RECOMMENDATION ALGORITHM

19/05/2015

1

NMF + TV

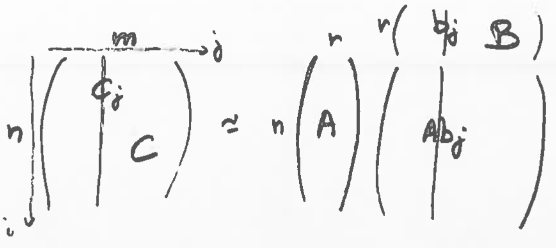
Observation Mask : $\Omega_{ij} = \begin{cases} 1 & \text{if } ij \text{ observed} \\ 0 & \text{otherwise} \end{cases}$

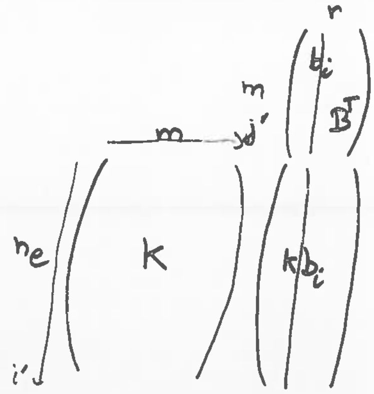
Hadamard product

- $$\min_{A,B \geq 0} \underbrace{\mathcal{D}\left(\Omega \circ (C \,\|\, AB) \right)}_{\substack{\text{Kullback-Leibler distance} \\ \text{between } C \text{ and } AB}} + \gamma \underbrace{\|B\|_{TV}}_{\| \underset{n_e \times m}{K} \underset{m \times r}{B^T} \|_1}$$

$$\Downarrow$$

$$\underset{n\times m}{C} \simeq \underset{n \times r}{A} \; \underset{r \times m}{B}$$





K = graph gradient operator
$n_e \times m$
n_e: # edges, m: # nodes

$$\min_{A,B \geq 0} \underbrace{- \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\Omega_{ij} C_{ij} \left(\log \frac{(AB)_{ij}}{C_{ij}} + 1 \right) + \Omega_{ij} (AB)_{ij} \right]}$$

$$+ \underbrace{\sum_{i=1}^{r} \| K b_i \|_1}$$

$$\sum_{i=1}^{r} \left| \sum_{i'=1}^{n_e} \sum_{j'=1}^{m} K_{i'j'} b_{ij'} \right|$$

$$\sum_{i=1}^{r} \sum_{i'=1}^{n_e} \left| \sum_{j'=1}^{m} K_{i'j'} b_{ij'} \right|$$

- Let fix A,

$$\min_{\substack{B \geq 0 \\ \Updownarrow \\ b_j \geq 0}} \sum_{j=1}^{m} \underbrace{\left[\sum_{i=1}^{n} - (\Omega_j)_i \cdot (C_j)_i \left(\log \frac{(Ab_j)_i}{(C_j)_i} + 1 \right) + (\Omega_j)_i \cdot (Ab_j)_i \right]}_{F(Ab_j)}$$

$$+ \sum_{i=1}^{r} \underbrace{\| K b_i \|_1}_{G(K b_i)}$$

$$\min_{b_j \geq 0} \quad \sum_{j=1}^{m} \underbrace{F(Ab_j)}_{\substack{= \\ \max\limits_{\underbrace{y_{1j}}_{n\times 1}} \langle \underbrace{\underset{n\times r}{A}\,\underset{r\times 1}{b_j}}_{n\times 1},\, y_{1j} \rangle - F^*(y_{1j})}} \quad + \quad \sum_{i=1}^{r} \underbrace{G(Kb_i)}$$

$$\max_{\underbrace{y_{2i}}_{n_e\times 1}} \quad \langle \underbrace{\underset{n_e\times m}{K}\,\underset{m\times 1}{b_i}}_{n_e\times 1},\, y_{2i} \rangle - G^*(y_{2i}$$

$$\min_{\underset{r\times 1}{b_j \geq 0}} \quad \max_{\underset{n\times 1}{y_{1j}},\ \underset{n_e\times 1}{y_{2i}}} \quad \sum_{j=1}^{m} \langle Ab_j, y_{1j} \rangle - F^*(y_{1j}) + \langle Kb_i, y_{2i} \rangle - G^*(y_{2i}$$

