Item Cold-Start Recommendations: Learning Local Collective Embeddings

Proof for Theorem 1

Martin Saveski, Amin Mantrach
Yahoo! Labs, Barcelona

The objective function J in Equation 2 is bounded from below by zero and the update rules for $\mathbf{H_s}$ and $\mathbf{H_u}$ are the same as in the original NMF formulation. Thus, we only need to prove that the objective J is non-increasing under the update step for $\mathbf{W}$, *i.e.*, Equation 9. We will follow a procedure based on auxiliary functions, similar to the one described in [1].

Definition. $G(w, w')$ is an auxiliary function for $J(w)$ if the conditions:

$$G(w, w') \geq J(w), \quad G(w, w) = J(w)$$

are satisfied. The auxiliary function is very useful because of the following lemma.

Lemma. If G is an auxiliary function of J, then J is non-increasing under the update:

$$w^{(t+1)} = \arg\min_{w} G(w, w^{(t)}) \tag{1}$$

Proof.

$$J(w^{(t+1)}) \leq G(w^{(t+1)}, w^{(t)}) \leq G(w^{(t)}, w^{(t)}) = J(w^{(t)}).$$

We rewrite the objective function of the Local Collective Embeddings (LCE) model (Equation 2) as follows:

$$\begin{aligned} \min : J = \ & \frac{1}{2}(\alpha \sum_{i=1}^{N}\sum_{j=1}^{M}(x^s_{ij} - \sum_{k=1}^{K} w_{ik}h^s_{kj})^2 + (1-\alpha)\sum_{i=1}^{N}\sum_{j=1}^{F}(x^u_{ij} - \sum_{k=1}^{K} w_{ik}h^u_{kj})^2 + \\ & +\beta \sum_{k=1}^{K}\sum_{j=1}^{N}\sum_{l=1}^{N} w_{jk}[\mathbf{L}]_{jl}w_{lk} + \lambda_W \sum_{i=1}^{N}\sum_{j=1}^{K} w_{ij}^2 + \lambda_{H_s}\sum_{i=1}^{K}\sum_{j=1}^{M}(h^s_{ij})^2 + \lambda_{H_u}\sum_{i=1}^{K}\sum_{j=1}^{F}(h^u_{ij})^2. \end{aligned}$$

Considering any element w_{ab} of $\mathbf{W}$, we use J_{ab} to denote the part of J which is only relevant to w_{ab}. It is easy to check that:

$$J'_{ab} = \left[\frac{\partial J}{\partial \mathbf{W}}\right]_{ab} = [\alpha \mathbf{W}\mathbf{H_s}\mathbf{H_s}^{\mathrm{T}} - \alpha \mathbf{X_s}\mathbf{H_s}^{\mathrm{T}} + (1-\alpha)\mathbf{W}\mathbf{H_u}\mathbf{H_u}^{\mathrm{T}} - (1-\alpha)\mathbf{X_u}\mathbf{H_u}^{\mathrm{T}} + \beta \mathbf{L}\mathbf{W} + \lambda_W \mathbf{W}]_{ab}.$$

$$J''_{ab} = \alpha[\mathbf{H_s}\mathbf{H_s}^{\mathrm{T}}]_{bb} + (1-\alpha)[\mathbf{H_u}\mathbf{H_u}^{\mathrm{T}}]_{bb} + \beta[\mathbf{L}]_{aa} + \lambda_W$$

Since our update is essentially element-wise, it is sufficient to show that each J_{ab} is non-increasing under the update step of Equation 9.

We define:

$$G(w, w_{ab}^{(t)}) = J_{ab}(w_{ab}^{(t)}) + J'_{ab}(w_{ab}^{(t)})(w - w_{ab}^{(t)}) +$$
$$+ \frac{\alpha[\mathbf{W}\mathbf{H_s}\mathbf{H_s}^{\mathrm{T}}]_{ab} + (1-\alpha)[\mathbf{W}\mathbf{H_u}\mathbf{H_u}^{\mathrm{T}}]_{ab} + \beta[\mathbf{D}\mathbf{W}]_{ab} + \lambda_W[\mathbf{W}]_{ab}}{w_{ab}^{(t)}}(w - w_{ab}^{(t)})^2$$

is an auxiliary function for J_{ab}, the part of J which is only relevant to w_{ab}.

