The Infinite Factorial Hidden Markov Model

Jurgen Van Gael JV279@CAM.AC.UK
Yee Whye Teh YWTEH@GATSBY.UCL.AC.UK
Zoubin Ghahramani ZOUBIN@ENG.CAM.AC.UK

Abstract

The infinite factorial hidden Markov model is a non-parametric extension of the factorial hidden Markov model. Our model defines a probability distribution over an infinite number of independent binary hidden Markov chains which together produce an observable sequence of random variables. Central to our model is a new type of non-parametric prior distribution inspired by the Indian Buffet Process which we call the *Indian Buffet Markov Process*.

1. Introduction

Hidden Markov models (HMM) are among our most successful models for dealing with time series data and the machine learning community has extended this model in many orthogonal directions. The factorial hidden Markov model (FHMM) (Ghahramani & Jordan, 1997) is an extension of the classical hidden Markov model in which the hidden state is factored. In other words, the information from the past is propagated in a distributed way through a set of parallel Markov chains. This model has been used in vision, audio processing and natural language processing.

Orthogonal to this work, the infinite hidden Markov model (iHMM) (Beal et al., 2002) is a non-parametric extension of the hidden Markov model where the number of hidden chains can be potentially unbounded. At the heart of this model is the hierarchical Dirichlet process (Teh et al., 2006) which formalizes the notion of an infinite number of states.

In our work, we "complete the square" by extending the factorial hidden Markov model to allow for an unbounded number of hidden chains. Central to our model is a new non-parametric prior which is inspired by the Indian Buffet Process but adding Markov dynamics. We call this the *Indian Buffet Markov Process* (IBMP). In section 2 we describe the construction of our new non-parametric prior and how it can be used in the most natural full model which we refer to as the *Infinite Factorial Hidden Markov Model* (iFHMM). Section 3 discusses recent work on how to do inference for the iFHMM. We end in section 4 with some preliminary results using the iFHMM model.

Presented at the *Nonparametric Bayes Workshop* at the *25'th International Conference on Machine Learning*, Helsinki, Finland, 2008.

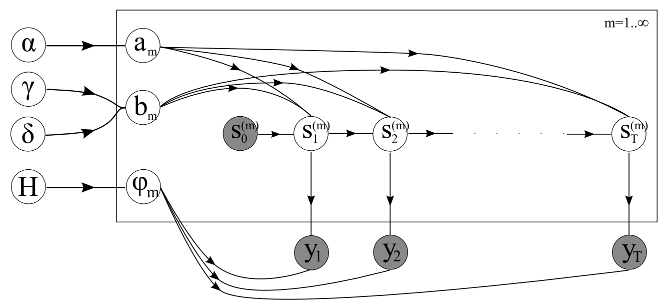

Figure 1. The Infinite Factorial Hidden Markov Model

2. The Indian Buffet Markov Model

We start by describing a finite generative model, carefully choosing the parameters to depend on the model size so that we can take the infinite limit; this is analogous to the derivation of the IBP in (Griffiths & Ghahramani, 2006). Our finite model will consist of M binary Markov chains with a transition matrix $\boldsymbol{W}_m$ for each chain parametrized by a_m and b_m as follows

$$\mathbf{W}_m = \begin{pmatrix} 1-a_m & a_m \\ 1-b_m & b_m \end{pmatrix}, \qquad (1)$$

where we sample the parameter $a_m \sim \mathsf{Beta}(\alpha/M, 1)$ and $b_m \sim \mathsf{Beta}(\gamma, \delta)$. We denote with Z_{tm} the state of the m'th hidden Markov chain at time t. We assume every chain starts off in a dummy 0 state ($Z_{0m} = 0$ for all $m \in M$) and then evolves according to the transition matrix $\boldsymbol{W}_m$ for T timesteps.

If we denote with $c_m^{00}, c_m^{01}, c_m^{10}, c_m^{11}$ the number of transitions from $0 \to 0, 0 \to 1, 1 \to 0$ and $1 \to 1$ in binary chain m (we count the transition from the dummy state to the first state too), we can then write

$$p(\boldsymbol{Z}|\boldsymbol{a},\boldsymbol{b}) = \prod_{m=1}^{M} (1-a_m)^{c_m^{00}} a_m^{c_m^{01}} (1-b_m)^{c_m^{10}} b_m^{c_m^{11}}. \quad (2)$$

Combining this with the priors on $\boldsymbol{a}$ and $\boldsymbol{b}$ we can marginalize out the parameters and find

$$p(\boldsymbol{Z}|\alpha,\gamma,\beta) = \prod_{m=1}^{M} \frac{\frac{\alpha}{M}\Gamma(\frac{\alpha}{M}+c_m^{01})\Gamma(c_m^{00}+1)}{\Gamma(\frac{\alpha}{M}+c_m^{00}+c_m^{01}+1)} \frac{\Gamma(\gamma+\delta)\Gamma(\delta+c_m^{10})\Gamma(\gamma+c_m^{11})}{\Gamma(\gamma)\Gamma(\delta)\Gamma(\gamma+\delta+c_m^{10}+c_m^{11})}.$$

