Online Learning for Mixed Membership Network Models

Prem Gopalan, David Mimno, Michael J. Freedman and David M. Blei

9 September 2011

Introduction. In this era of "Big Data", there is intense interest in analyzing large networks using statistical models. Applications range from community detection in online social networks to predicting the functions of a protein. MMSB [1] is a powerful mixed-membership model for learning communities and their interactions. It assigns nodes to multiple communities rather than simple clusters. Posterior inference for MMSB is intractable, and approximate inference algorithms such as variational inference or MCMC sampling are applied. However, these methods require multiple passes through the data, and do not easily work with streaming data. Inspired by the recent work on online variational Bayes for LDA [2], we develop an online variational Bayes algorithm for MMSB based on stochastic optimization.

A mixed-membership model. MMSB is a Bayesian probabilistic model of relational data that assumes context dependent membership of nodes in K groups, and that each interaction can be explained by two interacting groups. Given the groups, the generative process draws a K-dimensional mixed-membership vector $\pi_a \sim \text{Dir}(\alpha)$ for each node a and per-interaction membership indicators $z_{a\to b}$, $z_{a\leftarrow b}$ for each binary pair $y_{a,b}$. The indicators are used to index into a blockmodel matrix $B_{K\times K}$ of Bernoulli rates, and $y_{a,b}$ is drawn from it. The only observed variables in this model are $y_{a,b}$.

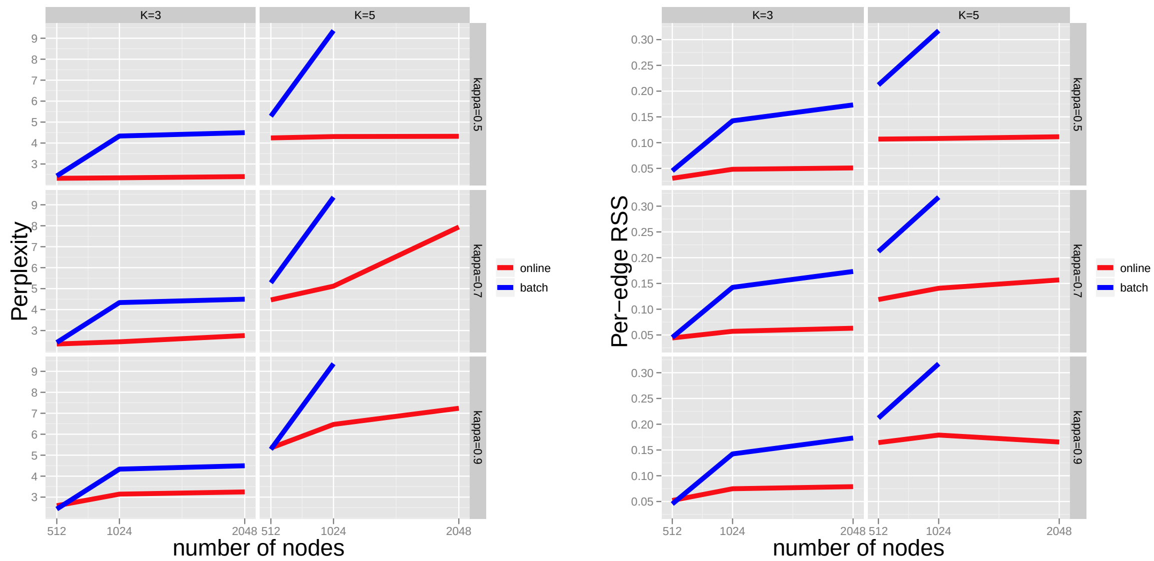
There is a degree of non-identifiability in MMSB due to both π_p and B competing to explain reciprocated interactions. If communities are known to be densely connected internally with sparse external interactions, or have only reciprocated interactions, as is the case with some online social networks, then a simpler model suffices. We replace B with K intragroup interaction rates $\beta_k \sim \text{Beta}(\eta_k)$ and a small, fixed interaction rate ε between distinct groups.

Posterior inference. In variational Bayes for MMSB, the true posterior is approximated by a simpler distribution $q(\beta, \pi, \mathbf{z}_\to, \mathbf{z}_\leftarrow | \gamma, \phi^\to, \phi^\leftarrow, \lambda)$. Following [1], we choose a fully factorized distribution q of the form $q(z_{a\to b} = k) = \phi_{a\to b,k}$, $q(\pi_p) = \text{Dir}(\pi_p; \gamma_p)$ and $q(\beta_k) = \text{Dir}(\beta_k; \lambda_k)$. We then apply stochastic optimization to the variation objective. We subsample the interactions $y_{a,b}$, compute an approximate gradient and follow the gradient with decreasing step-size. Options for subsampling include selecting a node or a pair of interactions uniformly at random, or sampling by exploration where we first select a node uniformly at random and then explore its neighbors. We derive a first-order stochastic natural gradient algorithm for MMSB below assuming random pair sampling.

