2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2010)
August 29 – September 1, 2010, Kittilä, Finland

MANIFOLD-RESPECTING PROBABILISTIC MATRIX TRI-FACTORIZATION

Shounan An [1], *Jiho Yoo* [1], *and Seungjin Choi* [1,2]

[1] Department of Computer Science
[2] Division of IT Convergence Engineering
Pohang University of Science and Technology, Korea
{anshounan,zentasis,seungjin}@postech.ac.kr

ABSTRACT

Probabilistic latent semantic analysis (PLSA) is a popular topic model for factor analysis of dyadic data, which is closely related to nonnegative matrix factorization (NMF) that seeks a 2-factor decomposition of a nonnegative data matrix. We previously proposed probabilistic matrix tri-factorization (PMTF) which is a probabilistic model for a 3-factor decomposition of a nonnegative data matrix, extending PLSA and NMF for co-clustering simultaneously columns and rows of dyadic data matrix. However, these methods do not take the local manifold structure of dyadic data into account. In this paper we present a method for manifold-respecting probabilistic matrix tri-factorization (MPMTF) where we incorporate a local manifold structure into PMTF, imposing smoothness constraints on posterior distributions over latent variables. We develop an EM algorithm to learn MPMTF. Our model handles both unlabeled and labeled data points, while existing methods considered unlabeled data only. Numerical experiments on document and image datasets confirm the useful behavior of our proposed method in the task of clustering.

Index Terms— Manifold regularization, nonnegative matrix factorization, probabilistic latent semantic analysis

1. INTRODUCTION

Nonnegative matrix factorization (NMF) is a method for multivariate analysis of nonnegative data, seeking a 2-factor decomposition of a nonnegative data matrix where two factor matrices are restricted to be nonnegative [9]. One of prominent applications of NMF, which is of our interest, is document clustering that plays an important role in dyadic data analysis [11]. Dyadic data refers to a domain with two finite sets of objects in which observations are made for pairs with one element from either set [8]. A term-document matrix is an example of dyadic data, consisting of co-occurrence frequencies of word-document pairs. When NMF is applied to a term-document matrix, the matrix is decomposed into a product of two factor matrices, where one corresponds to cluster centers and the other is associated with cluster indicator variables [11]. Orthogonality constraints were imposed on factor matrices in the decomposition [5, 4], where a clear link between k-means clustering and NMF is made in such a case. Multiplicative updates for orthogonal NMF were developed, directly exploiting gradient information on Stiefel manifolds [4, 12].

NMF is closely related to probabilistic latent semantic analysis (PLSA) [7] which is a widely-used topic model. PLSA defines a proper generative model, known as *the aspect model*, for factor analysis of dyadic data, viewing a document as a distribution over topics where each topic is described by a different distribution over words. An interesting link between PLSA and NMF was revealed in [6], where the multiplicative update algorithm for NMF with KL-divergence was shown to be equivalent to the EM algorithm for PLSA.

Nonnegative matrix tri-factorization (NMTF) is a 3-factor decomposition of a nonnegative data matrix, which has been successfully applied to *co-clustering*, the goal of which is to simultaneously group columns and rows of dyadic data matrix [5]. Co-clustering was shown to improve the accuracy of document clustering [5]. Previously we proposed probabilistic matrix tri-factorization (PMTF) [13] which is a probabilistic model for a 3-factor decomposition of a nonnegative data matrix, extending PLSA and NMF for simultaneously co-clustering columns and rows of a dyadic data matrix.

Most of existing methods neglected the low-dimensional manifold structure of dyadic data. Recently manifold regularization [1] was incorporated into NMF and PLSA [3], assuming that the posterior distributions over latent variables vary smoothly along geodesics in the underlying geometry. In this paper we present *manifold-respecting PMTF* (MPMTF) where we consider low-dimensional manifold structure of data as well as of attributes in the framework of PMTF, such that latent structure of dyadic data is discovered by matrix tri-factorization and local and global consistency is satisfied. In addition, we also demonstrate that our MPMTF can easily incorporate labeled data, which was not used in most of previous work. MPMTF inherits the useful behavior of PMTF such as probabilistic co-clustering and automatic detection of the number of clusters. On the top of that, MPMTF improves the clustering accuracy by taking the low-dimensional manifold structure of dyadic data into account. Numerical experiments on document and image datasets confirm the useful behavior of our proposed method in the task of clustering.

