Supplementary Material for Comment-based Multi-View Clustering of Web 2.0 Items

Xiangnan He[1] Min-Yen Kan[1] Peichu Xie[2] Xiao Chen[3]
[1]School of Computing, National University of Singapore
[2]Department of Mathematics, National University of Singapore
[3]Institute of Computing Technology, Chinese Academy of Sciences
{xiangnan, kanmy}@comp.nus.edu.sg xie@nus.edu.sg chenxiao3310@ict.ac.cn

This supplemental material provides convergence proof of the CoNMF algorithm proposed by the paper [1]. Specifically, we prove that in paper [1], the iterative solution Eq. (10) provides a local-minimum solution of the pair-wise CoNMF (Eq. (5)). The convergence of the cluster-wise CoNMF solution can be proved in a similar way.

1. CONVERGENCE PROOF

Recall that the objective function of the pair-wise CoNMF is written as:

$$J = \sum_{s=1}^{n_v} \lambda_s ||V^{(s)} - W^{(s)}H^{(s)}|| + \sum_{s,t} \lambda_{st} ||W^{(s)} - W^{(t)}||, \quad s.t. \quad W^{(s)} \geq 0, H^{(s)} \geq 0. \tag{1}$$

where $\|\cdot\|$ denotes the squared Frobenius norm. The update rules are written as:

$$H^{(s)} \leftarrow H^{(s)} \odot \frac{W^{(s)T}V^{(s)}}{W^{(s)T}W^{(s)}H^{(s)}}, \quad W^{(s)} \leftarrow W^{(s)} \odot \frac{\lambda_s V^{(s)}H^{(s)T} + \sum_{t=1}^{n_v}\lambda_{st}W^{(t)}}{\lambda_s W^{(s)}H^{(s)}H^{(s)T} + \sum_{t=1}^{n_v}\lambda_{st}W^{(s)}}. \tag{2}$$

In the followings, we present the proof of the non-increasing property of update rules Eq. (2) using the auxiliary function method [2]. Note that there are two parts of CoNMF objective function, the NMF part (i.e., the combination of NMF in individual matrics), and the co-regularization part. It is clear that the NMF part has been already proved by [2], in this material, we focus on the co-regularization part.

To construct the auxiliary function[1], we first calculate the gradient of the objective function J. Taking gradient with respective to w, we have:

$$\nabla_w J = 2\lambda_s\left(-vH^T + wHH^T\right) + 2\sum_{t=1}^{n}\lambda_{st}\left(w - w^t\right), \tag{3}$$

where $w = W^{(s)}_{\alpha\cdot}$, $w^t = W^{(t)}_{\alpha\cdot}$, $v = V^{(s)}_{\alpha\cdot}$ for the α-th line of $W^{(s)}$, $W^{(t)}$, $V^{(s)}$ taken separately as row vectors.

Given certain iteration of our data W and H, we denote the current value of w as $\bar{w}$. The auxiliary function with respect to the row vector w is given as:

$$G_{\bar{w}}(w) = J(\bar{w}) + \nabla_w J|_{\bar{w}} \cdot (w-\bar{w})^T + \frac{1}{2}(w-\bar{w})\cdot K_{\bar{w}} \cdot (w-\bar{w}), \tag{4}$$

[1]For the conditions to be satisfied by auxiliary function, please refer to [2].

where

$$K_{\bar{w}} = diag\left\{2\bar{w}_\beta^{-1}\left(\lambda_s\bar{w}HH^T + \sum_{t=1}^{n}\lambda_{st}\bar{w}\right)_\beta\right\}_{\beta=1}^{K} = 2\ diag\left\{\bar{w}_\beta^{-1}\left(\lambda_s\bar{w}HH^T\right)_\beta\right\}_{\beta=1}^{K} + 2\sum_{t=1}^{n}\lambda_{st}\cdot I. \tag{5}$$

Due to the quadraticity of J with respect to w, we have:

$$J(w) = J(\bar{w}) + \nabla_w J|_{\bar{w}} \cdot (w-\bar{w})^T + \frac{1}{2}(w-\bar{w})\cdot\nabla_w^2 J|_{\bar{w}}\cdot(w-\bar{w})^T, \tag{6}$$

in which the Hessian can be evaluated:

$$\nabla_w^2 J|_{\bar{w}} = 2\lambda_s HH^T + 2\sum_{t=1}^{n}\lambda_{st}\cdot I. \tag{7}$$

Therefore, we have:

$$K_{\bar{w}} - \nabla_w^2 J|_{\bar{w}} = 2diag\left\{\bar{w}_\beta^{-1}\left(\lambda_s\bar{w}HH^T\right)_\beta\right\}_{\beta=1}^{K} - 2\lambda_s HH^T. \tag{8}$$

Note that the entry-wise positiveness of matrices is enforced. Therefore we can obtain that:

$$G_{\bar{w}}(w) \geq J(w), \tag{9}$$

which is necessary and sufficient for $G.(\cdot)$ to be an auxilliary function, regarding the fact that $G_{\bar{w}}(\bar{w}) = J(\bar{w})$. Based on the above argument, the renewal taken at the arg-minimum of $G_{\bar{w}}$ each time will be non-increasing, more concretely:

$$\arg\min_w G_{\bar{w}} = \bar{w} - K_{\bar{w}}^{-1}\left(\nabla_w J|_{\bar{w}}\right) = \bar{w}\odot\frac{\lambda_s vH^T + \sum_{t=1}^{n}\lambda_{st}w^t}{\lambda_s\bar{w}HH^T + \sum_{t=1}^{n}\lambda_{st}\bar{w}}. \tag{10}$$

As such, our algorithm is exactly reproduced in its row-wise form. The convergence of the algorithm is thus proved.

2. REFERENCES

[1] X. He, M.-Y. Kan, P. Xie, and X. Chen. Comment-based multi-view clustering of web 2.0 items. In *Proc. of WWW '14*, 2014.

[2] D. Seung and L. Lee. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, 13:556–562, 2001.