability and respectively the distribution of $\mathbf{x}_t$ conditioned on state z_t, $p(\mathbf{x}_t|z_t)$ denotes the emission distribution. The model is usually learned with the EM algorithm with the forward-backward recursion (Bauch-Welch algorithm) (Rabiner & Juang, 1993; 1986) that estimates the transition matrix $\boldsymbol{\pi}$, as well as the emission matrix $\boldsymbol{\Theta}$.

However, HMM have the limitation of model selection, the number of states K being necessary known a priori. Therefore, the existence of the Infinite Gaussian Mixture Model (Rasmussen, 2000) where the number of classes are estimated automatically, makes natural to have the necessity of the infinite HMM (IHMM) where the number of states will be inferred during the learning algorithm.

3. Infinite Hidden Markov Model

Bayesian Non-Parametric (BNP) (Robert, 1994; Hjort et al., 2010) alternative offers a principled way to tackle the challenging problem of model selection. (Beal et al., 2002) shows the possibility of extending the Hidden Markov Model by having a possible infinite number of hidden states, were the theory of Dirichlet Process (DP)(Antoniak, 1974; Ferguson, 1973) was used to define priors over the transition matrix. Because the transitions of states is given independent priors, there is no coupling across transitions between different states (Beal et al., 2002), therefore DP (Ferguson, 1973) is not sufficient to extend HMM to an infinite model. (Teh et al., 2006) shows in more detail the Hierarchical Dirichlet Process (HDP) for the Hidden Markov Model (HDP-HMM).

3.1. Hierarchical Dirichlet Process

Hierarchical modeling is an important tool for Bayesian statistics, where the parameters are sampled according to some distributions, involving other parameters (named hyper-parameters). By adding distributions over the hyper-parameters the probabilistic models becomes more richer. This kind of recursion gives the idea of a hierarchical model.

Hierarchical Dirichlet Process (HDP) extends DP where groups of data are generated by unique generative process (each group of data has it's own model). The HDP consists of a set of DPs G_j coupled through a base DP G_0, that can be interpreted as a mean of G_j. The basic generative process for HDP is given by the equation (2).

$$\begin{aligned} G_0|\gamma, \mathcal{H} &\sim \text{DP}(\gamma, \mathcal{H}) \\ G_j|\alpha, G_0 &\sim \text{DP}(\alpha, G_0),\ \forall j = 1 \ldots J \end{aligned} \tag{2}$$

where J is the number of groups of data. The HDP is then used for the prior distribution over the parameters of the mixture. Supposing for each of the group of data J, the iid parameters $\theta_{j1}, \theta_{j2}, \ldots, \theta_{jt}$ corresponding to a single observation x_{ji} the likelihood of the data by using the HDP is then given by:

$$\begin{aligned} \theta_{jt}|G_j &\sim G_j \\ \mathbf{x}_{jt}|\theta_{jt} &\sim F(\theta_{jt}) \end{aligned} \tag{3}$$

where $F(\theta_{jt})$ is a data likelihood (e.g. a Gaussian distribution).

By giving a stick-breaking construction and replacing the DP, the Hierarchical Dirichlet Process (2) becomes:

$$\begin{aligned} G_0 &= \sum_{k=1}^{\infty} \beta_k \delta_{\boldsymbol{\theta}_k} \\ G_j &= \sum_{k=1}^{\infty} \pi_{jk} \delta_{\boldsymbol{\theta}_k} \end{aligned} \tag{4}$$

where the parameters of the model are supposed to be drawn according to a conjugate distribution [1] $\boldsymbol{\theta}_k \sim \mathcal{H}$ independently. For the Gaussian mixtures the parameters mean and covariance matrix are drawn according to the Normal-Inverse-Wishart distribution $\mathcal{NIW}(\mu_0, \kappa_0, \nu_0, \Lambda_0)$ where

