Insightful Additions for BSMF

Chaoqi Yang
chaoqiy2@illinois.edu

October 26, 2020

1 Property of Orthogonality

The basic idea of BSMF is to decompose $X^{MS} \approx UBM^\top$. The property of row set of U, $\{U_k\}$, is discussed below.

Theorem 1. *After BSMF factorization, $\{U_k\}$ are approximately orthogonal.*

Proof. We start the analysis by the recapitulation of loss function,

$$\begin{aligned} J &= \|X^{MS} - UBM^\top\|_F^2 + \lambda_1\|U\|_F^2 + \lambda_1\|M\|_F^2 + \lambda_2\|U\|_1 + \lambda_2\|M\|_1. \\ &= tr({X^{MS}}^\top X^{MS}) - 2tr({X^{MS}}^\top UBM^\top) + tr(UBM^\top MB^\top U^\top) \\ &+ \lambda_1 tr(U^\top U) + \lambda_1 tr(M^\top M) + \lambda_2\|U\|_1 + \lambda_2\|M\|_1. \end{aligned} \tag{1}$$

Let us first ignore the L_1 and L_2 term. Since $tr({X^{MS}}^\top X^{MS})$ is a constant, the problem of minimizing Equation (1) is equivalent to,

$$\max_{U,M\geq 0} tr({X^{MS}}^\top UBM^\top), \tag{2}$$

$$\text{and} \quad \min_{U,M\geq 0} tr(UBM^\top MB^\top U^\top). \tag{3}$$

The first objective, Equation (2), is similar to K-means type clustering after expanding the trace function, which maximize within-cluster similarities,

$$\max \sum_{i,j} u_i^\top B m_j x_{ij}. \tag{4}$$

The second objective is to enforce orthogonality approximately. Because $UBM^\top \approx X^{MS}$, let us add L_2-norm of U and M now, so that the scale of U, M are constrained [1]. Since B is positive and fixed, $tr(U^\top U) = \|U\|_F^2$ and $tr(BM^\top MB^\top) = \|MB^\top\|_F^2 \leq \|M\|_F^2\|B\|_F^2$ are approximately constants. So Equation (3) becomes,

$$\min_{U,M\geq 0} tr\left((U^\top U - I)(BM^\top MB^\top - I)\right). \tag{5}$$

By Cauchy's inequality, we further have,

$$tr\left((U^\top U - I)(BM^\top MB^\top - I)\right) \tag{6}$$

$$= \sum_{i=1,\dots,K;\ j=1,\dots,K} (U^\top U - I)_{ij}(BM^\top MB^\top - I)_{ji} \tag{7}$$

$$\leq \|U^\top U - I\|_F \cdot \|BM^\top MB^\top - I\|_F. \tag{8}$$

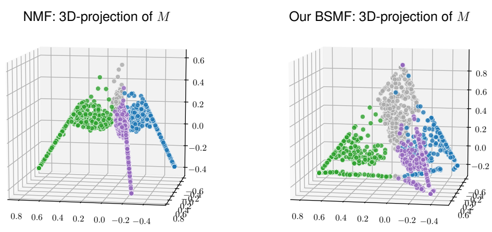

Figure 1: Visualization of M for NMF (Left) and Our BSMF (Right)

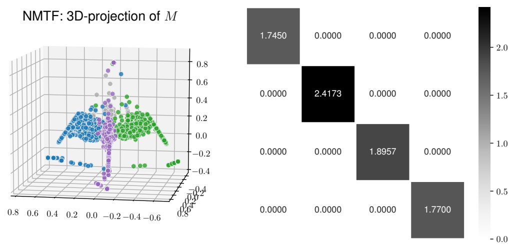

Figure 2: Visualization of M (Left) and $\tilde{B}$ (Right) for NMTF

Thus, the overall second objective is bounded by,

$$\min_{U,M\geq 0} \|U^\top U - I\|_F \cdot \|BM^\top MB^\top - I\|_F \tag{9}$$

$$\Rightarrow \min_{U\geq 0}\|U^\top U - I\|_F \cdot \min_{M\geq 0}\|BM^\top MB^\top - I\|_F. \tag{10}$$

which enforces the orthogonality of pre-belief bases $\{U_k\}$. □

2 Visualization of Synthetic Results

To augment the synthetic data experiments, we add the visualization results in the material to compare methods in more depth. Specifically, after running our BSMF and other methods, the estimated M is projected into a 3-D space in Figure 1 and Figure 2, where each data point represents a message. In each figure, all of the data points seem to lie in a regular tetrahedron (should be regular K-polyhedron for more general K-belief cases). It is interesting that for NMF, most of the data points cluster around the upper corner. It is obviously difficult to draw a boundary for the crowded mass. NMTF performs a little bit better: data points with different beliefs stretch apart, making their beliefs more separable. We also plot the estimated $\tilde{B} \in \mathbb{R}^{4\times 4}$ for NMTF, and it turns out to be an SVD-like diagonal matrix, which means that pure NMTF only learns the independent variances aligned with each belief.

The projection result of BSMF is surprising: data points are evenly located and grouped by colors. They approximately form a regular tetrahedron. We hypothesize that in the 4-dim space, data points should be perfectly aligned with one of the belief bases/parts, and these four bases are conceivably orthogonal in that space. This result suggests that our BSMF disentangles the latent manifold and leads to a better separation of messages by belief sets.

References

[1] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 6, pp. 1336–1353, 2012.