2. PLSA AND PMTF

We are given a term-document matrix $\boldsymbol{X} \in \mathbb{R}_{+}^{M\times N}$, where observations X_{ij} are co-occurrence frequencies of dyads (w_i, d_j) (i.e., the significance of term (word) w_i in document d_j) for two sets of objects, $\mathcal{W} = \{w_1, \ldots, w_M\}$ and $\mathcal{D} = \{d_1, \ldots, d_N\}$.

PLSA makes use of a statistical latent class model for factor analysis of dyadic data [7]. Given the latent class variable $z \in \{1, \ldots, K\}$, the likelihood is given by

$$p(\mathcal{X}) = \prod_i \prod_j \left(\sum_k p(w_i|z_k)p(d_j|z_k)p(z_k) \right)^{C_{ij}}, \tag{1}$$

where C_{ij} are empirical counts for dyads (w_i, d_j) in $\mathcal{X} = \{(w_i, d_j)\}$.

In contrast, PMTF introduces two dependent latent variables y_l and z_k, each of which corresponds to cluster indicator variable for words and documents, respectively. The term-document joint distribution is factorized as

$$\begin{aligned} p(w_i, d_j) &= \sum_{l=1}^{R}\sum_{k=1}^{K} p(w_i, d_j|y_l, z_k)p(y_l, z_k) \\ &= \sum_{l=1}^{R}\sum_{k=1}^{K} p(w_i|y_l)p(d_j|z_k)p(y_l, z_k), \end{aligned} \tag{2}$$

where R is the number of row clusters, K is the number of column clusters, and $p(y_l, z_k)$ is the joint prior probability for term cluster y_l and document cluster z_k. Given these two latent variables, the likelihood is given by

$$p(\mathcal{X}) = \prod_i \prod_j \left(\sum_l \sum_k p(w_i|y_l)p(d_j|z_k)p(y_l, z_k) \right)^{C_{ij}}, \tag{3}$$

where C_{ij} are empirical counts for dyads (w_i, d_j) in the training set $\mathcal{X} = \{(w_i, d_j)\}$.

3. MANIFOLD-RESPECTING PMTF

The EM algorithm for PMTF was derived in [13], where the following expected complete-data log-likelihood $\mathcal{L}_c$ was considered:

$$\mathcal{L}_c = \sum_{l,k,i,j} C_{ij} p(y_l, z_k|w_i, d_j) \log[p(w_i|y_l)p(d_j|z_k)p(y_l, z_k)],$$

where the posterior distributions over latent variables are computed by

$$p(y_l, z_k|w_i, d_j) = \frac{p(w_i|y_l)p(d_j|z_k)p(y_l, z_k)}{\sum_l \sum_k p(w_i|y_l)p(d_j|z_k)p(y_l, z_k)}.$$

In the task of document clustering, posterior distributions $p(z_k|d_j)$ indicate cluster membership information. A cluster assumption states that nearby documents are likely to have the same cluster membership (local consistency) and documents on the same structure (cluster or sub-manifold) are likely to have the same cluster membership (global consistency). In PMTF, cluster membership for documents and words is contained in $p(z_k|d_j)$ and $p(y_l|w_j)$, respectively. These posterior probabilities can be computed by using $p(d_j|z_k)$ and $p(w_j|y_l)$ with appropriately marginalized $p(y_l, z_k)$. Thus, we assume that these posterior distributions vary smoothly along geodesics in the underlying geometry. To this end, we add a dual manifold regularizer to the PMTF model, in order to satisfy local and global consistency for low-dimensional manifolds of documents and words. Then the objective function to be maximized is given by

$$\mathcal{J} = \mathcal{L}_c - \lambda\Psi - \mu\Omega, \tag{4}$$

where λ and μ are the manifold regularization parameters , Ψ and Ω are manifold regularizers for documents and words, which are described in Section 3.1.