$$\Updownarrow$$

$$\min_{\underset{r\times m}{B \geq 0}} \quad \max_{\underset{n\times m}{Y_1},\ \underset{n_e\times r}{Y_2}} \quad \mathrm{Tr}\Big(\underbrace{\underbrace{(\underset{n\times r}{A}\,\underset{r\times m}{B})^T}_{m\times n} \cdot \underset{n\times m}{Y_1}}_{m\times m} \Big) - F^*(Y_1) + \mathrm{Tr}\Big(\underbrace{\underbrace{(\underset{n_e\times m}{K}\,\underset{m\times r}{B^T})^T}_{r\times n_e} \cdot \underset{n_e\times r}{Y_2}}_{r\times r} \Big) - G^*($$

$$- \frac{1}{2\sigma_1} \| Y - Y_1^n \|_F^2 - \frac{1}{2\sigma_2} \| Y - Y_2^n \|_F^2 + \frac{1}{2\tau} \| B - B^n$$

$$\Downarrow$$

(1) $$\max_{Y_1} \quad \mathrm{Tr}\big((AB)^T \cdot Y_1 \big) - F^*(Y_1) - \frac{1}{2\sigma_1} \| Y_1 - Y_1^n \|_F^2$$

$$\Updownarrow$$

$$\max_{y_{1j}} \quad \underbrace{\sum_{j=1}^{m} \langle Ab_j, y_{1j} \rangle - F^*(y_{1j}) - \frac{1}{2\sigma_1} \| y_{1j} - y_{1j}^n \|_2^2}$$

$$\underbrace{\max_{y_{1j}} \quad \sum_{j=1}^{m} \mathrm{prox}_{\sigma_1 F^*} \Big(\underset{n\times 1}{y_{1j}} + \sigma_1 \underbrace{\underset{n\times r}{A}\,\underset{r\times 1}{b_j}}_{n\times 1} \Big)}$$

$$\max_{Y_1} \quad \mathrm{prox}_{\sigma_1 F^*} \Big(\underset{n\times m}{Y_1} + \sigma_1 \underbrace{\underset{n\times r}{A}\,\underset{r\times m}{B}}_{n\times m} \Big)$$

(2)

$$\max_{Y_2} \quad Tr\left((KB^T)^T \cdot Y_2 \right) - G^*(Y_2) - \frac{1}{2\sigma_2} \| Y_2 - Y_2^n \|_F^2$$

$$\Updownarrow$$

$$\max_{y_{2i}} \quad \sum_{i=1}^{r} \langle Kb_i, y_{2i} \rangle - G^*(y_{2i}) - \frac{1}{2\sigma_2} \| y_{2i} - y_{2i}^n \|_2^2$$

$$\max_{y_{2i}} \quad \sum_{i=1}^{r} \quad prox_{\sigma G} \left(\underset{n_e \times 1}{y_{2i}} + \sigma_2 \underbrace{\underset{n_e \times m}{K} \underset{m \times 1}{b_i}}_{n_e \times 1} \right)$$

$$\max_{Y_2} \quad prox_{\sigma_2 G} \left(\underset{n_e \times r}{Y_2} + \sigma_2 \underbrace{\underset{n_e \times m}{K} \underset{m \times r}{B}}_{n_e \times r} \right)$$

(3)

$$\min_{B \geq 0} \quad \underbrace{Tr\left((AB)^T \cdot Y_1 \right)}_{\sum_{j=1}^{m} \underbrace{\langle Ab_j, y_{1j} \rangle}_{\langle b_j, A^T y_{1j} \rangle}} + \underbrace{Tr\left((KB^T)^T \cdot Y_2 \right)}_{\sum_{i=1}^{r} \underbrace{\langle Kb_i, y_{2i} \rangle}_{\langle b_i, K^T y_{2i} \rangle}} + \frac{1}{2\tau} \| B - B^n \|_F^2$$

$\Downarrow$ optimality condition

$$A^T y_{1j} \qquad\qquad K^T y_{2i}$$

$$\underbrace{\underset{r \times n}{A^T} \underset{n \times m}{Y_1}}_{r \times m} + \underbrace{\Big(\underbrace{\underset{m \times n_e}{K^T} \underset{n_e \times r}{Y}}_{m \times r} \Big)^T}_{r \times m} + \frac{1}{\tau} \Big(\underset{r \times m}{B - B^n} \Big) = 0$$

$$\Rightarrow \quad B^{n+1} = \Big(B^n - \tau A^T Y_1 - \tau (K^T Y_2)^T \Big)_+$$

