Coupled Probabilistic Latent Semantic Analysis

John McCrae

October 26, 2012

1 Preliminaries

Assume the observed variables are

$$X = (w^l_{jn}); w^l_{jn} \in \{1, \dots V\}$$

Where there are J documents each of length N^l_j and $l \in \{1, 2\}$. The latent variables are

$$Y = (z^l_{jn}); z^l_{jn} \in \{1, \dots K\}$$

2 Model definition

Assume we have the following model variables

$$\Theta = (\phi^l_{wk}, \theta^l_{kj})$$

And that the model is given as

$$p_\Theta(X, Y) = \prod_j^J \prod_l^2 \prod_n^{N^l_j} \phi^l_{w^l_{jn} z^l_{jn}} \theta^l_{z^l_{jn} j}$$

And we have the marginal distribution as follows:

$$p_\Theta(X) = \prod_j^J \prod_l^2 \prod_n^{N^l_j} \prod_k^K \phi^l_{w^l_{jn} k} \theta^l_{kj}$$

Following [1] we define the posterior condition as

$$\begin{aligned}
\pi_{jk}(w_j, z_j) &= \frac{\sum_n^{N^2_j} \mathbb{1}(z^1_{jn} = k)}{N^1_j} - \frac{\sum_n^{N^2_j} \mathbb{1}(z^2_{jn} = k)}{N^2_j} \\
&= \frac{N^1_{jk}}{N^1_j} - \frac{N^2_{jk}}{N^2_j}
\end{aligned}$$

And that

$$\pi(X, Y) = \sum_j^J \sum_k^K \pi_{jk}(w_j, z_j)$$

Hence, the problem is to find Θ^* for some ϵ such that

$$\Theta^* = \arg\max_{\Theta} p_\Theta(Y|X)$$

Such that

$$E_\Theta[\pi(X,Y)] \leq \epsilon$$

3 Solution

By [2], this can be solved by an E-M procedure where

$$\textbf{E:} \quad q^{t+1} = \arg\min_{q} KL(q(Y)||p_{\Theta^t}(Y|X))$$

$$\textbf{M:} \quad \Theta^{t+1} = \arg\max_{\Theta} E_{q^{t+1}}[log_\Theta(p(X,Y))]$$

Where KL is as usual:

$$KL(p(X)||q(X)) = \sum_X p(x) \log\left(\frac{p(x)}{q(x)}\right)$$

And $q(Y)$ is constrained by

$$q(Y) : E_q[\pi(X,Y)] \leq \epsilon$$

By Lagrangian duality, the problem of solving

$$\arg\min_{q} KL(q(Y)||p_\Theta(Y|X))$$

Subject to

$$E_q[\pi(X,Y)] \leq \xi; ||\xi|| < \epsilon$$

Is equivalent to finding

$$q^*(z_j) = \frac{p_\Theta(z_j|X)\exp(-\lambda^*\theta(X,Y))}{\zeta(\lambda^*)}$$

Where

$$\lambda^* = \arg\max_{\lambda \geq 0} -log(\zeta(\lambda)) - \epsilon\lambda$$

And

$$\zeta(\lambda) = \sum_Y p_\Theta(Y|X)\exp(-\lambda\pi(X,Y))$$

$$= \sum_j^J \sum_l^2 p_\Theta(z_j^l|X)e^{-\lambda\pi_j(X,z_j)}$$

It follows by differentiation

$$-\frac{\zeta'(\lambda^*)}{\zeta(\lambda^*)} - \epsilon = 0$$

And the solution to this can be found by Newtonian gradient descent

Given we have λ^* then we can apply

$$q^*(z_j) \propto p_\Theta(z_j|X)exp(-\lambda^*\pi(X,Y))$$

And hence we can use a Gibbs-like sampling

$$p_\Theta(z_{jn}^{tl} = k|X) \propto \phi_{x_{jn}k}^{tl}\theta_{kj}^{tl}exp(-\lambda^*\pi_{jk}(w_j,z_j))$$

Finally the maximization step is trivial

$$\phi_{wk}^l = \frac{N_{wk}^l}{N_k^l}$$

$$\theta_{kj}^l = \frac{N_{kj}^l}{N_k^l}$$

4 Dirichlet assumption

The Dirichlet assumption is that

$$\theta_j^l \sim Dir(\alpha)$$

$$\phi_k^l \sim Dir(\beta)$$

The result of which is that the maximization step is generalized to [3]

$$\theta_{kj}^l = \frac{N_{kj}^l + \alpha}{N_k^l + K\alpha}$$

$$\phi_{wk}^l = \frac{N_{wk}^l + \beta}{N_k^l + V\beta}$$

5 Initialization

1. Sort words by frequency $\{w'_1, \dots w'_V\}$
2. Find $K-1$ values, κ_i, such that $\kappa_i < \kappa_{i+1}$ and $\sum_{j=\kappa_i}^{\kappa_{i+1}} w'_j \leq \frac{N}{K}$
3. Initialize $z_{jn}^l = k$ where $w_{jn}^l = w'_i$ and $\kappa_k \leq i < \kappa_{i+1}$
4. Calculate initial Θ

6 Notes

[1] used 500 iterations at $\alpha = 1.1$, $\beta = 1.01$.

References

[1] John C. Platt and Kristina Toutanova and Wen-tau Yih (2010). Translingual Document Representations from Discriminative Projections.

[2] Kuzman Ganchev and João Graça and Jennifer Gillenwater and Ben Taskar (2010). Posterior Regularization for Structured Latent Variable Models.

[3] Ian Porteous and David Newman and Alexander Ihler and Arthur Asuncion and Padhraic Smyth and Max Welling (2008). Fast Collapsed Gibbs Sampling For Latent Dirichlet Allocation.