3.1. Dual Regularization

We first consider the underlying geometry of the document manifold. The same technique is applied to the word manifold. As in [1], we approximate the geometry of the manifold by a neighborhood graph of data points (sampled from the underlying manifold). In order to construct a neighborhood graph, we define the edge weight matrix by

$$W_{ij}^d = \begin{cases} 1 & \text{if } d_i \in \mathcal{N}(d_j) \text{ or } d_j \in \mathcal{N}(d_i) \\ 0 & \text{otherwise} \end{cases}$$

where $\mathcal{N}(d_j)$ denotes k-nearest neighbors of d_j, usually computed from the Euclidean distance between the documents. The edge weight matrix for words W_{ij}^w can be computed in a similar way, using the distance between the words.

For the local and global consistency, we consider the following manifold regularizer based on the symmetrized KL-divergence between $p(z|d_i)$ and $p(z|d_j)$:

$$\begin{aligned} \Psi &= \frac{1}{2}\sum_i \sum_j W_{ij}^d \{\mathrm{KL}\left[p(z|d_i)\,||\,p(z|d_j)\right] \\ &\quad + \mathrm{KL}\left[p(z|d_j)\,||\,p(z|d_i)\right]\}, \end{aligned} \tag{5}$$

where $\mathrm{KL}\left[p(z|d_i)\,||\,p(z|d_j)\right] = \sum_k p(z_k|d_i) \log \frac{p(z_k|d_i)}{p(z_k|d_j)}$. In fact, the regularizer (5) was incorporated into PLSA in [3] where only the document manifold regularization is considered.

In a similar manner, the manifold regularizer for words is given by

$$\Omega = \frac{1}{2}\sum_i\sum_j W^w_{ij}\{\mathrm{KL}\left[p(y|w_i)\,||\,p(y|w_j)\right] + \mathrm{KL}\left[p(y|w_j)\,||\,p(y|w_i)\right]\}, \quad (6)$$

where $\mathrm{KL}\left[p(y|w_i)\,||\,p(y|w_j)\right] = \sum_l p(y_l|w_i)\log\frac{p(y_l|w_i)}{p(y_l|w_j)}$.

In the dual regularization, we can easily incorporate partially-labeled data, by constructing the edge weight matrix W^d in the following way:

$$W^d_{ij} = \begin{cases} 1 & \text{if } \gamma(d_i) = \gamma(d_j), \\ \exp\left\{-\frac{\|d_i - d_j\|^2}{\beta}\right\} & \text{if } d_i \in \mathcal{N}(d_j) \text{ or } d_j \in \mathcal{N}(d_i), \\ 0 & \text{otherwise,} \end{cases}$$

where $\gamma(d_i)$ denotes the label of d_i (when available) and the parameter β was set as the averaged distance over all pairs of k-nearest neighbors of data points. In the case of semi-supervised learning (where a few labeled data points are available), we use this edge weight matrix. In the case of supervised learning, the edge weight matrix is filled with 0 or 1, depending on whether $\gamma(d_i)$ is equal to $\gamma(d_j)$ or not, for pairs of documents in the training set.

We define row(or column) sum of $\boldsymbol{W}^d$ as $\boldsymbol{D}^d$, and define the unnormalized graph laplacian $\boldsymbol{L}^d = \boldsymbol{D}^d - \boldsymbol{W}^d$, which is a discrete approximation to the Laplace-Beltrami operator on the data manifold [1]. The unnormalized graph laplacian $\boldsymbol{L}^w$ could be computed in a similar manner.

3.2. EM Algorithm

The regularized expected complete-data log-likelihood for MPMTF is given in (4), where the dual regularization is described in the previous section. Here we derive an EM algorithm to find a solution that maximizes (4), which is summarized in Algorithm 1.