- <u>Algorithm I</u> :

$$
\begin{cases}
Y_1^{n+1} = \mathrm{prox}_{\sigma_1 F^*}\left(Y_1^n + \sigma_1^n A \bar{B}^n \right) \\
Y_2^{n+1} = \mathrm{prox}_{\sigma_2 G^*}\left(Y_2^n + \sigma_2^n K \bar{B}^n \right) \\
B^{n+1} = \left(B^n - \tau A^T Y_1^{n+1} - \tau (K^T Y_2^{n+1})^T \right)_+ \\
\theta^{n+1} = \dfrac{1}{\sqrt{1+2\gamma \tau^n}} , \quad \tau^{n+1} = \theta^{n+1} \tau^n , \quad \sigma^{n+1} = \sigma^n / \theta^{n+1} \\
\bar{B}^{n+1} = B^{n+1} + \theta^{n+1} \left(B^{n+1} - B^n_{old} \right) \\
B^{n+1}_{old} = B^{n+1}
\end{cases}
$$

$$
\text{with} \quad \begin{cases}
\sigma_1^{n=0} \cdot \tau^{n=0} \leq \dfrac{1}{\|A\|^2} \\
\sigma_2^{n=0} \cdot \tau^{n=0} \leq \dfrac{1}{\|K\|^2}
\end{cases}
$$

- Computations of prox solutions :

(1) $\mathrm{prox}_{\sigma_1 F^*}(y) \overset{\text{Moreau's identity}}{=} y - \sigma_1 \underbrace{\mathrm{prox}_{F/\sigma_1}(y/\sigma_1)}$

$$\arg\min_z \left\{ \frac{\sigma_1}{2} \| z - y/\sigma_1 \|_2^2 + \underbrace{F(z)}_{-\Omega c\left(\log \frac{z}{c} + 1\right) + \Omega z} \right\}$$

$\Downarrow$ optimality

$$\sigma_1 z - y - \frac{\Omega c}{z} + \Omega = 0 \qquad \times z$$

$$\underbrace{\sigma_1}_{a} z^2 + \underbrace{(\Omega - y)}_{b} z \underbrace{- \Omega c}_{c} = 0$$

$\Downarrow$

$$z = \frac{y - \Omega + \sqrt{(y-\Omega)^2 + 4\sigma_1 \Omega c}}{2\sigma_1}$$

$$\mathrm{prox}_{\sigma_1 F^*}(y) = y - \cancel{\sigma_1}\left(\frac{y - \Omega + \sqrt{(y-\Omega)^2 + 4\sigma_1\Omega c}}{2\cancel{\sigma_1}} \right)$$

$$= \frac{1}{2}\left(y + \Omega - \sqrt{(y-\Omega)^2 + 4\sigma_1 \Omega c} \right)$$

$\Rightarrow \mathrm{prox}_{\sigma_1 F^*}(Y) =$

$$\frac{1}{2}\left(Y + \Omega - \sqrt{(Y-\Omega)\circ(Y-\Omega) + 4\sigma_1 \Omega \circ C} \right)$$

(2) $$\text{prox}_{\sigma_2 G^*}(y) = y - \sigma_2 \underbrace{\text{prox}_{G/\sigma_2}(y/\sigma_2)}$$

$$\underbrace{\underset{z}{\text{argmin}} \left\{ \frac{\cancel{\sigma_2}}{2} \| z - y/\sigma_2 \|_2^2 + \gamma/\sigma_2 \| z \|_1 \right\}}_{\text{shrink}_{\gamma/\sigma_2}(y/\sigma_2)}$$

$$\text{prox}_{\sigma_2 G^*}(y) = y - \sigma_2 \, \text{shrink}_{\gamma/\sigma_2}(y/\sigma_2)$$

$$\Rightarrow \text{prox}_{\sigma_2 G^*}(Y) = Y - \sigma_2 \, \text{shrink}_{\gamma/\sigma_2}(Y/\sigma_2)$$

<u>Algorithm II</u> :

