ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 5, Issue 1 , January 2018

A Semi- Non-Negative Matrix Factorization and Principal Component Analysis Unified Framework for Data Clustering

V.Yuvaraj, N.SivaKumar

Assistant Professor, Department of Computer Science, K.S.G college of Arts and Science, (Affiliated to Bharathiar University), Coimbatore-641015. Tamilnadu, India.
Assistant Professor/HOD, Department of Information Technology, K.S.G college of Arts and Science, (Affiliated to Bharathiar University), Coimbatore-641015. Tamilnadu, India.

ABSTRACT: Clustering has received a significant amount of attention as an important problem with many applications, and a number of different algorithms have emerged over the years. Recently, the use of Non-Negative Matrix Factorization (NMF) for partitional clustering has attracted much interest. However, the popularity of NMF has significantly increased the proposed multiplicative NMF algorithms which they applied to image data. At present, NMF and its variants have already found a wide spectrum of applications in several areas such as pattern recognition and feature extraction, dimensionality reduction, segmentation and clustering, text mining and neurobiology. Nonnegative Matrix Factorization (NMF) is a popular matrix decomposition method with various applications in e.g. machine learning, data mining, pattern recognition, and signal processing. The non negativity constraints have been shown to result in parts-based representation of the data, and such additive property can lead to the discovery of data's hidden structures that have meaningful interpretations.

KEYWORDS: Data mining, , non-negative matrix factor, Clustering, applications of NMF

I.INTRODUCTION

Clustering has received a significant amount of attention as an important problem with many applications, and a number of different algorithms have emerged over the years. Recently, the use of Non-Negative Matrix Factorization (NMF) for partitional clustering has attracted much interest. However, the popularity of NMF has significantly increased the proposed multiplicative NMF algorithms which they applied to image data. At present, NMF and its variants have already found a wide spectrum of applications in several areas such as pattern recognition and feature extraction, dimensionality reduction, segmentation and clustering, text mining and neurobiology. Nonnegative Matrix Factorization (NMF) is a popular matrix decomposition method with various applications in e.g. machine learning, data mining, pattern recognition, and signal processing. The non negativity constraints have been shown to result in parts-based representation of the data, and such additive property can lead to the discovery of data's hidden structures that have meaningful interpretations.

A. NON-NEGATIVE MATRIX FACTOR

In linear algebra, a Matrix Factorization (MF) is a decomposition of a matrix into a product of matrices. Let the input data matrix $X = (x_1, \ldots . x_n)$ contain n data vectors of dimensionality m, $W = (w_1, \ldots . w_r)$, and $H = (h_1, \ldots . h_n)$. To factorize matrix X into the product of matrices W and H, one can write:

$$X = WH \quad (1)$$

In conventional MF, both the input matrix X and the factorized matrices Wand H can contain either positive or negative entries. The idea of Nonnegative Matrix Factorization (NMF) in which they introduced a factor analysis method called Positive Matrix Factorization (PMF). Given an observed positive data matrix X, PMF solves the following weighted factorization problem with nonnegativity constraints:

$$min_{W \geq 0, H \geq 0} \|A \odot (X - WH)\|_F \quad (2)$$

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 1 , January 2018

where $\|\cdot\|_F$ denotes Frobenius norm, $\odot$ denotes Hadamard (element-wise) product, A is the weighting matrix, and W, H are factor matrices that are constrained to be nonnegative.

NMF attract more research attentions and gain more applications in various fields. Given a nonnegative input data matrix $X \in \mathbb{R}_+^{m \times n}$, NMF finds two nonnegative $W \in \mathbb{R}_+^{m \times r}$ and $H \in \mathbb{R}_+^{r \times n}$ such that

$$X \approx WH \quad (3)$$

The rank r is often chosen so that $r < \min(m, n)$, An appropriate selection of the value r is critical in practice, but its choice is usually problem dependent. Let us write $X_i \approx Wh_i = \sum_{k=1}^{r} w_k \cdot h_{ki}$. One can see that NMF approximates each nonnegative input data vector in X by an additive linear combination of r nonnegative basis columns in W, with nonnegative coefficients in the corresponding column in H. Therefore, the matrix factor W is usually regarded as the basis matrix, the factor H as the coefficient matrix, and the product term WH is called the compressed version of the X or the approximating matrix of X. The additive nature of NMF can often generate parts-based data representation that conveys physical meanings.