Following the work in [3], we derive update rules for the M-step. We provide the update rule only for $p(w_i|y_l)$, since the same technique is applied to derive the update rule for $p(d_j|z_k)$. Denote by $\mathcal{F}$ the terms which involve only $p(w|y)$ in (4):

$$\mathcal{F} = \mathcal{L}_c - \mu\Omega. \quad (11)$$

The distribution over words conditioned on latent variable should satisfy the sum-to-one constraint, leading to the following Lagrangian:

$$\mathcal{E} = \mathcal{L}_c - \mu\Omega + \sum_{l=1}^{R}\alpha_l\left[1 - \sum_{i=1}^{M}p(w_i|y_l)\right]. \quad (12)$$

Solve the stationary point equation $\frac{\partial\mathcal{E}}{\partial p(w_i|y_l)} = 0$:

$$\begin{aligned} &\frac{\sum_{j,k} C_{ij}p(y_l, z_k|w_i, d_j)}{p(w_i|y_l)} - \alpha_l \\ &-\frac{\mu}{2}\sum_s\left[\log\frac{p(w_i|y_l)}{p(w_s|y_l)} + 1 - \frac{p(w_s|y_l)}{p(w_i|y_l)}\right]W^w_{is} = 0, \\ &\frac{\sum_{j,k} C_{ij}p(y_l, z_k|w_i, d_j)}{p(w_i|y_l)} - \alpha_l \\ &-\frac{\mu}{p(w_i|y_l)}\sum_s\left[p(w_i|y_l) - p(w_s|y_l)\right]W^w_{is} = 0, \quad (13) \end{aligned}$$

where we used Bayes formula $p(y_l|w_i) \propto p(w_i|y_l)p(y_l)$ and the property of log function, $\log(x) \approx 1 - \frac{1}{x}$ when x approaches 1.

Note that

$$\begin{aligned} &\sum_s\left[p(w_i|y_l) - p(w_s|y_l)\right]W^w_{is} \\ &= p(w_i|y_l)\sum_s W^w_{is} - \sum_s p(w_s|y_l)W^w_{is} \\ &= p(w_i|y_l)D^W_{ii} - \sum_s p(w_s|y_l)W^w_{is} \\ &= \left[\boldsymbol{L}^w[p(w|y)]_{:,l}\right]_i, \end{aligned}$$

where $\boldsymbol{L}^w$ is the graph Laplacian for the word neighborhood graph, i.e., $\boldsymbol{L}^w = \boldsymbol{D}^w - \boldsymbol{W}^w$ where $\boldsymbol{D}^w$ is the degree matrix which is a diagonal matrix with $[\boldsymbol{D}^w]_{ii} = \sum_{j=1}^{M}[\boldsymbol{W}^w]_{ij}$. We define diagonal matrices with its diagonal entries filled with Lagrangian multipliers,

$$\begin{aligned} \boldsymbol{\Upsilon}_l &= \mathrm{diag}(\alpha_l, \alpha_l, \ldots, \alpha_l) \in \mathbb{R}^{M\times M}, \\ \boldsymbol{\Phi}_k &= \mathrm{diag}(\beta_k, \beta_k, \ldots, \beta_k) \in \mathbb{R}^{N\times N}. \end{aligned}$$

We treat $[p(w|y)]$ as an $M \times R$ matrix and its (i,j)-entry is represented by $[p(w|y)]_{ij} = p(w_i|y_j)$. With these definitions, we construct stationary point equations (13) for $i = 1, \ldots, M$,

$$\begin{bmatrix} \sum_{j,k} C_{ij}p(y_l, z_k|w_1, d_j) \\ \cdots \\ \sum_{j,k} C_{Mj}p(y_l, z_k|w_M, d_j) \end{bmatrix} - (\boldsymbol{\Upsilon}_l + \mu\boldsymbol{L}^w)[p(w|y)]_{:,l} = 0,$$

yielding the Lagrangian multiplier α_l computed as

$$\alpha_l = \sum_i\sum_j C_{ij}\sum_k p(y_l, z_k|w_i, d_j) - \mu\sum_i\left[\boldsymbol{L}^w[p(w|y)]_{:,l}\right]_i.$$

Substituting this α_l back into (13), the update rule (9) in Algorithm 1 is obtained after a few simple mathematical manipulations. Remaining update rules (8) and (10) are derived in the similar manner, by solving corresponding stationary point equations.

Algorithm 1 EM algorithm for MPMTF.

1: **E-step**: Compute the posterior distributions over latent variables:

$$p(y_l, z_k|w_i, d_j) = \frac{p(w_i|y_l)p(d_j|z_k)p(y_l, z_k)}{\sum_l \sum_k p(w_i|y_l)p(d_j|z_k)p(y_l, z_k)}, \tag{7}$$

for $l = 1, \ldots, R$ and $k = 1, \ldots, K$.