[1] In Bayesian statistics, the conjugate distributions are when the prior distribution $p(\boldsymbol{\theta})$ and the posterior distributions $p(\mathbf{X}|\boldsymbol{\theta})$ are in the same family.

the hyper-parameters describes the shapes and the position for each mixture densities: μ_0 describing where the mean of the mixtures should be, κ_0 the number of pseudo-observations supposed to be attributed and the hyper-parameters ν_0, Λ_0 being similarly for the mixture densities of the covariance matrix (Wood & Black, 2008). More details on this conjugate prior over the parameters of the GMM is given in (Andrew Gelman & Rubin, 2003). Also note in the HPD with stick-breaking construction equation (4) the hyper-parameter $\boldsymbol{\beta} = \{\beta\}_{k=1}^{\infty} \sim \text{GEM}(\gamma)$ where $\text{GEM}(\gamma)$ is the stick-breaking construction for the DP (Sethurman, 1994), the process being described by the following steps:

1. Start with stick with length equal to 1.
2. Sample $\beta_1 \sim \text{Beta}(1, \gamma)$ where Beta meaning Beta distribution.
3. Break the stick at β_1 and set π_1 to the length of the stick on the left.la
4. Take the stick on the right and sample $\beta_2 \sim \text{Beta}(1, \gamma)$. Breaking the stick β_2, π_2 will be set.
5. Continue the process infinite times.

3.2. Model Definition

Hierarchical Dirichlet Process gives the possibility to have distributions over hyper-parameters by making the models more flexible. The coupling between transition matrix makes possible to give a higher level to DP prior over the parameters.

$$
\begin{aligned}
\boldsymbol{\beta} &\sim \text{DP}(\gamma/K, \ldots, \gamma/K) \\
\boldsymbol{\pi}_j &\sim \text{DP}(\alpha\boldsymbol{\beta})
\end{aligned}
\tag{5}
$$

$\boldsymbol{\pi}_j$ being the transition matrix for the state k and $\boldsymbol{\beta}$ the prior hyper-parameter.

Letting G_j describe both, the transition matrix $\boldsymbol{\pi}_{jk}$ and the emission matrix $\boldsymbol{\Theta}$. The infinite Hidden Markov Model can be given by the following generative process described as follows:

$$
\begin{aligned}
\boldsymbol{\beta}|\gamma &\sim \text{GEM}(\gamma) \\
\boldsymbol{\pi}_j|\alpha, \boldsymbol{\beta} &\sim \text{DP}(\alpha, \boldsymbol{\beta}) \\
z_{jt}|\boldsymbol{\pi}_j &\sim \text{Mult}(\boldsymbol{\pi}_j) \\
\theta_k|\mathcal{H} &\sim \mathcal{H} \\
x_{jt}|z_{jt}, \{\theta_k\}_{k=1}^{\infty} &\sim F(\theta_{z_{jt}})
\end{aligned}
\tag{6}
$$

z_{jt} being the indicator variable of the HDP-HMM, and θ_k taking the different values of $\theta_{z_{jt}}$ distributed according to G_j with some probability π_{jk}. $F(\theta_{z_{jt}})$ is a data likelihood (e.g. a Gaussian distribution) with mean $\boldsymbol{\mu}_{z_{jt}}$ and covariance matrix $\boldsymbol{\Sigma}_{z_{jt}}, \mathcal{N}(\mathbf{x}_{jt}; \boldsymbol{\mu}_{z_{jt}}, \boldsymbol{\Sigma}_{z_{jt}})$).

the graphical model for the infinite Hidden Markov Model having the representation as in figure 1.

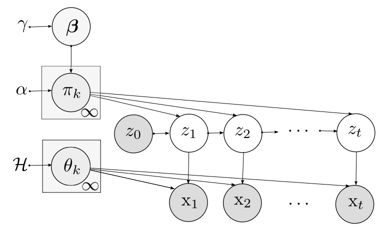

Figure 1. Infinite Hidden Markov Model (IHMM) graphical representation

Bacause of the lack of the strong beliefs of the hyper-parameters α and γ, (Teh et al., 2006) gives a Gamma distribution that where also used in other works (Beal et al., 2002; Van Gael et al., 2008).

