ARTICLE IN PRESS

Graphical Models xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Graphical Models

journal homepage: www.elsevier.com/locate/gmod

Swendsen-Wang Cuts sampling for spatially constrained Dirichlet process mixture models

Xiangrong Wang [a,b], Jieyu Zhao [a,*]

[a] Research Institute of Computer Science and Technology, Ningbo University, Ningbo, China

[b] College of Information Engineering, Ningbo Dahongying University, Ningbo, China

ARTICLE INFO

Article history:
Received 3 March 2014
Accepted 10 March 2014
Available online xxxx

Keywords:
Dirichlet process
Markov random fields
Swendsen-Wang Cuts
Collapsed Gibbs sampling
Bayesian nonparametric models

ABSTRACT

Spatially constrained Dirichlet process mixture models are springing up in image processing in recent years. However, inference for the model is NP-hard. Gibbs sampling which is a generic Markov chain Monte Carlo technique is commonly employed for the model inference. It needs to traverse all the nodes of the constructed graph in each iteration. The sampling process hardly crosses over the intermediate low probabilistic state. In addition, it is not well informed by the spatial relationship in the sampling process. In this paper, a spatially dependent split-merge algorithm for sampling the MRF/DPMM model based on Swendsen-Wang Cuts is proposed. It is a state of the art algorithm which combines the spatial relationship to direct the sampling, and lessen the mixing time drastically. In this algorithm, a set of nodes are being frozen together according to the discriminative probability of the edges between neighboring nodes. The frozen nodes update their states simultaneously in contrast to the single node update in a Gibbs sampling. The final step of the algorithm is to accept the proposed new state according to the Metropolis Hasting scheme, in which only the ratio of posterior distribution needs to be calculated in each iteration. Experimental results demonstrated that the proposed sampling algorithm is able to reduce the mixing time considerably. At the same time, it can obtain comparably stable results with a random initial state.

1. Introduction

The Dirichlet Process (DP) [1] draws much attention in machine learning such as natural language processing [2], object recognition [3,4], image retrieval [5], topic modeling [6], and image segmentation [7,8] in recent years. A notorious property that a DP model owns is the model selection. Algorithms such as K-means clustering, histogram clustering and Gaussian mixture model clustering need determine the number of clusters in advance. However, the fixed number of clusters always leads to undesirable results as the observed data varies in a wide range. A Dirichlet process framework is an advanced model that can determine the number of clusters from the observed data dynamically.

The Dirichlet process mixture model (DPMM) [9] is a practical model of DP. It is employed to perform model selections in various applications [2–8]. However, in some specific domains such as image processing, a DPMM is always combined with other spatial relationship constrained models since the spatial relationship always plays a significant role in the model selection. Models presented in literature [7–15] are several typical spatially constrained Dirichlet process mixture models (SCDPMM). With the model constructed, inference is crucial to obtain the final solution. Exact inference for the DPMM is impractical since it is difficult to handle the infinite number of components

* Corresponding author.
E-mail addresses: xiang_rong_wang@163.com (X. Wang), zhao_jieyu@nbu.edu.cn (J. Zhao).

http://dx.doi.org/10.1016/j.gmod.2014.03.008
1524-0703/

Please cite this article in press as: X. Wang, J. Zhao, Swendsen-Wang Cuts sampling for spatially constrained Dirichlet process mixture models, Graph. Models (2014), http://dx.doi.org/10.1016/j.gmod.2014.03.008

in specific applications. Naturally, various approximations methodologies have been proposed to solve the inference problems. All of these methodologies can be divided into two categories: the Markov Chain Monte Calo (MCMC) [9,16] and the variational inference [17,18]. In a variational inference, a gradient descent scheme is designed to search the local minima and its efficiency is largely dependent on the choice of the initial point. Unlike the variational inference, the MCMC approach employs a probabilistic scheme to draw a set of samples from the target distribution and achieves the final solution through statistical methods. Markov chains are designed in these algorithms to traverse the state space and derives relatively accurate estimates of the solutions.

In this paper, attention will be paid to the MCMC sampling methods. The MCMC sampler draws a set of samples from the posterior distribution and derives the empirical posterior distribution from these drawn samples. In order to sample the complicated true posterior distribution efficiently, a Markov chain with the target posterior as the limiting distribution is constructed. The widely used MCMC sampling for the DPMM model is an incremental Gibbs sampling [19–22], which is the simplest version of the MCMC. However, the Gibbs sampling updates only one variable at a time and it is hard to cross over the intermediate low-probability states, which leads to slow convergence. A spit-merge sampling algorithm [23–25] is proposed to surmount the obstacle by splitting or merging a set of variables, as well as it can keep detail balance of the Markov chain by employing the Metropolis-Hasting scheme [26,27].

