

Infinite Hidden Markov Models via the Hierarchical Dirichlet Process

Matthew J. Beal* (presenting author)

Yee Whye Teh†

Michael I. Jordan†

beal@cs.toronto.edu

ywteh@eecs.berkeley.edu

jordan@eecs.berkeley.edu

* Department of Computer Science, University of Toronto, 6 King’s College Road, Toronto, M5S 3G4 CANADA

† Computer Science Division, University of California at Berkeley, Berkeley, CA 94709 USA

In this presentation, we propose a new formalism under which we study the *infinite hidden Markov model (iHMM)* of Beal et al. [2]. The iHMM is a hidden Markov model (HMM) in which the number of hidden states is allowed to be countably infinite. This is achieved using the formalism of the Dirichlet process. In particular, a two-level urn model is used to determine the transition probabilities of the HMM (an urn model is a particular aspect of the Dirichlet process.) At the first level, the probability of transitioning from a state u to a state v is proportional to the number of times the same transition is observed at other time steps, while with probability proportional to α_0 an “oracle” process is invoked. At this second level, the probability of transitioning to state v is proportional to the number of times state v has been chosen by the oracle (regardless of the previous state), while the probability of transitioning to a previously unseen (i.e. novel) state is proportional to γ . The inclusion of the oracle was motivated by a need to tie together the transition models to have common destination states. Beal et al. presented an approximate Gibbs sampling algorithm for inference in this model, but without an explicit generative model the scope of the model was limited.

A fresh aspect on the iHMM is provided by the recent development of the *Hierarchical Dirichlet Process (HDP)* of Teh et al. [1]. The HDP framework considers problems involving related *groups of data*. In particular, each group of data is modelled by a Dirichlet process (DP) mixture model, with the common base measure of the DPs being itself distributed according to a global DP. This is a hierarchical Bayesian model, and is necessary to ensure that the different DP mixtures are able to share a common set of mixture components. Note that Beal et al. defined a notion of “hierarchical Dirichlet process”, but this is not hierarchical in the Bayesian sense—involving a distribution on the parameters of a Dirichlet process—but is instead a description of the two level urn model.

Nevertheless, it turns out that it is possible to formulate the iHMM within the HDP framework of Teh et al. (Figure 1 shows the relevant hierarchical Bayesian model for the iHMM), by viewing it as another instance of the grouped data formulation in the following way. We assign observations to groups, where the groups are indexed by the value of the previous state variable in the sequence; then the current state and emission distributions define a group-specific mixture model. In this way, the sequential ordering of the data is essential since the hidden state sequence implicitly defines a partition into groups. Further, the *Chinese restaurant franchise (CRF)* aspect of the HDP turns out to be equivalent to the two level urn model of Beal et al. This realisation of the iHMM in the HDP framework seems straightforward, but there are some striking non-trivialities to resolve.

First, the HDP framework was developed assuming a *fixed* partition of the data into groups; but for the iHMM, the above definition of groups means that the assignment of the sequence of observations into groups is itself random. To clarify this point, consider the CRF aspect of the HDP. Unlike the CRF, in the iHMM the assignment of customers to restaurants is dependent on the dish that the previous customer ate. Moreover, since the group assignment depends on the identity of the previous state, which is from a countably infinite set, there are now a countably infinite number of restaurants.

The second point to note is that for the HDP there currently exist two sampling algorithms for inference: the first based on the CRF description, where the allocations of customers to tables and of dishes to tables are sampled; the second is an auxiliary variable method [1]. It turns out that the CRF description, which was the natural one in Beal et al. , is extremely awkward for setting up an inference algorithm for the iHMM, as it involves substantial book-keeping. Indeed Beal et al. did not present an MCMC inference algorithm for the iHMM, proposing instead a heuristic approximation to Gibbs sampling. Fortunately however, the auxiliary variable method is more readily applicable to the iHMM, and has allowed us to derive novel correct algorithms for sampling in the iHMM. This is a major contribution over Beal et al.

To summarise, in this presentation we consider a new framework for modelling time-series data, based on an extension of the recently introduced HDP framework [1]. We show how previous work on the iHMM [2] can be encompassed and clarified within the HDP framework, and as a result we can provide effective new inference algorithms for the iHMM. We report on experiments applying the iHMM to real world data, including one on Descartes’ Meditations.

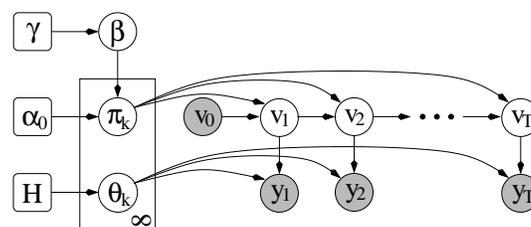


Figure 1: Hierarchical Bayesian model for the iHMM.

[1] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet Processes. Technical Report 653, Department of Statistics, University of California, Berkeley, 2004.

[2] M. J. Beal, Z. Ghahramani, and C. E. Rasmussen. The infinite hidden Markov model. In *Advances in Neural Information Processing Systems 14*, Cambridge, MA, 2002. MIT Press.

Category: graphical models.