Bayesian Nonparametric Priors for Hidden Markov Random Fields

Hongliang Lu, Julyan Arbel, Florence Forbes

Bayesian Nonparametric Priors for Hidden Markov Random Fields

Hongliang Lü & Julyan Arbel & Florence Forbes

Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

Résumé.
L'un des problèmes centraux en statistique et apprentissage automatique est de savoir comment choisir un modèle adéquat qui peut automatiquement s'adapter à la complexité des données observées. L'approche bayésienne non paramétrique est une solution intéressante pour gérer cette difficulté. Basés sur un espace de paramètres en dimension infinie, les modèles bayésiens non paramétriques sont flexibles et peuvent être relativement facilement utilisés pour apprendre des jeux de données complexes. Dans ce travail, nous abordons le problème de la détermination automatique du nombre de groupes en classification non supervisée lorsque les données à classer ne sont pas independantes mais modelisées à l'aide d'un champ de Markov. Plus precisément, l'estimation du nombre de groupes est évitée en considérant un modèle qui suppose un nombre infini de groupes. Nous montrons comment un champ aléatoire de Markov peut être combiné avec différentes lois a priori non paramétriques. Nous illustrons cela à l'aide d'un modèle de Potts combiné à un processus de Dirichlet et à un processus de Pitman-Yor. L'inférence de ces modèles est basée sur l'algorithme d'expectation-maximization variationnel en raison de son coût de calcul plus faible que l'approche Monte-Carlo par chaînes Markov (MCMC). L'approche proposée est appliquée à la segmentation d'images et quelques comparaisons et résultats préliminaires sont présentés et discutés.

Mots-clés. Champs de Markov cachés, méthodes bayésiennes non paramétriques, approximation variationelle, clustering, données spatiales, segmentation d'images.

Abstract. One of the central issues in statistics and machine learning is how to select an adequate model that can automatically adapt its complexity to the observed data. Bayesian nonparametric methods are thought of as one of the most promising candidates that are capable of handling such tasks. Based on an infinite-dimensional parameter space, Bayesian nonparametric models are highly flexible and thus can be readily used for parameterizing or learning about complex datasets. In the present work, we consider the issue of determining from the data the number of groups in a clustering task of non independent observations. The required guess on the number of clusters is avoided by considering models with an infinite number of clusters as suggested in Dirichlet Process Mixture models. We propose to combine a Markov random field model with different Bayesian nonparametric priors and illustrate such a combination on a Potts model combined with a Dirichlet process and a Pitman-Yor process. As regards inference, the variational expectation-maximization algorithm is adopted due to its lower computational cost with respect to its MCMC counterpart. Finally, the proposed framework is applied to image segmentation and some preliminary comparisons and results are presented and discussed.

Keywords. Hidden Markov random fields, Bayesian nonparametric methods, variational approximation, clustering, spatial data, image segmentation.

1 Introduction

Hidden Markov random field (HMRF) models are widely used for image segmentation or more generally for clustering data under spatial constraints. They can be seen as spatial extensions of independent mixture models. As for standard mixtures, one concern is the automatic selection of the proper number of components in the mixture, or equivalently the number of states in the hidden Markov field. Several criteria exist to select this number automatically based on penalized likelihood (eg. AIC, BIC, ICL etc.) but they usually require to run several models for different number of classes to choose the best one. Other techniques use a fully Bayesian setting including a prior on the class number. The most commonly used method in this case is reversible jump Markov chain Monte Carlo. Although simplifications in the inference have been proposed recently in [11], the computational cost of reversible jump techniques remains high. In this work, we propose to investigate alternatives based on the field of Bayesian nonparametrics. In particular, Dirichlet process mixture models (DPMM) have emerged as promising candidates for clustering applications where the number of clusters is unknown. Most applications of DPMM involve observations which are supposed to be independent. For more complex tasks such as unsupervised image segmentation with spatial relationships or dependencies between the observations, DPMM are not satisfying. In many applications, data points are not independent and require models that account for dependencies. When clustering data, the goal is to produce a partition of the observations (*eg.* via a labelling of each data point) that accounts for the groups existing in the observed data. To address this goal, a probabilistic missing data framework can be considered assuming the existence of a set of missing variables representing a set of labels. Dependencies between labels are modelled via Markov random field (MRF) models. Starting from a Potts Markov random field (MRF) modelisation of spatial dependencies and following the line of DPMM, we propose to extend the standard finite state space Potts model to a countable infinite number of states. This requires the use of an appropriate prior on the state variable in the MRF formulation.

2 Bayesian Nonparametric Hidden Markov models

Attempts to build such infinite HMRF have been made using BNP priors before. In particular, we can distinguish attempts such as [5, 4, 6] from the work in [1, 12, 7, 13]. The approach in [5, 4, 6] differs in that it is not based on a generalization of the Potts model but on a transformation of an inference algorithm. More specifically in [5, 4], a mean field approximation is first considered and then transformed to account for an infinite number of states. In that sense it is closer to an Iterated Conditional Mode (ICM) algorithm, but does not provide a spatial generalization of the DPMM. Typically the simple Potts model considered in [5] (Section IV) cannot be extended to an infinite number of states. Other attempts to combine Potts and DP modelling includes the work in [10] but there the number of states is known to be three and the DP is used instead to model intensity distributions non parametrically.