2: **M-step**: We re-estimate parameters $p(w_i|y_l)$, $p(d_j|z_k)$, and $p(y_l, z_k)$ such that (4) is maximized where the expectation is computed with respect to the posterior distributions (7). Denote by $[p(w|y)] \in \mathbb{R}^{M\times R}$ and $[p(d|z)] \in \mathbb{R}^{N\times K}$ parameter matrices and their (i, j)-entry is represented by $p(w_i|y_l)$ and $p(d_j|z_k)$. Column and row vectors are represented by $[p(w|y)]_{:,l}$ and $[p(w|y)]_{i,:}$, respectively. Update rules are given by

$$p(y_l, z_k) \leftarrow \frac{\sum_i \sum_j C_{ij} p(y_l, z_k|w_i, d_j)}{\sum_i \sum_j C_{ij}}, \tag{8}$$

$$[p(w|y)]_{:,l} \leftarrow [p(w|y)]_{:,l} \odot \left\{ \sum_j \left[\frac{X}{[p(w,d)]}\right]_{:,j} \right\} \odot \left\{ \mu\alpha_l^{-1} \sum_j \left[[p(d|z)] \left[[p(y,z)]_{l,:}\right]^\top \right]_j L^w [p(w|y)]_{:,l} \right\}, \tag{9}$$

$$[p(d|z)]_{:,k} \leftarrow [p(d|z)]_{:,k} \odot \left\{ \left[\sum_i \left[\frac{X}{[p(w,d)]}\right]_{i,:} \right]^\top \right\} \odot \left\{ \lambda\beta_k^{-1} \sum_i \left[[p(w|y)][p(y,z)]_{:,k}\right]_i L^d [p(d|z)]_{:,k} \right\}, \tag{10}$$

where $\odot$ represents Hadamard product (element-wise multiplication) and $\left(\frac{X}{[p(w,d)]}\right)$ is computed by element-wise division. α_l's and β_k's are Lagrangian multipliers to update $[p(w|y)]$ and $[p(d|z)]$, respectively. $L^w \in \mathbb{R}^{M\times M}$ and $L^d \in \mathbb{R}^{N\times N}$ are graph Laplacians associated with word and document neighborhood graphs, i.e., $L^w = D^w - W^w$, for instance.

4. NUMERICAL EXPERIMENTS

We used several benchmark data sets, the statistics of which are summarized in Table 1, where COIL20[1] is an image data set and the others (CSTR[2], K1a, K1b, Waps[3], Webkb4s[4], NG2 and NG10[5]) are document data sets.

Table 1. Description of data sets.

Data sets	#data	#attributes	#classes
COIL20	1440	1024	20
CSTR	602	3791	4
K1a	13879	2340	20
K1b	13879	2340	6
Waps	8460	1560	20
Webkb4s	1200	1000	4
NG2	7568	500	2
NG10	5346	497	10

We evaluate the clustering accuracy as a performance measure, which is computed using the true label information:

$$\text{AC} = \frac{1}{N} \sum_{i=1}^{N} \delta(\text{map}(\gamma(d_i)), t_i),$$

where $\gamma(d_i)$ is the estimated cluster index for document d_i and t_i is the true label. In the clustering problems, we do not have specific index of the clusters, so we have to match the estimated cluster indices to the true cluster labels. For this purpose, we use the mapping 'map$(\cdot)$', which is searched over all possible permutations. The mapping function is implemented by Kuhn-Munkres algorithm [10] to solve the perfect matching problem. $\delta(x, y)$ is delta function equal to one when $x = y$ and otherwise equal to zero.

We compared several existing methods such as PLSA [7], GNMF [2], PMTF [13], LTM [3] to our MPMTF, in the unsupervised clustering tasks. Clustering accuracy averaged over 20 independent trials is summarized in Table 2, where MPMTF outperforms other methods in most of cases. In experiments, we set K as the number of true clusters. The regularization parameters were determined from several tries on different parameter values, and we used $\lambda = 1000$, $\mu = 1000$ for all the datasets. We used 10-nearest neighbors in constructing a neighborhood graph. Fig. 1 shows the comparison of MPMTF to GNMF and LTM in the task of semi-supervised clustering where 'CSTR' data set is used.