3.3. Inference of the infinite Hidden Markov Model

In the IHMM model we estimate the state sequences $\mathbf{z} = (z_1, \ldots z_t)$ and the hyper-parameters $(\alpha, \boldsymbol{\beta}, \gamma)$ that defines the transition and the emission matrix. The Gibbs sampling is described shortly in the algorithm 1 that computes $\mathcal{O}(K)$ probabilities for each of t states giving to the inference for the infinite Hidden Markov Models with Gibbs sampling a $\mathcal{O}(TK)$ computational complexity. The Beam sampling, having the worst complexity $\mathcal{O}(TK^2)$ is briefly described, but we refer the reader for more details to (Van Gael et al., 2008).

The Gibbs sampling for the infinite Hidden Markov Model consisting in sampling of the states z_t that needs two factors, where the first is the conditional likelihood of x_t given $\mathbf{z}$, $\mathbf{X}$ and $\mathcal{H}$.

$$
p(z_t = k|\mathbf{z}_{\setminus t}, \boldsymbol{\beta}, \alpha) \propto
\begin{cases}
(n_{z_{t-1},k} + \alpha\beta_k)\frac{n_{k,z_{t+1}} + \alpha\beta_{z_{t+1}}}{n_{k\cdot} + \alpha} & \text{if} \quad k \leq K,\ k \neq z_{t-1} \\
(n_{z_{t-1},k} + \alpha\beta_k)\frac{n_{k,z_{t+1}} + 1 + \alpha\beta_{z_{t+1}}}{n_{k\cdot} + 1 + \alpha} & \text{if} \quad k = z_{t-1} = z_{t+1} \\
(n_{z_{t-1},k} + \alpha\beta_k)\frac{n_{k,z_{t+1}} + \alpha\beta_{z_{t+1}}}{n_{k\cdot} + 1 + \alpha} & \text{if} \quad k = z_{t-1} \neq z_{t+1} \\
\alpha\beta_k\beta_{z_{t+1}} & \text{if} \quad k = K + 1
\end{cases}
\tag{7}
$$

The second factor $p(z_t|\mathbf{z}_{\setminus t}, \boldsymbol{\beta}, \alpha)$ will be computed as in the equation 7,where n_{ij} is the number of transitions from state i to the state j, excluding the time steps t and $t-1$; $n_{\cdot i}$ and $n_{i\cdot}$ being the number of transition in and respectively out of state i and K is the number of distinct states in $\mathbf{z}_{-t}$.

Algorithm 1 Gibbs sampling for Infinite Hidden Markov Models

Input: The observations $\mathbf{X} = (x_1, \ldots, x_T)$, the number of samplings n_s.
Initialize a random hidden state sequence $\mathbf{z}_0 = (z_1, \ldots, z_T)$.
for $q = 1$ **to** n_s **do**
  for $t = 1$ **to** T **do**
    1. Sample the state $p(z_t = k|\mathbf{X}, \mathbf{z}_{\backslash t}, \boldsymbol{\beta}, \alpha, \mathcal{H}) \propto p(x_t|\mathbf{x}_{\backslash t}, z_t = k, \mathbf{z}_{\backslash t}, \mathcal{H})\, p(z_t = k|\mathbf{z}_{\backslash t}, \boldsymbol{\beta}, \alpha)$.
    2. Update count matrices to reflect new z_t. This step may change the number of represented hidden states K.
  end for
  3. Update the hyper-parameters $\alpha, \boldsymbol{\beta}, \gamma$.
end for

