Extensions to Variational Bayesian Approach to Movie Rating Prediction

Yew Jin Lim and Yee Whye Teh

August 14, 2007

1 Future Extensions

1.1 Adapting to the Probe Set by Linear Regression

The distribution of the training set which gave us M is different from that for the probe set, which is in fact the same distribution as for the qualifying set. Thus it is important to adapt the prediction for the probe set.

Here we use the notion that the full training set is used to construct representations U and V for the users and movies respectively; then using these representations to learn about the probe set using linear regression. We will keep the variational Bayesian formalism, treating uncertainty in U, V and the linear regression weights λ properly.

For each (ij) in the probe set, we model m_{ij} as:

$$p(m_{ij}|U, V, \lambda) = \frac{1}{\sqrt{2\pi\xi^2}} \exp\left(-\frac{1}{2}\frac{(m_{ij} - u_i^\top v_j - \lambda^\top(u_i \otimes v_j))^2}{\xi^2}\right) \tag{1}$$

That is, the major predictor for m_{ij} is still the usual $u_i^\top v_j$, but we have a variation term $\lambda^\top(u_i \otimes v_j)$ that is specific to the probe set. In other words, instead of modelling $M \approx U^\top V$, we model $M \approx U^\top V + U^\top \Lambda V$, where Λ is a $n \times n$ matrix with rows given by consecutive series of n entries in λ (conversely, λ is Λ row vectorized, i.e. $\lambda^\top$ is formed by first concatenating rows of Λ together. We write $\lambda = \mathrm{rvec}[\Lambda]$).

We give a prior for λ where each entry is independent and is a Gaussian with zero mean and variance ω^2. The prior term is:

$$p(\lambda) = \prod_{l=1}^{n^2} \frac{1}{\sqrt{2\pi\omega^2}} \exp\left(-\frac{1}{2}\frac{\lambda_l^2}{\omega^2}\right) \tag{2}$$

Now we keep the variational posterior for U and V fixed, and make the assumption that λ is independent from U and V in the posterior. To now get $Q(\lambda)$, we collect appropriate terms and take expectations with respect to $Q(U)$

and $Q(V)$. The appropriate terms are:

$$-\frac{1}{2}\left(\frac{\lambda^\top\lambda}{\omega^2}+\sum_{(ij)}\frac{(m_{ij}-u_i^\top v_j-\lambda^\top(u_i\otimes v_j))^2}{\xi^2}\right)$$

where (ij) now ranges only over the probe set. The above is, up to additive factors constant in λ,

$$\begin{aligned}
&=-\frac{1}{2}\left(\frac{\lambda^\top\lambda}{\omega^2}+\sum_{(ij)}\frac{\lambda^\top(u_i\otimes v_j)(u_i\otimes v_j)^\top\lambda-2\lambda^\top(m_{ij}-u_i^\top v_j)(u_i\otimes v_j)}{\xi^2}\right)\\
&=-\frac{1}{2}\left(\frac{\lambda^\top\lambda}{\omega^2}+\sum_{(ij)}\frac{\lambda^\top(u_iu_i^\top\otimes v_jv_j^\top)\lambda-2\lambda^\top(m_{ij}(u_i\otimes v_j)-\mathrm{rvec}[u_iu_i^\top v_jv_j^\top])}{\xi^2}\right)
\end{aligned}\tag{3}$$

Thus we have that $Q(\lambda)$ is Gaussian with covariance and mean given by:

$$\begin{aligned}
\Xi &= \left(\begin{pmatrix}\frac{1}{\omega^2} & & 0\\ & \ddots & \\ 0 & & \frac{1}{\omega^2}\end{pmatrix}+\sum_{(ij)}\frac{E_{Q(U)}[u_iu_i^\top]\otimes E_{Q(V)}[v_jv_j^\top]}{\xi^2}\right)^{-1}\\
&= \left(\begin{pmatrix}\frac{1}{\omega^2} & & 0\\ & \ddots & \\ 0 & & \frac{1}{\omega^2}\end{pmatrix}+\sum_{(ij)}\frac{(\Phi_i+\overline{u_iu_i}^\top)\otimes(\Psi_j+\overline{v_jv_j}^\top)}{\xi^2}\right)^{-1} \qquad (4)\\
\overline{\lambda} &= \Xi\left(\sum_{(ij)}\frac{m_{ij}(E_{Q(U)}[u_i]\otimes E_{Q(V)}[v_j])-\mathrm{rvec}[E_{Q(U)}[u_iu_i^\top]E_{Q(V)}[v_jv_j^\top]]}{\xi^2}\right)\\
&= \Xi\left(\sum_{(ij)}\frac{m_{ij}(\overline{u_i}\otimes\overline{v_j})-\mathrm{rvec}[(\Phi_i+\overline{u_iu_i}^\top)(\Psi_i+\overline{v_jv_j}^\top)]}{\xi^2}\right) \qquad (5)
\end{aligned}$$

