
Reversible jump MCMC for Bayesian NMF

Mikkel N. Schmidt

Morten Mørup

Cognitive Systems, Informatics and Mathematical Modeling, Technical University of Denmark
 Richard Petersens Plads, DTU Bldg. 321, DK-2800 Kgs. Lyngby, Denmark

In non-negative matrix factorization (NMF), a data matrix \mathbf{X} is modelled as the product of two matrices \mathbf{A} and \mathbf{B} with non-negative entries. Taking a Bayesian approach to NMF entails specifying an appropriate likelihood and priors for the factorizing matrices and the number of components K (columns of \mathbf{A} and rows of \mathbf{B} .) Related to Bayesian NMF is also models based on the Indian buffet process [5, 4, 2]. For a given value of K , inference can be performed using standard methods such as Gibbs sampling [7] or variational approximation [1]; however, inferring K is more challenging. Computationally expensive approaches including Chib’s method [7], thermodynamic integration, and reversible jump independence MCMC [8] have been proposed.

Inspired by Jain and Neal’s split-merge procedure [3] for Dirichlet process mixture models, we have developed a procedure [6] for generating efficient cross-dimensional (reversible jump) Metropolis-Hastings proposals in Bayesian NMF, in which we consider two different move types: birth-death and split-merge moves. To outline the basic idea, consider for example a merge move: We first remove the two components that are to be merged. Then we generate one new component from the prior and refine it through a number of Gibbs sampling rounds restricted to the new component. These rounds of restricted Gibbs sampling serve the purpose of moving the new component into a region of high posterior probability, and the resulting component constitutes what is referred to as the launch state. Then, one final round of restricted Gibbs sampling is performed, keeping track of all transition probabilities. Thus, the proposal is a Gibbs random walk starting at the launch state. Next, to evaluate the Metropolis-Hastings acceptance probability, we must compute the reverse transition probability, i.e. the probability of splitting the new component into the two original ones. This is done by removing the newly added component and generating and refining two launch state components. In the final round of “Gibbs sampling” for the two launched components, they are forced to end up equal to the original components and the reverse transition probability can thus be computed.

We believe the idea of using random walk proposals from highly probable launch states is generally applicable in a number of variable-dimension latent feature models. Our future research includes utilizing numerical optimization algorithms for computing launch states; considering more general remove- i -add- j moves; and applying the ideas to other latent feature models.

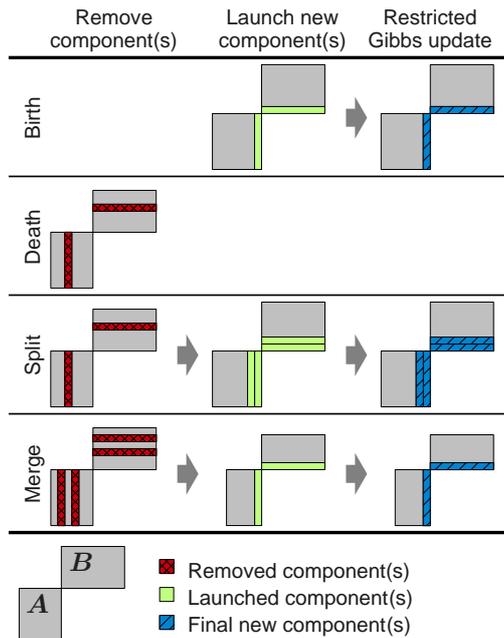


Figure 1: Illustration of birth-death and split-merge proposals: Components are removed, and new components are generated and refined through restricted Gibbs sampling. The final proposed components are generated through one last restricted Gibbs sweep. The proposal density is computed by keeping track of all transition probabilities in the final Gibbs sweep.

References

- [1] Ali Taylan Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, 2009.
- [2] Nan Ding, Yuan (Alan) Qi, Rongjing Xiang, Ian Molloy, and Ninghui Li. Nonparametric Bayesian matrix factorization by power-EP. In *Artificial Intelligence and Statistics, International Conference on (AISTATS)*, 2010.
- [3] Sonia Jain and Radford M. Neal. A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. *Journal of Computational and Graphical Statistics*, 13(1):158–182, 2004.
- [4] David Knowles and Zoubin Ghahramani. Infinite sparse factor analysis and infinite independent components analysis. In *Independent Component Analysis and Signal Separation, International Conference on (ICA)*, volume 4666 of *Lecture Notes in Computer Science Series (LNCS)*, pages 381–388. Springer, 2007.
- [5] Edward Meeds, Zoubin Ghahramani, Radford M. Neal, and Sam T. Roweis. Modeling dyadic data with binary latent factors. In *Advances in Neural Information Processing Systems (NIPS)*, volume 19, pages 977–984, 2007.
- [6] Mikkel N. Schmidt and Morten Mørup. Infinite non-negative matrix factorization. In *European Signal Processing Conference (EUSIPCO)*, Aug 2010.
- [7] Mikkel N. Schmidt, Ole Winther, and Lars Kai Hansen. Bayesian non-negative matrix factorization. In *Independent Component Analysis and Signal Separation, International Conference on (ICA)*, volume 5441 of *Lecture Notes in Computer Science (LNCS)*, pages 540–547. Springer, 2009.
- [8] Mingjun Zhong and Mark Girolami. Reversible jump MCMC for non-negative matrix factorization. In *Artificial Intelligence and Statistics, International Conference on (AISTATS)*, volume 5, pages 663–670, 2009.