# **Extension** of CTR

Valerio Perrone

March 10, 2017

## 1 Model

Assume there are $K$ topics $\beta := \beta_{1:K}$ and that these are the same for users and items.

1. For each user $i$,
   (a) Draw topic proportions $\theta_i^u \sim \text{Dirichlet}(\alpha)$.
   (b) Draw item latent offset $\epsilon_i^u \sim \text{N}(0, \lambda_u^{-1} I_k)$ and set the item latent vector as $\boldsymbol{u_i = \epsilon_i^u + \theta_i^u}$.
   (c) For each word $w_{in}^u$.
      i. Draw topic assignment $z_{im}^u \sim \text{Mult}(\theta^u)$.
      ii. Draw word $w_{im}^u \sim \text{Mult}(\beta_{z_{im}^u})$.
2. For each item $j$,
   (a) Draw topic proportions $\theta_j^v \sim \text{Dirichlet}(\alpha)$.
   (b) Draw item latent offset $\epsilon_j^v \sim \text{N}(0, \lambda_v^{-1} I_k)$ and set the item latent vector as $v_j = \epsilon_j^v + \theta_j^v$.
   (c) For each word $w_{jn}^v$.
      i. Draw topic assignment $z_{jn}^v \sim \text{Mult}(\theta)$.
      ii. Draw word $w_{jn}^v \sim \text{Mult}(\beta_{z_{jn}})$.
3. For each user-item pair $(i, j)$, draw the rating $r_{ij} \sim \text{N}(u_i^T v_j, c_{ij}^{-1})$.

## 2 Inference

This is an EM-style algorithm to learn MAP estimates. The goal is to maximize the complete log likelihood

$$L = -\frac{\lambda_u}{2}\sum_i \boldsymbol{(u_i - \theta_i^u)^T(u_i - \theta_i^u)} - \frac{\lambda_v}{2}\sum_j (v_j - \theta_j^v)^T(v_j - \theta_j^v)$$
$$+ \boldsymbol{\sum_i \sum_m log(\sum_k \theta_{ik}^u \beta_{k,w_{im}^u})} + \sum_j \sum_n log(\sum_k \theta_{jk}^v \beta_{k,w_{jn}^v}) - \sum_{i,j} \frac{c_{i,j}}{2}(r_{ij} - u_i^T v_j)^2.$$

The updates are the following:

1. Given $\theta_j^v, \theta_i^u \rightarrow$ update $u_i$ and $v_j$:

$$u_i \leftarrow (VC_iV^T + \lambda_u I_k)^{-1}(VC_iR_i \boldsymbol{+ \lambda_u \theta_i^u})$$
$$v_i \leftarrow (UC_jU^T + \lambda_v I_k)^{-1}(UC_jR_j + \lambda_v \theta_j^v)$$

   where $C_i$ is the diagonal matrix with $c_{ij}$ on the diagonal and $R_i = (r_{ij})_{j=1}^J$ for user $i$.

   Note: the following is only needed if we want to update $\theta$.

2. Given $U, V \rightarrow$ update $\theta_j^v$ and $\theta_i^u$ by projection gradient using the following[1]:

$$L(\theta_j^v) \geq -\frac{\lambda_v}{2}(v_j - \theta_j^v)^T(v_j - \theta_j^v) + \sum_n \sum_k \phi_{jnk}^v(log\theta_{jk}^v \beta_{k,w_{jn}^v} - log\phi_{jnk}^v)$$
$$\boldsymbol{L(\theta_i^u) \geq -\frac{\lambda_u}{2}(u_i - \theta_i^u)^T(u_i - \theta_i^u) + \sum_m \sum_k \phi_{ink}^u(log\theta_{ik}^u \beta_{k,w_{im}^u} - log\phi_{ink}^u)}$$

   given that the optimal $\phi_{jnk}^v$, $\phi_{imk}^u$ and $\beta$ satisfy:

$$\phi_{jnk}^v \propto \theta_{jk}^v \beta_{k,w_{jn}^v}$$
$$\boldsymbol{\phi_{imk}^u} \propto \boldsymbol{\theta_{ik}^u \beta_{k,w_{im}^u}}$$
$$\boldsymbol{\beta_{kw}} \propto C_1 \sum_j \sum_n \phi_{jnk} 1[w_{jn} = w] \boldsymbol{+ C_2 \sum_i \sum_m \phi_{imk} 1[w_{im} = w]}$$

   where $C1$ and $C2$ follow from Appendix A.4.1 in the Latent Dirichlet Allocation paper.

[1] After defining $\phi_{jnk}$ as the probability that word $z_{jn}$ (i.e. the $n$th word in document $j$) is assigned to topic $k$.