Finally, we can also update the variances as:

$$\begin{aligned}
\omega^2 &= \frac{1}{n^2-1}\sum_{l=1}^{n^2}\Xi_{ll}+\overline{\lambda_l}^\top \qquad (6)\\
\xi^2 &= \frac{1}{K_p-1}\sum_{(ij)}m_{ij}^2+\mathrm{Tr}[(\Phi_i+\overline{u_iu_i}^\top)(\Psi_j+\overline{v_jv_j}^\top)]+\mathrm{Tr}[(\Xi+\overline{\lambda\lambda}^\top)((\Phi_i+\overline{u_iu_i}^\top)\otimes(\Psi_j+\overline{v_jv_j}^\top))]\\
&\qquad -2m_{ij}\overline{u_iv_j}^\top-2m_{ij}\overline{\lambda}^\top(\overline{u_i}\otimes\overline{v_j}))+2\overline{\lambda}^\top\,\mathrm{rvec}[(\Phi_i+\overline{u_iu_i}^\top)(\Psi_i+\overline{v_jv_j}^\top)] \qquad (7)
\end{aligned}$$

where K_p is the number of entries in the probe set.

1.2 Variational Linear Regression

The idea is to perform linear regression to learn the weights of a feature matrix. In addition, we address user-specific weights.

Assume $m_{ij} \sim \mathcal{N}(u_i v_j^T + w^T f_{ij} + u_i Z g_{ij}, \tau^2)$, where u_i is a $1 \times n$ vector, v_j is a $1 \times n$ vector, w is a $F \times 1$ vector, f_{ij} is a $F \times 1$ vector, Z is a $n \times F$ matrix and g_{ij} is a $F \times 1$ vector.

We also assume that
$u_{il} \sim \mathcal{N}(0, \sigma_l^2)$,
$v_{jl} \sim \mathcal{N}(0, \rho_l^2)$,
$w_m \sim \mathcal{N}(0, \theta^2)$,
$z_{lm} \sim \mathcal{N}(0, \xi_l^2)$

Furthermore, U, V, W and Z are assumed to be independent in the posterior $Q(U, V, W, Z) = Q(U)Q(V)Q(W)Q(Z)$. The variational free energy is given by

$$\begin{aligned}
\mathcal{F}(Q(U,V,W,Z)) &= \mathbb{E}_{Q(U,V,W,Z)}[\log\, p(M,U,V,W,Z) - \log\, Q(U,V,W,Z)] \\
&= -\frac{K}{2}\log(2\pi\tau) - \frac{I}{2}\sum_{l=1}^{n}\log(2\pi\sigma_l) - \frac{J}{2}\sum_{l=1}^{n}\log(2\pi\rho_l) \\
&\quad -\frac{F}{2}\log(2\pi\theta) - \frac{F}{2}\sum_{l=1}^{n}\log(2\pi\xi_l) - \frac{1}{2}\sum_{m=1}^{F}\frac{\mathbb{E}_{Q(W)}[w_m^2]}{\theta^2} \\
&\quad -\frac{1}{2}\sum_{l=1}^{n}\left[\frac{\sum_{i=1}^{I}\mathbb{E}_{Q(U)}[u_{il}^2]}{\sigma_l^2} + \frac{\sum_{j=1}^{J}\mathbb{E}_{Q(V)}[v_{jl}^2]}{\rho_l^2}\right] \\
&\quad -\frac{1}{2}\sum_{l=1}^{n}\left[\frac{\sum_{m=1}^{F}\mathbb{E}_{Q(Z)}[z_{lm}^2]}{\xi_l^2}\right] \\
&\quad -\frac{1}{2}\sum_{(ij)}\frac{\mathbb{E}_{Q(U,V,W,Z)}[(m_{ij} - u_i v_j^T - w^T f_{ij} - u_i Z g_{ij})^2]}{\tau^2} \\
&\quad -\log\, Q(U) - \log\, Q(V) - \log\, Q(W) - \log\, Q(Z)
\end{aligned}$$

$$\begin{aligned}
Q(U) \propto \exp(-\frac{1}{2}(\quad \textstyle\sum_{i=1}^{I}\sum_{l=1}^{n}\frac{u_{il}^2}{\sigma^2} + \sum_{j\in N(i)} \quad & \frac{u_i\mathbb{E}_{Q(V)}[v_j^T v_j]u_i^T - 2m_{ij}u_i\mathbb{E}_{Q(V)}[v_j^T]}{\tau^2} \\
&+ \frac{-2m_{ij}u_i\mathbb{E}_{Q(Z)}[Z]g_{ij} + 2u_i\mathbb{E}_{Q(V)}[v_j^T]\mathbb{E}_{Q(W)}[w^T]f_{ij}}{\tau^2} \\
&+ \frac{2u_i\mathbb{E}_{Q(V)}[v_j^T]u_i\mathbb{E}_{Q(Z)}[Z]g_{ij}}{\tau^2} \\
&+ \frac{2\mathbb{E}_{Q(W)}[w^T]f_{ij}u_i\mathbb{E}_{Q(Z)}[Z]g_{ij} + (u_i\mathbb{E}_{Q(Z)}[Z]g_{ij})^T(u_i\mathbb{E}_{Q(Z)}[Z]g_{ij})}{\tau^2}))
\end{aligned}$$

