Supplementary material for: "Content-Based Social Recommendation with Poisson Matrix Factorization"

Eliezer de Souza da Silva, Helge Langseth and Heri Ramampiaro

1 Variational inference

- **Variational updates for $q(\boldsymbol{Y}_{dv})$:**

  We know that the complete conditional of $\boldsymbol{Y}_{dv}$ and $\boldsymbol{Z}_{ud}$ are Multinomial distributions, so we will take a look in general form of the variational updates when we have a multinomial complete conditional.

$$
\begin{aligned}
p(\boldsymbol{x}|*) &= \text{Mult}(\boldsymbol{x}|m, p = (p_1, \cdots, p_n)) \\
\Rightarrow \log p(\boldsymbol{x}|*) &= \log m! + \sum_{i=1}^{n} x_i \log(p_i) - \log(x_i!) \\
\Rightarrow \text{E}_{q(p_1,\cdots,p_n)} \log p(\boldsymbol{x}|*) &= \log m! + \sum_{i=1}^{n} x_i \, \text{E}_{q(p_i)}[\log(p_i)] - \log(x_i!) \\
\Rightarrow \exp\{\text{E}_{q(p_1,\cdots,p_n)} \log p(\boldsymbol{x}|*)\} &= m! \prod_{i=1}^{n} \frac{\exp\{\text{E}_{q(p_i)}[\log(p_i)]\}^{x_i}}{x_i!} \\
\Rightarrow q(\boldsymbol{x}) &= \text{Mult}(\boldsymbol{x}|m, \boldsymbol{p}^* = (p_1^*, \cdots, p_n^*)) \\
\text{with } p_i^* &= \exp(\text{E}_{q(p_i)}[\log(p_i)]) \\
\text{and } \sum_i p_i^* &= 1
\end{aligned}
$$

The above result is saying that if the complete conditional of a variable is a multinomial, the approximate variational distribuition is also a multinomial with parameters proportional to $\exp(\text{E}_{q(p_i)}[\log(p_i)])$ (or alternatively, the log of the parameters proportional to $\text{E}_{q(p_i)}[\log(p_i)]$). Now we shall apply this result to $\boldsymbol{Y}_{dv}$.

$$
\begin{aligned}
\boldsymbol{Y}_{dv}|* &\sim \text{Mult}(W_{dv}; \boldsymbol{\phi}_{dv} = (\phi_{dv,1}, \cdots, \phi_{dv,K})) \\
\text{with } \phi_{dv,k} &= \frac{\beta_{vk}\theta_{dk}}{\sum_k \beta_{vk}\theta_{dk}} \\
\Rightarrow q(\boldsymbol{Y}_{dv}) &= \text{Mult}(W_{dv}; \boldsymbol{\phi}^*_{dv} = (\phi^*_{dv,1}, \cdots, \phi^*_{dv,K})) \\
\Rightarrow \log \phi^*_{dv,k} &= \text{E}_q[\log(\beta_{vk}) + \log(\theta_{dk})] - \text{E}_q[\log(\sum_k \beta_{vk}\theta_{dk})]
\end{aligned}
$$

From the multinomial properties we know that $\sum_k \phi^*_{dv,k} = 1$, implying that $e^{\text{E}_q[\log(\sum_k \beta_{vk}\theta_{dk})]} = \sum_k e^{\text{E}_q[\log(\beta_{vk})+\log(\theta_{dk})]}$, in other words, we can compute $\phi^*_{dv,k} \propto \exp\{\text{E}_q[\log(\beta_{vk}) + \log(\theta_{dk})]\}$ and normalize after. From the properties of gamma distribution we have the following results:

$$
\begin{aligned}
\mathrm{E}_q[\log(\beta_{vk})] &= \Psi(a_{\beta_{vk}}) - \log(b_{\beta_{vk}}) \\
\mathrm{E}_q[\log(\theta_{dk})] &= \Psi(a_{\theta_{dk}}) - \log(b_{\theta_{dk}}) \\
\Rightarrow \phi^*_{dv,k} &\propto \frac{\exp\{\Psi(a_{\beta_{vk}})}{b_{\beta_{vk}}} \frac{\Psi(a_{\theta_{dk}})\}}{b_{\theta_{dk}}} \\
\text{with } \sum_k \phi^*_{dv,k} &= 1
\end{aligned}
$$

Note that $\Psi(.)$ is the Digamma function.

