Stochastic Divergence Minimization for Biterm Topic Models

Zhenghang Cui
The University of Tokyo

Issei Sato
The University of Tokyo / RIKEN

Masashi Sugiyama
RIKEN / The University of Tokyo

With the emergence and the thriving development of social networks, a huge amount of short texts are accumulated and need to be processed. Inferring latent topics of collected short texts is useful for understanding its hidden structure. Latent Dirichlet allocation (Blei, Ng, and Jordan 2003) is a traditional generative probabilistic model, which approaches to model latent topics based on word co-occurrences. However, due to the lack of document-level word co-occurrences, LDA usually fails to show satisfactory performance on data sets with short document length. The biterm topic model (BTM) (Cheng et al. 2014) was recently proposed to overcome this sparseness of document-level word co-occurrences by modifying the word generating part of the model and directly modeling the generation process of word pairs. Instead of each word, the generation process of unordered combinations of two words, or biterms, is modeled under the BTM context. Compared to conventional topic models, this modification makes BTM less sensitive to the shortness of each document, and these word pairs are more stable to clearly reveal the relationship between words.

Since it is difficult to estimate latent topics as a closed form solution for BTM, a collapsed Gibbs sampling algorithm is used to approximate the true posterior (Cheng et al. 2014). Its main inference process is drawing samples from the posterior distribution of a single variable conditioned on all other variables. Based on the batch inference algorithm mentioned above, two online inference algorithms are also developed (Cheng et al. 2014). One is based on the idea of updating hyperparameters between time slices, which is inspired by the online LDA algorithm, while the other is based on the idea of resampling topics of observed biterms for a sufficient number of times after a new biterm is observed, which is inspired by an incremental Gibbs sampler for LDA. On the other hand, based on the idea of zero-order stochastic collapsed variational Bayesian inference (SCVB0) for LDA, a similar SCVB0 algorithm for BTM was proposed for better latent topics estimation (Awaya et al. 2016). However, the two online algorithms are not working very efficiently on memory usage, and the SCVB0 algorithm relies on very crude estimation.

To cope with the above problems, we proposed the stochastic divergence minimization (SDM) (Sato and Nakagawa 2015) inference algorithm for learning BTM. The original SDM for LDA was proposed to achieve better predictive performance than traditional inference algorithms. However, since its derivation is specific to LDA, we have to newly derive SDM for BTM. Similarly to idea of collapsed Gibbs sampling inference algorithms, we first marginalize out model parameters to obtain the distribution of the observed data and remaining hidden variables. Assuming the distribution we approximate can be fully factorized, we try to find the best distribution that has the minimum α-divergence from the true posterior. Since the minimization is intractable to perform analytically, we reparameterize the distribution we approximate as a product of multiple terms, and then use local divergence minimization to approximate each term independently. In the multiple terms we defined in reparameterization, one term can be used to restore other terms. By updating this term using fixed point iteration, we can derive a stochastic divergence minimization algorithm for BTM.

In this work, we reviewed a generative probabilistic model designed for analyzing short texts, the biterm topic model, and its three existing online inference algorithms. We then developed a stochastic divergence minimization algorithm, and demonstrated that it outperforms existing online inference algorithms for BTM in experiments. More details are available in the arXiv paper [1].

References

Awaya, N.; Kitazono, J.; Omori, T.; and Ozawa, S. 2016. Stochastic collapsed variational bayesian inference for biterm topic model. In *2016 International Joint Conference on Neural Networks (IJCNN)*, 3364–3370.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3:993–1022.

Cheng, X.; Yan, X.; Lan, Y.; and Guo, J. 2014. Btm: Topic modeling over short texts. *IEEE Transactions on Knowledge and Data Engineering* 26(12):2928–2941.

Sato, I., and Nakagawa, H. 2015. Stochastic divergence minimization for online collapsed variational bayes zero inference of latent dirichlet allocation. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, 1035–1044.

[1] `https://arxiv.org/abs/1705.00394`