Since $G(w, w) = J_{ab}(w)$ is obvious, we need only show that $G(w, w_{ab}^{(t)}) \geq J_{ab}(w)$. To do this, we compare the Tylor series expansion of $J_{ab}(w)$:

$$J_{ab}(w) = J_{ab}(w_{ab}^{(t)}) + J'_{ab}(w_{ab}^{(t)})(w - w_{ab}^{(t)}) + (\alpha[\mathbf{H_s}\mathbf{H_s}^{\mathrm{T}}]_{bb} + (1-\alpha)[\mathbf{H_u}\mathbf{H_u}^{\mathrm{T}}]_{bb} + \beta[\mathbf{L}]_{aa} + \lambda_W)(w - w_{ab}^{(t)})^2,$$

with $G(w, w_{ab}^{(t)})$ to find that $G(w, w_{ab}^{(t)}) \geq J_{ab}(w)$ is equivalent to:

$$\frac{\alpha[\mathbf{W}\mathbf{H_s}\mathbf{H_s}^{\mathrm{T}}]_{ab} + (1-\alpha)[\mathbf{W}\mathbf{H_u}\mathbf{H_u}^{\mathrm{T}}]_{ab} + \beta[\mathbf{D}\mathbf{W}]_{ab} + \lambda_W[\mathbf{W}]_{ab}}{w_{ab}^{(t)}} \geq$$
$$\geq \alpha[\mathbf{H_s}\mathbf{H_s}^{\mathrm{T}}]_{bb} + (1-\alpha)[\mathbf{H_u}\mathbf{H_u}^{\mathrm{T}}]_{bb} + \beta[\mathbf{L}]_{aa} + \lambda_W.$$

We have:

$$\alpha[\mathbf{W}\mathbf{H_s}\mathbf{H_s}^{\mathrm{T}}]_{ab} = \alpha \sum_{l=1}^{k} w_{al}^{(t)}[\mathbf{H_s}\mathbf{H_s}^{\mathrm{T}}]_{lb} \geq \alpha w_{ab}^{(t)}[\mathbf{H_s}\mathbf{H_s}^{\mathrm{T}}]_{bb},$$
$$(1-\alpha)[\mathbf{W}\mathbf{H_u}\mathbf{H_u}^{\mathrm{T}}]_{ab} = (1-\alpha) \sum_{l=1}^{k} w_{al}^{(t)}[\mathbf{H_u}\mathbf{H_u}^{\mathrm{T}}]_{lb} \geq (1-\alpha) w_{ab}^{(t)}[\mathbf{H_u}\mathbf{H_u}^{\mathrm{T}}]_{bb},$$
$$\beta[\mathbf{D}\mathbf{W}]_{ab} = \beta \sum_{j=1}^{N} [\mathbf{D}]_{aj} w_{jb}^{(t)} \geq \beta[\mathbf{D}]_{aa} w_{ab}^{(t)} \geq \beta[\mathbf{D} - \mathbf{A}]_{aa} w_{ab}^{(t)} = \beta[\mathbf{L}]_{aa}.$$

Thus, $G(w, w_{ab}^{(t)}) \geq J_{ab}(w)$ holds.

Replacing $G(w, w_{ab}^{(t)})$ in Equation 1 results in the update rule:

$$\begin{aligned} w_{ab}^{(t+1)} &= w_{ab}^{(t)} - w_{ab}^{(t)} \frac{J'_{ab}(w_{ab})}{\alpha[\mathbf{W}\mathbf{H_s}\mathbf{H_s}^{\mathrm{T}}]_{ab} + (1-\alpha)[\mathbf{W}\mathbf{H_u}\mathbf{H_u}^{\mathrm{T}}]_{ab} + \beta[\mathbf{D}\mathbf{W}]_{ab} + \lambda_W[\mathbf{W}]_{ab}} \\ &= w_{ab}^{(t)} \frac{[\alpha\mathbf{X_s}\mathbf{H_s}^{\mathrm{T}} + (1-\alpha)\mathbf{X_u}\mathbf{H_u}^{\mathrm{T}} + \beta\mathbf{A}\mathbf{W}]_{ab}}{[\alpha\mathbf{W}\mathbf{H_s}\mathbf{H_s}^{\mathrm{T}} + (1-\alpha)\mathbf{W}\mathbf{H_u}\mathbf{H_u}^{\mathrm{T}} + \beta\mathbf{D}\mathbf{W} + \lambda_W\mathbf{W}]_{ab}}. \end{aligned}$$

Since G is an auxiliary function, J_{ab} is non-increasing under this update rule. $\square$

References

[1] D. Cai, X. He, X. Wu, and J. Han. Non-negative matrix factorization on manifold. In *International Conference on Data Mining*, 2008.