Proof of convergence of SMV-NMF

anonymous

April 2019

1 First part of convergence

The aim of this part is to find anauxiliary function for SMV-NMF objective function as expressed in Eq. (8).

Definition 1. $G(h, h^{'})$ is an auxiliary function for our final function $J(h)$ if the following conditions are satisfied:

$$G\left(h', h\right) \geq J(h) \quad \text{and} \quad G(h, h) = J(h) \tag{1}$$

The auxiliary function is useful because of the following lemma, and the proof of Lemma 1 is given by [2].

Lemma 1. If G is an auxiliary function, then J is non-increasing under the update:

$$h^{t+1} = \arg\min_{h} G\left(h, h^{t}\right) \tag{2}$$

consequently, we have:

$$J\left(h^{t+1}\right) \leq G\left(h^{t+1}, h^{t}\right) \leq G\left(h^{t}, h^{t}\right) = J\left(h^{t}\right) \tag{3}$$

Lemma 2. If $K(h^t)$ is a diagonal matrix under the following definition,

$$K(h^t) = diag(Wdiag(v)W^T h./h) \tag{4}$$

where v is a column vector of $V = Y_p + \overline{Y}_p \odot \left(\lambda_1 Z_p + \lambda_2 Z_p' + \lambda_3 \mathbf{1}\right)$ then,

$$G\left(h, h^{t}\right) = J\left(h^{t}\right) + \left(h - h^{t}\right)^{T} \nabla J\left(h^{t}\right) + \frac{1}{2}\left(h - h^{t}\right)^{T} K\left(h^{t}\right)\left(h - h^{t}\right) \tag{5}$$

is an auxiliary function for $J(h)$.

Proof: Since $G(h, h) = J(h)$ is obvious, we need only show that $G(h, h^t) \geq J(h)$. To do this, we compare

$$J(h) = J\left(h^{t}\right) + \left(h - h^{t}\right)^{T} \nabla J\left(h^{t}\right) + \frac{1}{2}\left(h - h^{t}\right)^{T}\left(Wdiag(v)W^{T}\right)\left(h - h^{t}\right) \tag{6}$$

with Eq. (5) to find that $G(h, h^t) \geq J(h)$ is equivalent to

$$0 \leq \left(h-h^{t}\right)^{T}\left[K\left(h^{t}\right)-W diag(v) W^{T}\right]\left(h-h^{t}\right) \tag{7}$$

The next step is to prove $\left[K\left(h^{t}\right)-W \operatorname{diag}(v) W^{T}\right]$ is positive semi-definite. Let $Q = W diag(v) W^{T}$, then $\left[K\left(h^{t}\right)-W \operatorname{diag}(v) W^{T}\right]$ can be expressed as $[diag(Qh./h) - Q]$. As the Lemma 1 provided in [1], if Q is a symmetric non-negative matrix and h be a positive vector, then the matrix $\hat{Q} = diag(Qh./h) - Q \succeq 0$.

2 Second part of convergence

We can now demonstrate the convergence of method.

Proof: Replacing $G(h, h^t)$ in Eq. (2) by Eq. (5) results in the update rule:

$$h^{t+1}=h^{t}-K\left(h^{t}\right)^{-1} \nabla J\left(h^{t}\right) \tag{8}$$

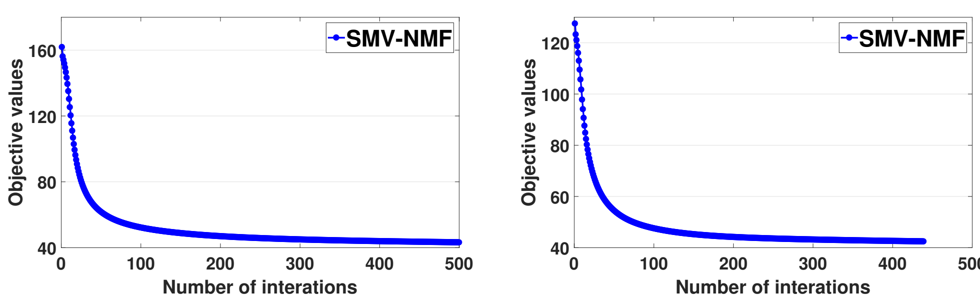
Since Eq. (5) is an auxiliary function, J is nonincreasing under this update rule, according to Lemma 1. Writing the components of this equation explicitly, we obtain

$$h_{a}^{t+1}=h_{a}^{t} \frac{(W x)_{a}}{\left(W\left(v \odot W^{T} h\right)\right)_{a}} \tag{9}$$

where x is the column vector of $X=\left(Y \odot X+\overline{Y} \odot\left(\lambda_{1} Z \odot X^{mv}+\lambda_{2} Z^{\prime} \odot X^{sv}+\lambda_{3} X^{knn}\right)\right)$.

By reversing the roles of W and H in Lemma 1 and 2, J can similarly be shown to be nonincreasing under the update rules for W.

Figures 1 (a) and (b) show the convergence trends of iterative model SMV-NMF on both largest and smallest datasets. It illustrates that our algorithm can converge into a local solution in terms of the objective value in a small amount of iterations.



(a) Conduct on the largest dataset (Melbourne) (b) Conduct on the smallest dataset (Perth)

Figure 1: Converge rate

References

[1] Mithun Das Gupta and Jing Xiao. Non-negative matrix factorization as a feature selection tool for maximum margin classifiers. In *CVPR 2011*, pages 2841–2848. IEEE, 2011.

[2] Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.