- **Variational updates for** $q(\beta_{vk})$: we will look with some detail to the variational updates equations for $q(\beta_{vk})$ because this is be used as a template for all the variational update for variables with Gamma distributed complete conditionals.

$$
\begin{aligned}
\log p(\beta_{vk}|*) &= \log \mathrm{Gamma}(\beta_{vk}|\overbrace{a + \sum_d Y_{dv,k}}^{a^*}, \overbrace{b + \sum_d \theta_{dk}}^{b^*}) \\
&= a^* \log b^* - \log \Gamma(a^*) + \log(\beta_{vk})(a^* - 1) - \beta_{vk} b^* \\
\Rightarrow \mathrm{E}_{q(-\beta_{vk})}[\log p(\beta_{vk}|*)] &= \mathrm{E}_q[a^* \log b^* - \log \Gamma(a^*)] + \log(\beta_{vk}) \mathrm{E}_q[a^* - 1] - \beta_{vk} \mathrm{E}_q[b^*] \\
\Rightarrow q(\beta_{vk}) &\propto \exp\{\mathrm{E}_q[a^* \log b^* - \log \Gamma(a^*)]\} \beta_{vk}^{\mathrm{E}_q[a^*-1]} \exp(-\beta_{vk} \mathrm{E}_q[b^*]) \\
&\propto \beta_{vk}^{\mathrm{E}_q[a^*]-1} \exp(-\beta_{vk} \mathrm{E}_q[b^*]) \\
\Rightarrow q(\beta_{vk}) &= \mathrm{Gamma}(\beta_{vk}|a_{\beta_{vk}}, b_{\beta_{vk}}) \\
\text{with, } a_{\beta_{vk}} &= \mathrm{E}_q[a^*] = a + \sum_d \mathrm{E}_q[Y_{dv,k}] \\
b_{\beta_{vk}} &= \mathrm{E}_q[b^*] = b + \sum_d \mathrm{E}_q[\theta_{dk}]
\end{aligned}
$$

At this point we already have the update equation for the parameter of the approximate variational distribution $q(\beta_{vk})$, however we need to calculate the expected value of variational distribution of $Y_{dv,k}$ and θ_{dk}. Because of the mean field, we assume that we know this distributions (Gamma and Multinomial). So $q(\theta_{dk}) = \mathrm{Gamma}(a_{\theta_{dk}}, b_{\theta_{dk}})$, implying that $E_q[\theta_{dk}] = \frac{a_{\theta_{dk}}}{b_{\theta_{dk}}}$. Also $q(\boldsymbol{Y}_{dv})$ is a multinomial with parameters $\phi^*_{dv,k}$, so $\mathrm{E}_q[Y_{dv,k}] = w_{dv}\phi^*_{dv,k}$.

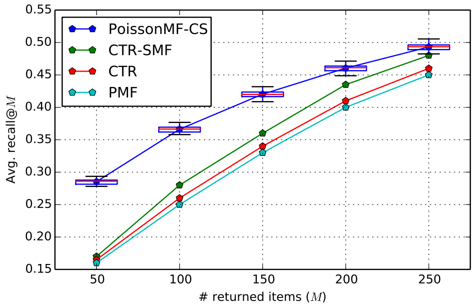
Now we may write the parameter coordinate ascent equation in a closed-form

$$
\begin{aligned}
q(\beta_{vk}) &= \mathrm{Gamma}(\beta_{vk}|a_{\beta_{vk}}, b_{\beta_{vk}}) \\
\text{with, } a_{\beta_{vk}} &= \mathrm{E}_q[a^*] = a + \sum_d w_{dv}\phi^*_{dv,k} \\
b_{\beta_{vk}} &= \mathrm{E}_q[b^*] = b + \sum_d \frac{a_{\theta_{dk}}}{b_{\theta_{dk}}}
\end{aligned}
$$

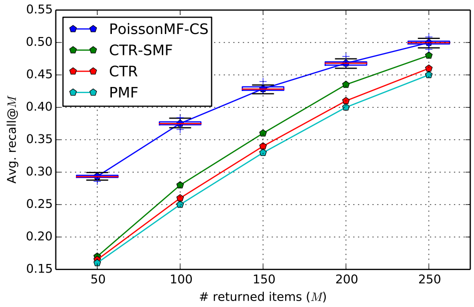
Note that this equations are valid for any variable with a gamma distributed complete conditional. In other words, the approximate variational distribution of a variable with a gamma distributed complete conditional is gamma distributed and the parameter of the approximate distribution is the expected value of the parameters of the complete conditional. So we can apply the same equation to all other latent variables, but making the necessary adjustments to the proper compute $\mathrm{E}_q[a^*]$ and $\mathrm{E}_q[b^*]$.

2 Experiments

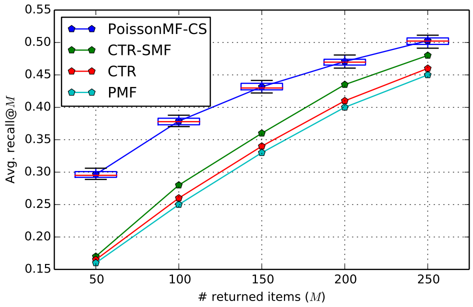
Here we present some plots omitted from the main article.