$$
\left\{
\begin{aligned}
\underset{n\times m}{Y_1^{n+1}} &= \operatorname{prox}_{\sigma_1 F^*}\left(Y_1^n + \sigma_1^n \underset{n\times r}{A}\, \underset{r\times m}{\overline{B}^n} \right) \\
\underset{n_e\times r}{Y_2^{n+1}} &= \operatorname{prox}_{\sigma_2 G^*}\left(Y_2^n + \sigma_2^n \underset{n_e\times m}{K}\, \underset{m\times r}{\overline{B}^n} \right) \\
\underset{r\times m}{B^{n+1}} &= \left(B^n - \tau_1 \underset{r\times n}{A^T}\, \underset{n\times m}{Y_1^{n+1}} - \tau_2 \left(\underset{m\times n_e}{K^T}\, \underset{n_e\times r}{Y_2^{n+1}} \right)^T \right)_+ \\
\theta_1^{n+1} &= \frac{1}{\sqrt{1+2\gamma\tau_1^n}} , \quad \tau_1^{n+1} = \theta_1^{n+1}\tau_1^n , \quad \sigma_1^{n+1} = \sigma_1^n/\theta_1^{n+1} \\
\theta_2^{n+1} &= \frac{1}{\sqrt{1+2\gamma\tau_2^n}} , \quad \tau_2^{n+1} = \theta_2^{n+1}\tau_2^n , \quad \sigma_2^{n+1} = \sigma_2^n/\theta_2^{n+1} \\
\overline{B}^{n+1} &= B^{n+1} + \frac{\theta_1^{n+1}}{2}\left(B^{n+1} - B^n_{old} \right) + \frac{\theta_2^{n+1}}{2}\left(B^{n+1} - B^n_{old} \right) \\
B^{n+1}_{old} &= B^{n+1}
\end{aligned}
\right.
$$

with
$$\sigma_1^{n=0} = \tau_1^{n=0} = \frac{1}{\|A\|}$$
$$\sigma_2^{n=0} = \tau_2^{n=0} = \frac{1}{\|K\|}$$

- Back to the problem:

Let fix B,

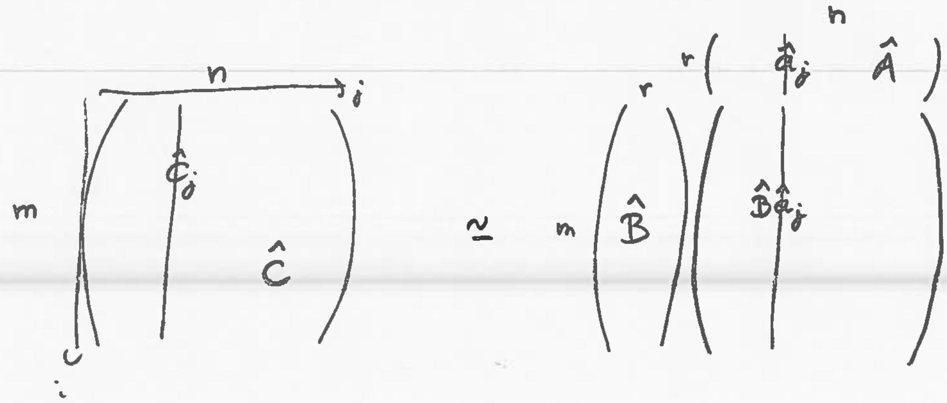
$$\min_{A \geq 0} \; D\left(\Omega \circ \left(C \,\|\, AB \right) \right)$$

$$\Updownarrow$$

$$\min_{A^T \geq 0} \; D\left(\Omega^T \circ \left(C^T \,\|\, \underbrace{(AB)^T}_{B^T A^T} \right) \right)$$

change of notation: $\begin{cases} \hat{\Omega} = \Omega^T \\ \hat{C} = C^T \\ \hat{A} = A^T \\ \hat{B} = B^T \end{cases}$

$$\min_{\hat{A} \geq 0} \; D\left(\underset{m \times n}{\hat{\Omega}} \circ \left(\underset{m \times n}{\hat{C}} \,\|\, \underset{m \times r}{\hat{B}} \; \underset{r \times n}{\hat{A}} \right) \right)$$



$$\hat{C} \simeq \hat{B}\hat{A}$$

$$\min_{\hat{A} \geq 0} \; \Leftrightarrow \; \min_{\hat{a}_j \geq 0} \; \sum_{j=1}^{n} \left[\underbrace{\sum_{i=1}^{m} -(\hat{\Omega}_j)_i \cdot (C_j)_i \left(\log \frac{(\hat{B}\hat{a}_j)_i}{(C_j)_i} + 1 \right)}_{F(\hat{B}\hat{a}_j)} + \underbrace{\sum_{i=1}^{m} (\hat{\Omega}_j)_i \cdot (\hat{B}\hat{a}_j)_i}_{G(\hat{a}_j)} \right.$$

$$\min_{\substack{\hat{a}_j \geq 0 \\ r\times 1}} \sum_{j=1}^{n} \max_{\substack{\hat{y}_j \\ m\times 1}} \langle \underbrace{\hat{B}_{m\times r} \hat{a}_{j\,r\times 1}}_{m\times 1}, \hat{y}_j \rangle - F^*(\hat{y}_j) + G(\hat{a}_j)$$