The Beam sampling for the infinite Hidden Markov Model consist in using the truncation of the transition matrix and using the dynamic programming to sample the states $\mathbf{z}$. Therefore it extends the Gibbs sampling by adding an auxiliary variable $\mathbf{u} = (u_1, \ldots, u_T)$ that has the following distribution:

$$u_t \sim \text{Uniform}(0, \pi_{z_t z_{t-1}}) \tag{8}$$

The states $\mathbf{z}$ of the model is sampled by a forward filtering-backward sampling. The idea is that only the finite number of states with probabilities $\pi_{z_{t-1} z_t} \geq u_t,\ \forall t \in T$ will have non-zero probabilities. The probability density for the state z_t will be computed as follows:

$$p(\mathbf{z}_t|\mathbf{x}, \mathbf{u}, \boldsymbol{\pi}, \boldsymbol{\Theta}) \propto p(\mathbf{x}_t|z_t) \sum p(z_{t-1}|\mathbf{x}_{\backslash t}, \mathbf{u}_{\backslash t}, \boldsymbol{\pi}, \boldsymbol{\Theta}) \tag{9}$$

In equation 9 the sum over an infinite number of terms was constrained by $u_t < \pi_{z_{t-1} z_t}$ and $p(z_{t-1}|\mathbf{x}_{\backslash t}, \mathbf{u}_{\backslash t}, \boldsymbol{\pi}, \boldsymbol{\Theta}) > 0$, making it to sum over the finite number of $\mathbf{z}_{t-1}$.

The sampling of the model parameters in the infinite Hidden Markov Model, in particular $\boldsymbol{\pi}$, $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are given in (Teh et al., 2006), however we give a brief introduction to these. Supposing n_{jk} the number of times that a state j transits to another state k, where $j, k \in 1 \ldots K$, K representing the number of different states in $\mathbf{z}$, the conditional distribution for the transition matrix $\boldsymbol{\pi}$ given the states $\mathbf{z}$, and the hyper-parameters $\boldsymbol{\beta}, \alpha$ is sampled by a Dirichlet distribution:

$$\boldsymbol{\pi}_j \propto \text{Dir}(n_{j1} + \alpha\beta_1, \ldots n_{jK} + \alpha\beta_K, \alpha \sum_{i=K+1}^{\infty} \beta_i) \tag{10}$$

where $\boldsymbol{\beta}$ is also sampled according to the Dirichlet distribution:

$$\boldsymbol{\beta} \propto \text{Dir}(m_{.1}, \ldots, m_{.K}, \gamma) \tag{11}$$

where $m_{.k}$ representing the number of clusters k in all the data groups J, respectively one can say $m_{.k} = \sum_{j=1}^{K} m_{jk}$. More details are could be found in (Teh et al., 2006; Antoniak, 1974). At the end the parameters $\boldsymbol{\theta}_k = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$ conditional on the data $\mathbf{X}$, states $\mathbf{z}$ and the prior distribution $\mathcal{NIW}(\mu_0, \kappa_0, \nu_0, \Lambda_0)$ are sampled according to their posterior distributions.

4. Experiments

In this experiment, we apply the infinite Hidden Markov Model to a challenging problem of humpback whale song decomposition. Humpback whales produce songs with a specific structure and the study of that songs is very challenging and very useful for bio-acousticians and scientists to namely understand how do whales song and communicate (possibly according to which vocabulary) and to have an idea about their origin, since the songs of whales from different origins can be different. The analysis of such complex signals that aims at discovering the call units (which can be considered as a kind of whale vocabulary), can be seen as a problem of unsupervised call units classification as in (Pace et al., 2010). We therefore reformulate the problem of whale song decomposition as a clustering problem. Contrary to the approach used in (Pace et al., 2010), in which the number of clusters (call units in this case) has been fixed manually, here, we apply the infinite Hidden Markov Model to find the sates of the whale song, and automatically infer the number of states from the data. The used data are available in the framework of our SABIOD project[2]. The data consist of MFCC parameters of 8.6 minutes of a Humpback whale song recordings produced at few meters distance from the whale in La Reunion - Indian Ocean. The 8.6 minutes of a Humpback whale song recordings where treated for this application. The Hierarchical Dirichlet Process for Hidden Markov Model was investigated by two approaches, the Gibbs sampling (Fox et al., 2008) with the original HDP-HMM (Teh et al., 2006) and the Beam sampling (Van Gael et al., 2008). The week limit approximation to the DP (Ishwaran & Zarepour, 2002) also named in literature as the truncation level (L) [3] equal to 30 was used for the two approaches. The algorithms runs for 29000 sweeps and the state sequence partitions with the spectrogram of the humpback whale song units highlight the interest of using the Bayesian non-parametric approaches over the Hidden Markov Model.