II CLUSTERING

Clustering is a combinatorial problem whose aim is to find the cluster assignment of data that optimizes certain objective. The aim of clustering is to group a set of objects in such a way that the objects in the same cluster are more similar to each other than to the objects in other clusters, according to a particular objective. Clustering belongs to the unsupervised learning scope that involves unlabeled data only, which makes it a more difficult and challenging problem than classification because no labeled data or ground truth can be used for training. Cluster analysis is prevalent in many scientific fields with a variety of applications. For example, image segmentation, an important research area of computer vision, can be formulated as a clustering problem.

A. PRINCIPAL COMPONENT ANALYSIS AND NON-NEGATIVE DATA

There has been published several papers where NMF outperforms PCA. Analyze the outcome of Principal Component Analysis (PCA) when the observation is nonnegative. This analysis shows that PCA will output only one purely positive component and the remaining components will contain both positive and negative elements.

In PCA a set of vectors $v_1, \ldots . v_m \in \mathbb{R}^n$ is projected to a r-dimensional space such that most variance is obtained. In other words, PCA finds a matrix $P_{PCA} \in \mathbb{R}^{r \times n}$ with orthonormal row vectors that fulfils

$$P_{PCA} = \arg max_{P \in \mathbb{R}^{r \times n}} \|PV\|_F^2 \quad (4)$$

Note, there are many solutions to the maximization problem. Therefore, argmax means that P_{PCA} is just one of the optimal matrices.

B. EVALUATION MEASURES

Effective evaluation measures are crucial for quantifying and comparing the performance of clustering methods. Generally, there are three different categories of evaluation criteria: internal, relative, and external. Internal criteria examine the resulting clusters directly from the original input data. Relative criteria compare several clustering structures, which can be produced by different algorithms, and decide which one may best characterize the data to certain extent. External criteria have been commonly used; they measure the clustering performance by using the known information (often referred to as ground truth). Two widely used external criteria are

- **Purity, defined as**

$$purity = \frac{1}{2} \sum_{k=1}^{r} max_{1 \le l \le q} \, n_k^l \quad (4)$$

where n_k^l is the number of vertices in the partition k that belong to the ground-truth class l. Purity is easy to understand, as it can be interpreted in a similar way as the classification accuracy in supervised learning. However, purity has a drawback in that it tends to emphasize the large clusters.

- **Normalized Mutual Information, defined as**

$$NMI = \frac{\sum_{i=1}^{K} \sum_{j=1}^{k'} n_{i,j} \; log \left(\frac{n_i, n_j}{n_i m_j} \right)}{\sqrt{\sum_{i=1}^{K} n_i \; log \left(\frac{n_i}{n} \right) \sum_{j=1}^{K'} m_j \log \left(\frac{m_j}{n} \right)}} \quad (5)$$

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 5, Issue 1 , January 2018

where K and K' respectively denote the number of clusters and classes; $n_{i,,j}$ is the number of data points agreed by cluster i and class j; $n_{i,}$ and $m_{i,}$ denote the number of data points in cluster i and class j respectively; and n is the total number of data points in the dataset. NMI examines the quality of clusters from an information-theoretic perspective. Compared with purity, NMI tends to be less affected by the cluster sizes due to the normalization step given by its denominator, but it is not that intuitive as purity for people to interpret.

III. SIGNIFICANCE OF THE SYSTEM

The concept of matrix factorization is used in a wide range of important applications and each matrix factorization relies on an assumption about its components and its underlying structures, it is an essential process in each application domain. Very often, the data sets to be analyzed are non-negative, and sometimes they also have a sparse representation. The spectral clustering, normalized cuts, and Kernel k-means are particular cases of clustering with NMF under a doubly stochastic constraint. They also considered the symmetric matrix decomposition under non-negativity constraints similar to those formulated.

IV. LITERATURE SURVEY

The propose system to improve the multi-view point algorithm in two actual ways. First, the current PVC algorithm is considered specifically for two-view datasets. We extend this algorithm for the k multi-view point. Second, extend our k multiple-view algorithm to include view-specific graph laplacian regularization. This enables the proposed algorithm to exploit the intrinsic geometry of the data distribution in each view. The compare propose method against existing clustering algorithm on both text and image datasets. The baseline methods and datasets used in this method are more exhaustive than what is used in the PVC work. The experiments show that the proposed GPMVC method outperforms PVC and other competitive baseline methods on all the different datasets. It also provides insights into how our algorithm performs when there is a skew in the distribution of partial examples across views.