$$\Phi_i = \left(\begin{pmatrix} \frac{1}{\sigma_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_n} \end{pmatrix} + \textstyle\sum_{j\in N(i)} \frac{\Psi_j + \overline{v_j^T v_j} + \overline{v_j^T}g_{ij}^T\overline{Z^T} + \Xi + \overline{Z}g_{ij}g_{ij}^T\overline{Z^T}}{\tau^2}\right)^{-1}$$

$$\overline{u_i} = \Phi_i \left(\sum_{j \in N(i)} \frac{m_{ij}\overline{v_j} + m_{ij}(\overline{Z}g_{ij})^T - \overline{v_j}\,\overline{w^T} f_{ij} - \overline{w^T} f_{ij} \overline{Z} g_{ij}}{\tau^2} \right)$$

$$Q(V) \propto \exp(-\frac{1}{2}(\quad \sum_{j=1}^{J} \sum_{l=1}^{n} \frac{v_{jl}^2}{\rho^2} + \sum_{i \in N(j)} \quad \frac{v_j \mathbb{E}_{Q(U)}[u_i^T u_i] v_j^T - 2m_{ij} v_j^T \mathbb{E}_{Q(u)}[u_i^T]}{\tau^2} + \frac{2\mathbb{E}_{Q(U)}[u_i] v_j^T \mathbb{E}_{Q(W)}[w^T] f_{ij} + 2\mathbb{E}_{Q(U)}[u_i] v_j^T \mathbb{E}_{Q(Z)}[Z] g_{ij}}{\tau^2}))$$

$$\Psi_i = \left(\begin{pmatrix} \frac{1}{\rho_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\rho_n} \end{pmatrix} + \sum_{i \in N(j)} \frac{\Phi_i + \overline{u_i^T u_i}}{\tau^2} \right)^{-1}$$

$$\overline{v_j} = \Psi_j \left(\sum_{i \in N(j)} \frac{m_{ij}\overline{u_i} - \overline{u_i}\,\overline{w^T} f_{ij} - \overline{u_i}\,\overline{Z} g_{ij}}{\tau^2} \right)$$

$$Q(W) \propto \exp(-\frac{1}{2}(\sum_{m=1}^{F} \frac{w_m^2}{\theta^2} + \sum_{(ij)} \quad \frac{(w^T f_{ij})^T (w^T f_{ij}) + 2w^T f_{ij} \mathbb{E}_{Q(U)}[u_i] \mathbb{E}_{Q(Z)}[Z] g_{ij}}{\tau^2} - \frac{2m_{ij} w^T f_{ij} + 2\mathbb{E}_{Q(U)}[u_i] \mathbb{E}_{Q(V)}[v_j^T] w^T f_{ij}}{\tau^2}))$$

$$\Omega = \left(\frac{1}{\theta^2} + \sum_{(ij)} \frac{f_{ij}^T f_{ij}}{\tau^2} \right)^{-1}$$

$$\overline{w} = \Omega \left(\sum_{(ij)} \frac{m_{ij} f_{ij} - f_{ij} \overline{u_i}\,\overline{Z} g_{ij} - \overline{u_i}\,\overline{v_j^T} f_{ij}}{\tau^2} \right)$$

$$Q(Z) \propto \exp(-\frac{1}{2}(\sum_{m=1}^{F} \sum_{l=1}^{n} \frac{z_{lm}^2}{\xi_l^2} + \sum_{(ij)} \quad \frac{(\mathbb{E}_{Q(U)}[u_i] Z g_{ij})^T (\mathbb{E}_{Q(U)}[u_i] Z g_{ij}) - 2m_{ij} \mathbb{E}_{Q(U)}[u_i] Z g_{ij}}{\tau^2} + \frac{2\mathbb{E}_{Q(U)}[u_i] \mathbb{E}_{Q(V)}[v_j^T] \mathbb{E}_{Q(U)}[u_i^T] Z g_{ij} + 2\mathbb{E}_{Q(W)}[w^T] f_{ij} \mathbb{E}_{Q(U)}[u_i] Z g_{ij}}{\tau^2}))$$

$$\Xi_m = \left(\begin{pmatrix} \frac{1}{\xi_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\xi_n} \end{pmatrix} + \sum_{(ij)} \frac{(\Phi_i + \overline{u_i^T u_i})(g_{ij}^{(m)})^2}{\tau^2} \right)^{-1}$$

$$\overline{z_m} = \Xi_m \left(\sum_{(ij)} \frac{m_{ij} \overline{u_i} g_{ij}^{(m)} - (\Phi_i + \overline{u_i^T u_i v_j^T}\, g_{ij}^{(m)} - \overline{w^T} f_{ij} \overline{u_i} g_{ij}^{(m)}}{\tau^2} \right)$$

where $g_{ij}^{(m)}$ represents the m component of the vector g_{ij}