Define $f(y_{a,b}, \beta_k) = \beta_k^{y_{a,b}}.(1-\beta_k)^{(1-y_{a,b})}$. Initialize γ, λ.
for $t = 0$ to ∞ **do**
    E step $\forall$ (a,b) *in mini-batch* S. Initialize $\phi^t_{a\to b}$, $\phi^t_{a\leftarrow b}$.
    repeat
        Set $g(\phi, k) = \sum_{i\neq k} \phi_i^{t-1} \log f(y_{a,b}, \varepsilon)$
        Set $\phi^t_{a\to b,k} \propto exp\{E_q[\log \pi_{a,k}] + \phi^{t-1}_{a\leftarrow b,k} E_q[\log f(y_{a,b}, \beta_k)] + g(\phi_{a\leftarrow b}, k)\}$ $\forall k$
        Set $\phi^t_{a\leftarrow b,k} \propto exp\{E_q[\log \pi_{b,k}] + \phi^{t-1}_{a\to b,k} E_q[\log f(y_{a,b}, \beta_k)] + g(\phi_{a\to b}, k)\}$ $\forall k$
    until convergence
    M step
    Compute $\tilde{\gamma}_{a,k} = \alpha_k + \frac{N(N-1)}{|S|} \sum_S (\phi^t_{a\to .,k} + \phi^t_{.\leftarrow a,k})$ $\forall k$, $\forall a$
    Compute $\tilde{\lambda}_{k,i} = \eta_{k,i} + \frac{N(N-1)}{|S|} \sum_S (\phi^t_{a\to b,k} \phi^t_{a\leftarrow b,k} y_{a,b,i}) \forall i \in (0,1), \forall k$
    Set $\gamma = (1-\rho_t')\gamma + \rho_t'\tilde{\gamma}$. Set $\lambda = (1-\rho_t)\lambda + \rho_t\tilde{\lambda}$
end for

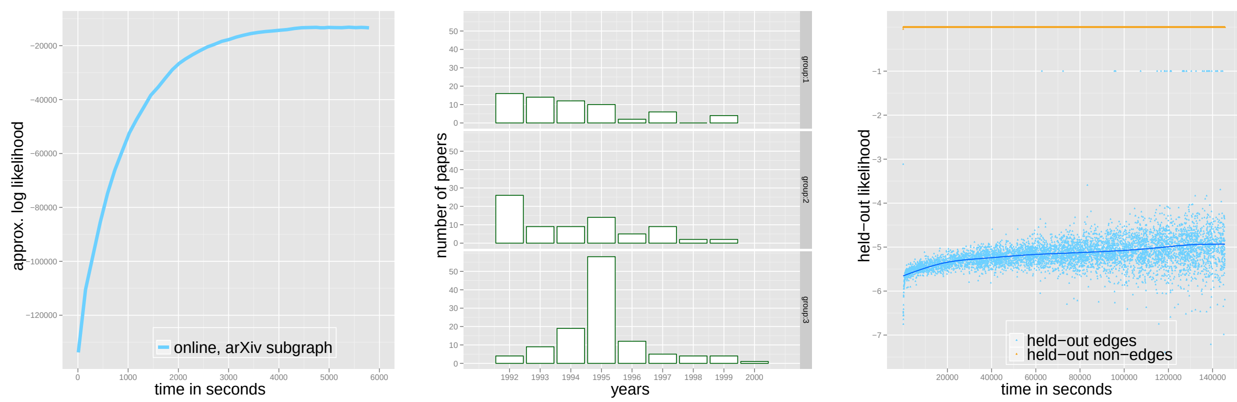
Figure 1: Online variational Bayes for MMSB

Despite conceptual and notational similarities with online LDA [2], online MMSB faces unique challenges. First, the online LDA E step finds locally optimal values of parameters associated with a selected document holding the topics fixed. In MMSB a given node a's γ_a cannot be optimized independently of other nodes. Therefore, we derive updates for both γ and λ. Second, the dimension of γ, the number of



(a) (b)

Figure 2: Perplexity (left) and accuracy (right) comparisons of batch and online MMSB on simulated networks for various node sizes and learning parameter κ, run for 4 hours each. Accuracy is measured as the RSS between two sets of Hellinger distances, one based on true π and the other based on $E[\pi]$. The distances are computed only for $y_{a,b} = 1$. In both plots, the lower the value the better the performance of the algorithms. For both K=3 and K=5, online has lower perplexity than batch as N increases, and approximately equal perplexity for N=512. Accuracy becomes significantly poorer for batch as N increases compared to online. Batch values for N=2048 are unavailable. They took long after 4 hours to compute.



(a) (b) (c)

Figure 3: (a) and (b) show the convergence and resulting groups (shown are histograms of publications by years) respectively on a 256 node subgraph of arXiv citation dataset. (c) shows the convergence of held-out likelihood computed over incoming mini-batches on a 4096 node subgraph of arxiv citation dataset. We set K = 3 in both cases.

nodes, can be very large in contrast to LDA's topics. Finally, MMSB can leverage efficient subsampling strategies, such as forest fire sampling, to exploit network structure.

Preliminary results on real datasets We ran online MMSB on a complete subgraph of 256 nodes from the Arxiv high-energy physics citation graph and obtained reasonable groups. The histogram of publications by years in each group is shown in Fig 3(b). Group 3 consists of many publications in 1995. The average frequency of certain top words (`Calabi-Yau, string, symmetry` etc.) was 150 – 180% of that in Groups 1 and 2. 1995 marked the beginning of the *second superstring revolution* with a flurry of research activity spurred by M-theory. There is a greater frequency of singular memberships in group 3. We are currently evaluating results on a subgraph of 4096 nodes.

References

[1] E. Airoldi, D. Blei, S. Fienberg, and E. Xing. *Mixed membership stochastic blockmodels*, Journal of Machine Learning Research, 9:1981–2014, 2008.

[2] M. Hoffman, D. Blei, and F. Bach. *Online learning for latent Dirichlet allocation*, Neural Information Processing Systems, 2010.