$$\Updownarrow$$

$$\min_{\substack{A \geq 0 \\ r\times n}} \max_{\substack{Y \\ m\times n}} \operatorname{Tr}\Big(\underbrace{\underbrace{(\hat{B}_{m\times r}\hat{A}_{r\times n})^T}_{n\times m} \cdot \underbrace{\hat{Y}}_{m\times n}}_{n\times n} \Big) - F^*(\hat{Y}) + G(\hat{A})$$

$$- \frac{1}{2\sigma} \| \hat{Y} - \hat{Y}^n \|_F^2 + \frac{1}{2\tau} \| \hat{A} - \hat{A}^n \|_F^2$$

$$\Downarrow$$

Algorithm III :

$$\begin{cases} \underset{m\times n}{\hat{Y}^{n+1}} = \operatorname{prox}_{\sigma F^*} \left(\hat{Y}^n + \sigma^n \underset{m\times r}{\hat{B}} \underset{r\times n}{\bar{\hat{A}}^n} \right) \\ \underset{r\times n}{\hat{A}^{n+1}} = \operatorname{prox}_{\tau G} \left(\hat{A}^n - \tau^n \underset{r\times m}{\hat{B}^T} \underset{m\times n}{\hat{Y}^{n+1}} \right) \\ \theta^{n+1} = \dfrac{1}{\sqrt{1+2\gamma\tau^n}} , \quad \tau^{n+1} = \theta^{n+1} \tau^n , \quad \sigma^{n+1} = \sigma^n / \theta^n \\ \bar{\hat{A}}^{n+1} = \hat{A}^{n+1} + \theta^{n+1} \left(\hat{A}^{n+1} - \hat{A}^n_{old} \right) \\ \hat{A}^{n+1}_{old} = \hat{A}^{n+1} \end{cases}$$

with $\sigma^{n=0} = \tau^{n=0} = \dfrac{1}{\|B\|}$

(1) $\text{prox}_{\sigma F^*}(y) = y - \sigma \underbrace{\text{prox}_{F/\sigma}(y/\sigma)}$

$$\underset{z}{\text{argmin}} \left\{ \frac{\sigma}{2} \| z - y/\sigma \|_2^2 + \underbrace{F(z)}_{-\hat{\Omega}\hat{c}\left(\log \frac{z}{\hat{c}} + 1\right)} \right\}$$

$\Downarrow$ optimality

$$\sigma z - y - \frac{\hat{\Omega}\hat{c}}{z} = 0 \qquad \times z$$

$$\underbrace{\sigma}_{a} z^2 \underbrace{- y}_{b} z \underbrace{- \hat{\Omega}\hat{c}}_{c} = 0$$

$\Downarrow$

$$z = \frac{y + \sqrt{y^2 + 4\sigma\hat{\Omega}\hat{c}}}{2\sigma}$$

$$\text{prox}_{\sigma F^*}(y) = y - \cancel{\sigma}\left(\frac{y + \sqrt{y^2 + 4\sigma\hat{\Omega}\hat{c}}}{2\cancel{\sigma}} \right)$$

$$= \frac{1}{2}\left(y - \sqrt{y^2 + 4\sigma\hat{\Omega}\hat{c}} \right)$$

$$\Rightarrow \text{prox}_{\sigma F^*}(Y) = \frac{1}{2}\left(Y - \sqrt{Y \circ Y + 4\sigma \hat{\Omega} \circ \hat{C}} \right)$$

(2) $$\mathrm{prox}_{\tau G}(\hat{a}) = \underset{z \geq 0}{\mathrm{argmin}} \left\{ \frac{1}{2\tau} \| z - \hat{a} \|_2^2 + \underbrace{G(z)}_{\hat{\Omega} \cdot (\hat{B} z)} \right\}$$

$\Downarrow$ optimality

$$\cancel{\frac{1}{\tau}} (z - \hat{a}) + \tau \hat{\Omega} \cdot (\hat{B} 1) = 0$$

$$= \left(\hat{a} - \tau \hat{\Omega} (\hat{B} 1) \right)_+$$

$$\Rightarrow \quad \mathrm{prox}_{\tau G}(\hat{A}) = \Big(\underset{r \times n}{\hat{A}} - \tau \underset{m \times n}{\hat{\Omega}} \circ (\underset{m \times r}{\hat{B}} \underset{r \times n}{1}) \Big)_+$$