Figure 2 illustrates the state sequences obtained by using the Gibbs sampling inference approach for the HDP-HMM developed in (Teh et al., 2006; Fox et al., 2008). One can

[2] Scaled Acoustic BIODiversity: http://sabiod.univ-tln.fr/data_samples.html

[3] L is the number that is bigger then the expected number of states in the model

see the number of states estimated in this context is equal to $K = 6$ being much smaller then the truncation level L. The figure 2 illustrating the maximum state sequence possible supposed a priori equal to $L = 30$, one can see that there are empty states.

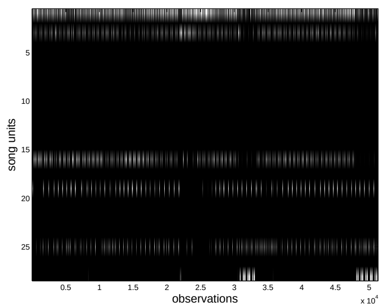

Figure 2. The state sequences obtained by the Gibbs sampling inference approach for Infinite Hidden Markov Model of the humpback whale song data.

In the figure 3 we illustrate the spectrogram for 3 states sequences of the song units obtained for the humpback whale song data by the bas HDP-HMM with the Gibbs sampling inference approach. On the vertical axes the frequency (0 to 22.05 kHz) is showed and on the horizontal axes we have the time, represented in seconds.

The Beam sampling was also investigated over the humpback whale song data, and the results where shown in the figure 4. The inferred state sequence representation on the left of the figure 4 was zoomed to 10 sequences, however the truncation level was taken to be the same as for the Gibbs sampling for the HDP-HMM $(L = 30)$. We notice the ninth sequence are clearly conveying information. This determines the humpback whale song data. On the right of the figure 4 observe the spectrogram of the whale song.

5. Conclusion

In this paper we relied on the Hierarchical Dirichlet Process for Hidden Markov Model for unsupervised learning from complex bioacoustic data. The two different approaches, the Gibbs sampling and the Beam sampling were investigated over the bioacoustic signals. The possible hidden whale song units of the humpback whale signals were determined in an automatic way. The obtaining results highlight the interest of using the Bayesian non-parametric approach for the Hidden Markov Model. In a future work we propose the possibility to make the eigenvalue decomposition for the covariance matrix for the emission density of the HMM, where more flexible models could appear in term of different volumes, orientations and shapes for each clusters.

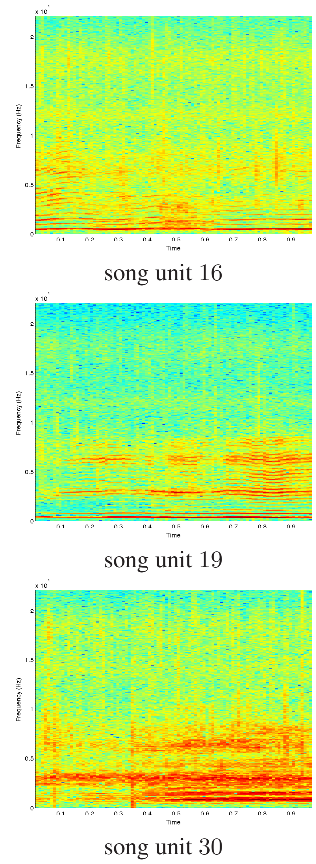

Figure 3. Spectrograms for the song units of the humpback whale obtained with the Gibbs sampling inference approach for Infinite Hidden Markov Model.