In this work we build on the approach in [1] which differs from [12, 7, 13] in that it uses a stick-breaking representation for the mixing weights while the latter use a partition model representation. In particular [7] generalizes [12] and proposes a more efficient MCMC inference based on the Swendsen-Wang algorithm, while [13] extends this idea to Hierarchical Dirichlet process priors for multiple image segmentation. Indeed, although the most straightforward prior is the Dirichlet process (DP) prior, it may not be the best choice for segmentation applications and number of components selection. DP induces a highly peaked prior on the number of components which requires a reliable prior information on this number and this is often unavailable. Also, DP tends to produce tiny extra clusters that may be difficult to interpret. Gibbs-type priors [2] are a recent generalization in particular of the Dirichlet and Pitman-Yor process priors. In this class of priors, the existence of additional parameters allows for a better control of the informativeness of the prior on the number of components, which is likely to provide much more flexibility. Segmentation with spatially dependent Pitman-Yor processes has been considered in [14] but using Gaussian processes.

In contrast to [7, 13], we will therefore use a stick-breaking representation for the mixing weights, [potentially] thus providing a richer representation than partition models which integrate out the process. Also stick-breaking representations lead naturally to variational approximations for inference [3] which have the advantage to lower the computational cost in complex data segmentation and not to suffer from label switching complications. For such a generalization, we need to consider the Potts model in a less standard formulation than in [5, 4, 6].

Our goal is therefore to consider several such Gibbs-type priors in an HMRF context and to design the associated infinite HMRF models using a stick-breaking formulation. We will then compare both their respective theoretical properties and their practical performance assessed on various spatial data and image segmentation tasks.

3 Inference using variational approximation

A sampling (MCMC) based inference of similar DP and extensions with MRF models has been proposed in [12, 7]. In this work, we rather consider inference of the proposed models using a variational Bayesian approach. Variational algorithms based on truncation [3, 8] are first considered but other (hybrid) variational approximations [15] and fast implementations [9] are also investigated. The reason for preferring a variational Bayesian approximation instead of Markov chain Monte Carlo inference, is the size and complexity of the data involved in common HMRF applications (e.g., image segmentation) which usually requires inference methods allowing for increased computational efficiency. However, it is interesting to compare such a variational procedure to its Markov chain Monte Carlo counterpart.

References

[1] Mohanad Albughdadi, Lotfi Chaâri, Jean-Yves Tourneret, Florence Forbes, and Philippe Ciuciu. A Bayesian non-parametric hidden Markov random model for hemodynamic brain parcellation. *Signal Processing*, 135:132–146, 2017.

[2] Pierpaolo De Blasi, Stefano Favaro, Antonio Lijoi, Ramsés H. Mena, Igor Prünster, and Matteo Ruggiero. Are Gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(2):212–229, 2015.

[3] David M. Blei and Michael I. Jordan. Variational inference for Dirichlet process mixtures. *Bayesian Anal.*, 1(1):121–143, 03 2006.

[4] Sotirios P. Chatzis. A Markov random field-regulated Pitman-Yor process prior for spatially constrained data clustering. *Pattern Recognition*, 46(6):1595–1603, 2013.

[5] Sotirios P. Chatzis and Gabriel Tsechpenakis. The infinite hidden Markov random field model. *IEEE Trans. Neural Networks*, 21(6):1004–1014, 2010.

[6] Adelino R. Ferreira da Silva. A Dirichlet process mixture model for brain MRI tissue classification. *Medical Image Analysis*, 11(2):169–182, 2007.

[7] R. Y. Da Xu, F. Caron, and A. Doucet. Bayesian nonparametric image segmentation using a generalized Swendsen-Wang algorithm. *ArXiv e-prints*, February 2016.

[8] Nicholas J. Foti, Jason Xu, Dillon Laird, and Emily B. Fox. Stochastic variational inference for hidden Markov models. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 3599–3607, 2014.

[9] Fangjian Guo, Xiangyu Wang, Kai Fan, Tamara Broderick, and David B. Dunson. Boosting variational inference. *CoRR*, abs/1611.05559, 2016.

[10] Timothy D Johnson, Zhuqing Liu, Andreas J Bartsch, and Thomas E Nichols. A Bayesian non-parametric Potts model with application to pre-surgical fMRI data. *Statistical Methods in Medical Research*, 22(4):364–381, 2013.

[11] Jeffrey W. Miller and Matthew T. Harrison. Mixture models with a prior on the number of components. *Journal of the American Statistical Association*, 0(0):1–17, 2017.

[12] Peter Orbanz and Joachim M. Buhmann. Nonparametric Bayesian image segmentation. *International Journal of Computer Vision*, 77(1-3):25–45, 2008.

[13] Jessica Sodjo, Audrey Giremus, Nicolas Dobigeon, and Jean-François Giovannelli. A generalized Swendsen-Wang algorithm for Bayesian nonparametric joint segmentation of multiple images. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1882–1886, La Nouvelle Orléans, LA, United States, March 2017. IEEE.

[14] Erik B. Sudderth and Michael I. Jordan. Shared segmentation of natural scenes using dependent Pitman-Yor processes. In *Advances in Neural Information Processing Systems 21, Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 8-11, 2008*, pages 1585–1592, 2008.

[15] Chong Wang and David M. Blei. Truncation-free stochastic variational inference for Bayesian nonparametric models. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, pages 413–421, 2012.