Scalable Inference in Dynamic Admixture Models

Patrick Jähnichen, Florian Wenzel, and Marius Kloft
{jaehnicp, wenzelfl, kloft}@hu-berlin.de

Machine Learning Group
Humboldt-University of Berlin
Germany

Dynamic probabilistic models are standard in various time-series applications, including weather forecasting, stock market analysis, and robotics. Typically such models consist of a diffusion model that governs the state of the system and a model of measuring this state. As an example consider the simple non-mixture time series model

$$\mu_t = \mu_{t-1} + v_t, v_t \sim \mathcal{N}(0, \nu^2)$$
$$x_t = \mu_t + w_t, w_t \sim \mathcal{N}(0, \sigma^2).$$

where μ_t is the state of the system at time t and x_t is a noisy measurement of that state. Note that this kind of model is akin to the well-known Kalman filter.

A drawback of such a simple model is that it does not capture data that is a mixture of several possible components in varying proportions. Such data emerges in e.g. corpora modeling where each document is comprised of words that are generated by different themes that are present in the corpus and underly a time dynamic. An example of a more complex model is the continuous time dynamic topic model (cDTM) [6]. In this model the time structure of the mixture components is modeled in terms of a Markov chain. We generalize this approach to general Gaussian processes (GPs). This allows for more flexible modeling of the diffusion process (time structure) by changing the GP covariance function, capturing a wider variety/combination of mixture component dynamics.

Inference in these models is a major challenge. The posterior we seek is generally intractable and we must appeal to an approximation. Up until recently, state-of-the-art approaches used variational inference as in [3, 6] and our own preliminary research [4]. As these approaches are limited to rather small datasets, [2] recently applied a stochastic gradient Langevin dynamics sampler [7] which allows for inference in these models using larger numbers of datapoints. However, [1, 5] have shown that this approach is amenable to considerable improvements.

We develop an inference method which is based on more evolved stochastic gradient based sampling techniques (as e.g. [2]) leading to a novel robust inference method which is applicable to millions of data points. Our preliminary empirical findings suggest that we can improve performance in terms of accuracy and speed over the state-of-the-art methods.

Bibliography

[1] S. Ahn, A. Korattikara, and M. Welling. Bayesian posterior sampling via stochastic gradient Fisher scoring. In *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, 2012.

[2] A. Bhadury, J. Chen, J. Zhu, and S. Liu. Scaling up Dynamic Topic Models. In *Proceedings of the 25th International . . .* , 2016.

[3] D. M. Blei and J. D. Lafferty. Dynamic topic models. *Proceedings of the 23rd International Conference on Machine Learning*, 2006.

[4] P. Jähnichen. *Time dynamic topic models.* PhD thesis, Leipzig University, Leipzig, Mar. 2016.

[5] S. Mandt, M. D. Hoffman, and D. M. Blei. A Variational Analysis of Stochastic Gradient Algorithms. *arXiv.org*, 2016.

[6] C. Wang, D. M. Blei, and D. Heckerman. Continuous time dynamic topic models. *Proc. of UAI*, 2008.

[7] M. Welling and Y.-W. Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th International Conference on Machine Learning*, pages 681–688, 2011.