It is even more complicated while sampling the SCDPMM such as the MRF/DPMM model since the extra interaction between the spatially adjacent observations makes them mutually dependent. The available sampling algorithms [7,10,9,16] for the DPMM in the literature are not suitable for the SCDPMM since these algorithms are not well informed by the input data to adjust the sampling course which leads to a long mixing time.

We propose a new Swendsen-Wang Cuts [28,29] based sampling algorithm for the MRF/DPMM model in this paper. In the MRF/DPMM model, a Dirichlet process prior is considered as an external field of a Markov random field (MRF), and functions as a model selector. From this viewpoint, a Dirichlet process prior does not violate the law of the MRF. Therefore, the MRF/DPMM model is still a Markov random field in essence. Sampling this model is very like sampling a MRF model. A Swendsen-Wang Cuts sampler is a reasonable choice for the model inference. In our proposed algorithm, the whole process is divided into three steps. The first step is to sample the edge variables introduced into the algorithm. Each edge will be frozen or broken up probabilistically. Then a set of connected components are formed according to the frozen edges. The second step is to sample the cluster label, in which a new candidate label is assigned to the selected component. In the last step the new label will be accepted or rejected according to the Metropolis-Hasting scheme. The DP prior in the model will affect the sampling path through the acceptance ratio which is the key to the SWC sampling algorithm. The first two steps of our algorithm are similar to those in the generic SWC algorithm. Experimental results show that the performance of the proposed SWC algorithm has been greatly improved in contrast to that of a Gibbs sampler or a split-merge sampler.

2. Dirichlet process related models

2.1. The Dirichlet process

The Dirichlet process is a Bayesian nonparametric model. It can be characterized by a base distribution G_0 and an innovation parameter α, denoted as $DP(G_0, \alpha)$. Let G be a discrete distribution which is drawn from the DP. $\{\Theta_i\}_{i=1}^n$ is a set of variables that follows the distribution G. Thus the DP model can be described as

$$G \mid G_0, \alpha \sim DP(G_0, \alpha) \tag{1}$$

$$\Theta_i \mid G \sim G \quad i = 1, \dots, n \tag{2}$$

From the formulation [11], we can see that a DP is a distribution placed over the distribution G, which is discrete with a probability 1. In order to sample this model conveniently, the random distribution G is integrated out from the model. With this operation, a new representation for the DP is derived. Specifically, given the $n-1$ samples drawn from G, $\{\Theta_i\}_{i=1}^{n-1}$, a new sample Θ_n can be selected from the existing $n-1$ samples or drawn from the base distribution G_0. Let $\{\Theta_c^*\}_{c=1}^K$ be the set of distinct values of the variables $\{\Theta_i\}_{i=1}^{n-1}$, and n_c^{n-1} be the number of Θ that take value Θ_c^*. The conditional probability of Θ_n takes the form [11]

$$p(\Theta_n \mid \{\Theta_i\}_{i=1}^{n-1}, G_0, \alpha) = \frac{\alpha}{\alpha + n - 1} G_0 + \sum_{c=1}^{K} \frac{n_c^{n-1}}{\alpha + n - 1} \delta_{\Theta_c^*} \tag{3}$$

where $\delta_{\Theta_c^*}$ denotes the mass probability concentrated at a single point Θ_c^*. From this representation of the DP, several important properties can be easily observed. Above all, the innovation parameter α decides how often to generate a new distinct Θ_c^*. As α increases, it is more likely to draw a new distinct parameter from the base distribution G_0. At its limit, G will approach G_0. On the contrary, when α approaches 0, all the samples tend to cluster into one class. This clustering effect of the model makes it universally applied in many clustering problems. The second important property is that the more the value is shared among the variables, the more probable it will be taken in the subsequent sampling.

In the Bayesian context, we can explain (3) another way [7]. Assume $\{\Theta_i\}_{i=1}^{n-1}$ is drawn independently from the Dirichlet process, and let $\widehat{G}_{n-1}$ be the empirical distribution of the variable set $\{\Theta_i\}_{i=1}^{n-1}$

$$\widehat{G}_{n-1} = \frac{1}{n-1} \sum_{i=1}^{n-1} \delta_{\Theta_i} \tag{4}$$

Then the posterior distribution G conditioned on the existing drawn samples is also a Dirichlet process

$$G \mid \{\Theta_i\}_{i=1}^{n-1} \sim DP(\alpha G_0 + (n-1)\widehat{G}_{n-1}) \tag{5}$$

Thus Eq. (3) is obtained. In order to draw the sequence $\{\Theta_i\}_{i=1}^n$ from the random measure $G \sim DP(G_0, \alpha)$, the first sample is drawn according to

Please cite this article in press as: X. Wang, J. Zhao, Swendsen-Wang Cuts sampling for spatially constrained Dirichlet process mixture models, Graph. Models (2014), http://dx.doi.org/10.1016/j.gmod.2014.03.008