Summary 6: Hierarchical Beta Processes and the Indian Buffet Process

Susan Cherry

This paper introduces an alternative to the classic mixture models that have a multinomial representation. Instead, it discuses factorial models, which associate each data point with a set of latent Bernoulli variables. The advantages of the factorial representation are that the Bernoulli variables have a natural interpretation of "featural" descriptions, representing objects in terms of sets of Bernoulli variables provides a natural way to define topologies on clusters, and the factorial approach is appropriate for instances with large numbers of clusters.

The Indian Buffet Process is a factorial analog of the Chinese Restaurant Process. While the de Finetti mixing distribution of the CRP is the Dirichlet Process, this paper identifies the de Finetti mixing distribution behind the Indian Buffett Process as the Beta Process. The beta process is often studied in survival analysis. The paper defines both the Beta Process and the Bernoulli Process. Next, it connects these to processes to the Indian Buffet Process. They present the two-parameter generalization of the IBP with concentration parameter c and mass parameter γ.

Next the paper introduces a algorithm to generate the beta process. Their algorithm is simple and efficient and is closely related to the stick breaking construction of Dirichlet Processes. The algorithm to construct $\hat{B_n}$ is as follows: Start with $\hat{B_0} = 0$. For each step $n \geq 1$, first sample K_n $Poi(\frac{c\gamma}{c+n-1})$. Second, sample K_n new locations ω_j from $\frac{1}{\gamma}B_0$ independently. Third, sample their weightp_j $Beta(1, c+n-1)$ independently. Fourth, $\hat{B_n} = \hat{B_{n-1}} + \sum_{j=1}^{K_n} p_j \delta_{\omega_j}$.

. Next, the paper discusses the hierarchical beta process, which is the parallel to the hierarchical Dirichlet process. They have the following model, in terms of document classification: Baseline: B $BP(c_0, B_0)$, Cagetories: A_j $BP(c_j, B)$ for all $j \leq n$, Documents: $X_{i,j}$ $BeP(A_j)$ for all $i \leq n_j$. Then they discuss Monte Carlo inference algorithm for hierarchies of arbitrarily many levels. First the talk about the discrete part and how to derive the posterior. The posterior is log concave and has a maximum which can be obtained by binary search. Then they discuss the continuous part, where all observations are equal to zero. Finally, the extend the model to larger hierarchies, so the algorithm can handle hierarchies of arbitrary depth.

Finally, the paper discusses application to document classification. Naive Bayes is often used to model documents as lists of features and assumes that features are independent given category. Laplace smoothing can be used to correct for unbalanced training data, but it makes rare features hurt performance, so usually naive Bayes with feature selection is used. This wastes information, so the authors propose hierarchical Beta Process (hBP). hBP allows the number of features to grow with data and gives a constant prior for varying amounts of data. The performance of hBP and naive Bayes on 100 posts from "Newsgroups" was compared and hBP had 58% accuracy, compared to 50% for naive Bayes. The hBP handles rare features well, so it is safer to include a larger set of features than possible for a naive Bayes model.