[1] http://www1.cs.columbia.edu/CAVE/software/softlib/coil-20.php
[2] http://www.cs.rochester.edu/trs/
[3] http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview
[4] http://www.cs.cmu.edu/afs/cs/project/theo-20/www/data/
[5] http://people.csail.mit.edu/jrennie/20Newsgroups/

Table 2. Clustering accuracy on eight data sets.

Datasets	PLSA	GNMF	PMTF	LTM	MPMTF
COIL20	0.3970	0.4505	0.3951	0.3975	**0.4662**
CSTR	0.7453	0.7813	0.7431	0.7674	**0.8651**
K1a	0.4093	0.4804	0.4467	0.4605	**0.5208**
K1b	0.5518	0.6727	**0.6903**	0.6621	0.6776
Waps	0.4038	0.4941	0.3745	0.4250	**0.5382**
Webkb4s	0.6541	0.6115	0.6783	0.6497	**0.7180**
NG2	0.6389	0.7276	0.6306	0.7236	**0.7363**
NG10	0.4725	0.5535	0.4843	0.4517	**0.5847**

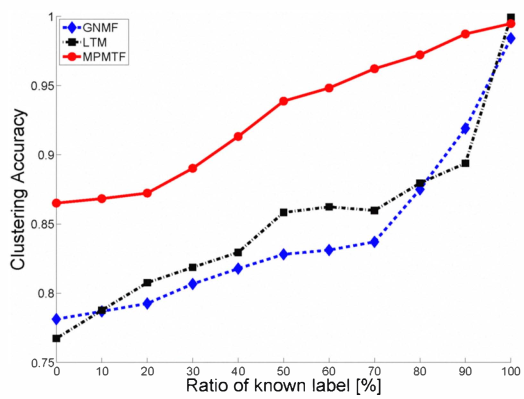

Fig. 1. Clustering accuracy averaged over 20 trials with different portions of labeled data for CSTR data set.

5. CONCLUSIONS

We have presented a method for co-clustering where we have incorporated low-dimensional manifold structure of data and attributes into probabilistic matrix tri-factorization. An EM algorithm was developed to learn the model. In addition, MPMTF could incorporate label information for partially-labeled data to achieve better performance. Experiments on document and image data sets confirmed the useful behavior of MPMTF in the task of clustering.

Acknowledgments: This work was supported by National Research Foundation (NRF) of Korea (KRF-2008-313-D00939, Project No. 2010-0018828) and NRF WCU Program (Project No. R31-2008-000-10100-0).

6. REFERENCES

[1] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, pp. 1373–1396, 2003.

[2] D. Cai, X. He, X. Wu, and J. Han, "Non-negative matrix factorization on manifold," in *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, Pisa, Italy, 2008, pp. 63–72.

[3] D. Cai, X. Wang, and X. He, "Probabilistic dyadic data analysis with local and global consistency," in *Proceedings of the International Conference on Machine Learning (ICML)*, Montreal, Canada, 2009, pp. 105–112.

[4] S. Choi, "Algorithms for orthogonal nonnegative matrix factorization," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, Hong Kong, 2008.

[5] C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix tri-factorizations for clustering," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, Philadelphia, PA, 2006.

[6] E. Gaussier and C. Goutte, "Relation between PLSA and NMF and implications," in *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, Salvador, Brazil, 2005.

[7] T. Hofmann, "Probablistic latent semantic analysis," in *Proceedings of the Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 1999.

[8] T. Hofmann, J. Puzicha, and M. I. Jordan, "Learning from dyadic data," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 11. MIT Press, 1999.

[9] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[10] L. Lovasz and M. Plummer, *Matching Theory*. Akademiai Kiado, 1986.

[11] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, Toronto, Canada, 2003.

[12] J. Yoo and S. Choi, "Orthogonal nonnegative matrix factorization: Multiplicative updates on Stiefel manifolds," in *Proceedings of the 9th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL)*, Daejeon, Korea, 2008.

[13] ——, "Probabilistic matrix tri-factorization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, 2009.