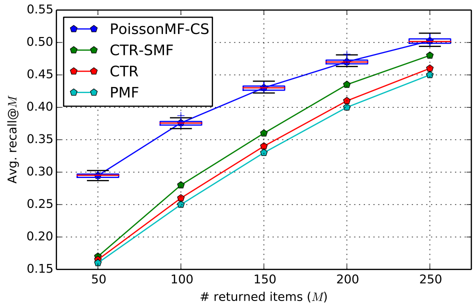
(a) K=5



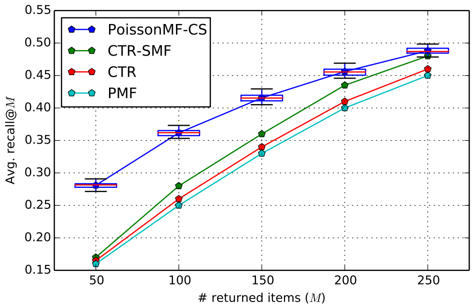
(b) K=10



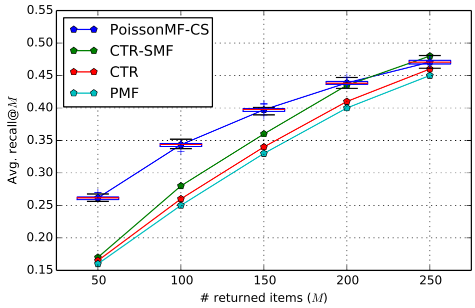
(c) K=15



(d) K=20



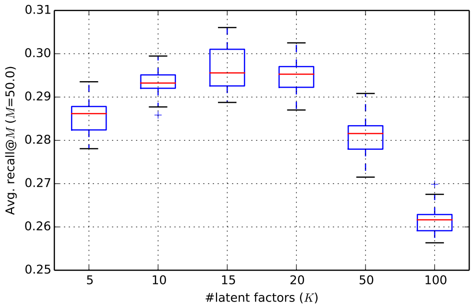
(e) K=50



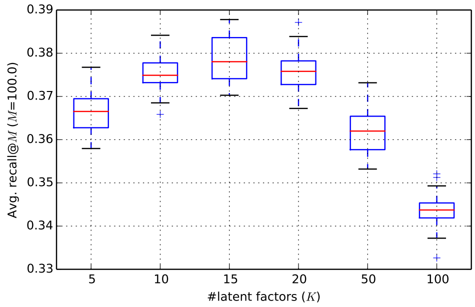
(f) K=100

Figure 1: Comparison of PoissonMF-SC with CTR with SMF, CTR and PMF. Each subplot is the result of running the PoissonMF-SC recommendation algorithm over 30 random splits of the *Hetrec2011-lastfm* dataset for a fixed number of latent features k (in this case, $K = 5, 15, 50, 100$). The values for CTR with SMF, CTR and PMF are taken from [?], and according to the reported results, they are the best values given a grid search of the parameters

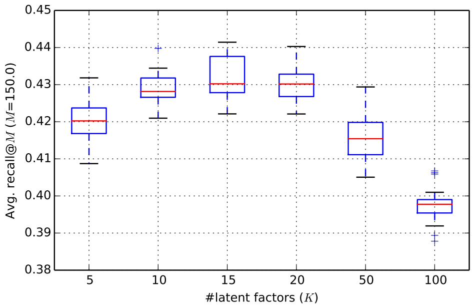
2.0.1 Impact of parameters



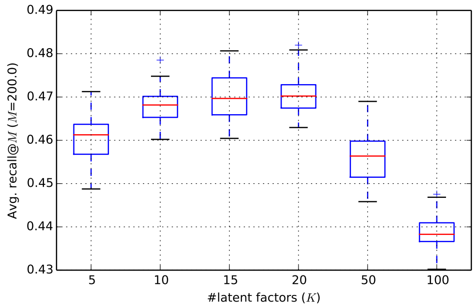
(a) m=50



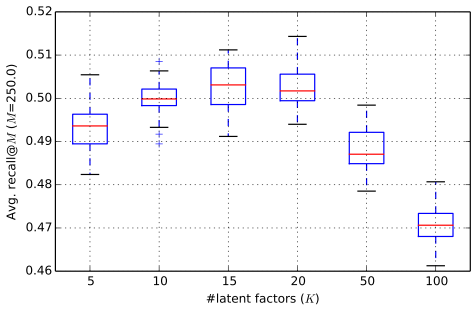
(b) m=100



(c) m=150



(d) m=200



(e) m=250

Figure 2: Impact of the number of latent variables (K) parameter on the Average recall at M metric for different number of returned items (M). Each subplot is the result of running the PoissonMF-SC recommendation algorithm over 30 random splits of the *Hetrec2011-lastfm* dataset with K varying in (5,10,15,20,50,100)