Next, we want to take the limit $M \to \infty$. Similar to the IBP, $p(\boldsymbol{Z}|\alpha,\gamma,\beta) = 0$ when M becomes infinitely large, however we are only interested in the probability of the left-ordered form equivalence class of Z (denoted by $[Z]$). After careful algebraic manipulation, we find that for $M \to \infty$

$$p([\boldsymbol{Z}]) = \frac{\alpha^{M_+}}{\prod_{h=0}^{2^T-1} M_h!} \exp\{-\alpha H_T\} \prod_{m=1}^{M_+} \frac{(c_m^{01}-1)!c_m^{00}!\Gamma(\gamma+\delta)\Gamma(\delta+c_m^{10})\Gamma(\gamma+c_m^{11})}{(c_m^{00}+c_m^{01})!\Gamma(\gamma)\Gamma(\delta)\Gamma(\gamma+\delta+c_m^{10}+c_m^{11})}.$$

This defines the IBMP prior distribution $p([\boldsymbol{Z}])$.

2.1. The iFHMM

The most natural application of the IBMP prior is to add a parameter ϕ_m for each chain which is sampled from a base measure H. We then add a likelihood model F which produces an observation y_t at every timestep t by combining the parameters $\boldsymbol{\phi}$ conditional on which chains are switched on at time t. This model, whose graphical model is shown in figure 1, we refer to as the iFHMM.

3. Inference

Because the iFHMM models time series data, we want to use some form of dynamic programming to do inference rather than a Gibbs sampler. Luckily, we can extend the stick breaking construction of the IBP (Teh et al., 2007) to the IBMP. This allows us to implement a slice sampling algorithm which will adaptively truncate our model into a finite FHMM. Sampling the hidden variables in a finite model can then be done by a blocked Gibbs sampler which samples a single *whole* Markov chain using a forward-filtering backward-sampling algorithm, conditional on keeping all other Markov chains fixed. Although this allows us to learn efficiently in the models we have considered so far, we are currently experimenting with adding Metropolis Hasting's steps to improve performance.

4. Experiments

A first model which we consider is a linear Gaussian model through time. We set the base measure H of our iFHMM to be a D dimensional normal distribution with variance σ_A^2. In other words, with each chain we associate aD dimensional vector. The likelihood function F just sums up all the vectors for the chains that are on and adds zero mean Gaussian noise with variance σ_X^2 to generate the D dimensional observations $\boldsymbol{Y}$.

We created a bars-in-time dataset by choosing the five 25 dimensional vectors shown in figure 2 and generated 5 random Markov chains of length 400. We ran our inference algorithm for 1000 iterations and recovered the bars as shown in figure 3. Our current work focusses on using the iFHMM for real world problem.

References

Beal, M. J., Ghahramani, Z., & Rasmussen, C. E. (2002). The infinite hidden markov model. *NIPS*, *14*, 577 – 584.

Ghahramani, Z., & Jordan, M. I. (1997). Factorial hidden markov models. *Machine Learning*, *29*, 245–273.

Griffiths, T. L., & Ghahramani, Z. (2006). Infinite latent feature models and the indian buffet process. *NIPS*, *18*, 475–482.

Teh, Y. W., Görür, D., & Ghahramani, Z. (2007). Stick-breaking construction for the indian buffet process. *AI-Stat*, *11*.

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006). Hierarchical dirichlet processes. *JASA*, *101*, 1566–1581.

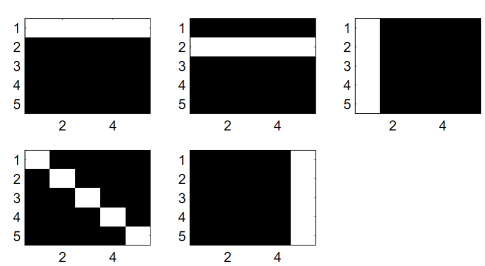

Figure 2. Generating Bars

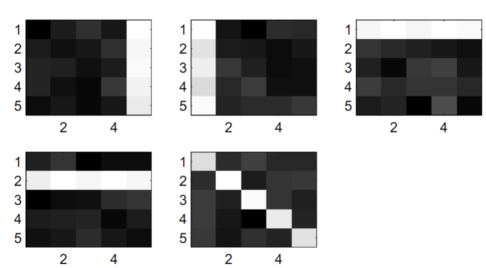

Figure 3. Bars sample after 1000 iterations.