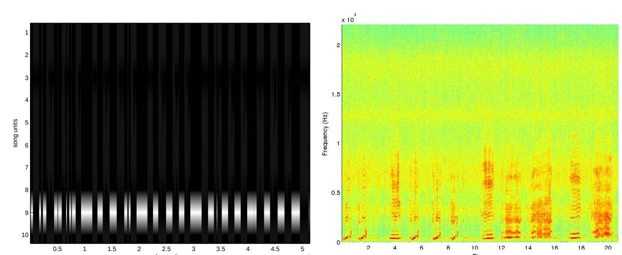

Figure 4. The state sequences (on left) and the spectrogram (on right), (the ninth unit) for the song of humpback whale obtained by the the Beam sampling inference approach for the Infinite Hidden Markov Model

Acknowledgments

This work is supported by the SABIOD project: `http://sabiod.univ-tln.fr`.

References

Andrew Gelman, John B. Carlin, Hal S. Stern and Rubin, Donald B. *Bayesian Data Analysis*. Chapman and

Hall/CRC, 2003.

Antoniak, C. E. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2(6):1152–1174, 1974.

Beal, Matthew J., Ghahramani, Zoubin, and Rasmussen, Carl E. The infinite hidden markov model. In *Machine Learning*, pp. 29–245. MIT Press, 2002.

Derrode, S. Benyoussef, L. and Pieczynski, W. Contextual estimation of hidden markov chains with application to image segmentation. In *ICASSP*, pp. 15–19, Toulouse, May 2006.

Ferguson, Thomas S. A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.

Fox, Emily B., Sudderth, Erik B., Jordan, Michael I., and Willsky, Alan S. An HDP-HMM for systems with state persistence. In *ICML 2008: Proceedings of the 25th international conference on Machine learning*, pp. 312–319, New York, NY, USA, 2008. ACM.

Hjort, N., C., Holmes, Muller, P., and Waller, S. G. *Bayesian Non Parametrics*. Cambrige University Press, 2010.

Ishwaran, Hemant and Zarepour, Mahmoud. Exact and approximate sum representations for the Dirichlet process. *Can J Statistics*, 30(2):269–283, 2002.

Pace, Federica, Benard, Frederic, Glotin, Herve, Adam, Olivier, and White, Paul. Subunit definition and analysis for humpback whale call classification. *Applied Acoustics*, 71(11):1107 – 1112, 2010.

Rabiner, L. and Juang, B.-H. *Fundamentals of speech recognition*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

Rabiner, L. R. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

Rabiner, L. R. and Juang, B. H. An introduction to hidden markov models. *IEEE ASSp Magazine*, 1986.

Rasmussen, C. The infinite gaussian mixture model. *Advances in neuronal Information Processing Systems*, 10: 554 – 560, 2000.

Robert, Christian P. *The Bayesian choice: a decision-theoretic motivation*. Springer-Verlag, 1994.

Sethurman, J. A constructive definition of dirichlet priors. *Statistica Sinica*, 4(2):639–650, 1994.

Teh, Yee Whye, Jordan, Michael I., Beal, Matthew J., and Blei, David M. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

Van Gael, J., Saatci, Y., Teh, Y.W., and Ghahramani, Z. Beam sampling for the infinite hidden Markov model. In *Proceedings of the 25th international conference on Machine learning*, pp. 1088–1095. ACM New York, NY, USA, 2008.

Wood, F. and Black, M. J. A nonparametric bayesian alternative to spike sorting. *Journal of neuroscience methods*, 173(1):1–12, 2008.