A. NON-NEGATIVE MATRIX FACTORIZATION

NMF with the sum of formed error cost meaning is equal to a comfortable K-means clustering, the most this algorithm used for unsupervised dataset learning. The NMF with the I-divergence cost function is matched into probabilistic latent semantic indexing analysis, unsupervised learning method mostly used for text analysis. Many existing data mining and machine learning algorithm can be used to solve the NMF problem.

Here consider the input data row and column matrix denoted by $X = (X_1, X_2, \dots X_n)$contain the *n* no ofdata column vectors. factorize$X$ into two matrices,

$$X \approx FG^T \quad (1)$$

where$X \in \mathbb{R}^{p\times n}$, $F \in \mathbb{R}^{p\times k}$ and $G \in \mathbb{R}^{n\times k}$. Generally, $p < n$ and the rank of matrices$F,G$is much lower than the rank of X, i.e., $k \ll \min(p;n)$. $F;G$are obtained by minimizing a cost function. The most common cost function is the sum of squared errors,

$$Min_{F,G\geq 0} J_{sse} = \|X - FG^T\|^2 \quad (2)$$

the matrix norm is indirectly assumed to be the Frobenius norm. A rank non-deficiency disorder is assumed for F, G. the cost function is the so-called I-divergence:

$$Min_{F,G\geq 0} J_{ID} = \sum_{i=1}^{m}\sum_{j=1}^{n}\left[X_{ij} log \frac{X_{ij}}{(FG^T)_{ij}} - X_{ij} + (FG^T)_{ij}\right] \quad (3)$$

It's easy to show that the dissimilarity I(x) = x log x−x+1 ≥ 0 holds when x ≥ 0; the equivalence holds when x = 1. The quantity I(u;v) = (u=v) log(u=v)−u=v+1 is called I-divergence,

NMF and K-means Clustering

Here proposed K-means based clustering algorithm is one of the best clustering algorithm for high dimension dataset. Let consider X= $(X_1, X_2, \dots X_n)$ be n data points. The divider them into K equally disjoint clusters. The K-means clustering objective can be written as

$$J_{kmeans} = \sum_{i=1}^{n} min_{1\leq K\leq k}\|x_i - f_k\|^2 = \sum_{k=1}^{k}\sum_{i\in c_k}\|x_i - f_k\|^2 \quad (4)$$

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 5, Issue 1 , January 2018

The following theorem shows that NMF is inherently related to K-means clustering algorithm

$$min_{F\geq 0, G\geq 0}\|X - FG^T\|^2 \quad st\ G^TG = 1 \quad (5)$$

is equivalent to K-means clustering. The k-means if X and F have mixed-sign accesses. appreciate this relationship in this similarity. Here c is the no of cluster to be consider the $C = (C_1, C_2, \dots C_k)$be the cluster centroids found via base cluster K-means clustering. Let H be the cluster indicators: i.e., $h_{ki} = 1$ if $x_i$belongs to cluster $c_k; h_{ik} = 0$ otherwise. We can write the K-means cluster objective as $J = \sum_{i=1}^{n}\sum_{k=1}^{K} h_{ik}\|x_i - c_k\|^2 = \|X - CH^T\|^2$ From this analogy, in NMF F has the meaning of cluster centroids and G is the cluster indicator. Thus K-means and NMF have similar objective function but with different constraints. The originally K-means objective function can be expressed to ignore the nonnegativity constraint while keeping the orthogonality restriction, the principal component is the solution. On the other hand, if ignore the orthogonality while keeping the nonnegativity, NMF is the solution.

B. MULTI-VIEW CLUSTERING

The multi-view point clustering algorithm to analysis based on data point in multi views would be assigned to the same cluster with high likelihood. To apply this instinct in a NMF setting the coefficient matrices (V_i) learnt from dissimilar views are softly regularized towards a common agreement matrix (V^*). This agreement matrix is considered to reflect the latent structure shared by different views. The multi-point NMF clustering setup divergence between the ith coefficient matrix and agreement matrix i.e. k $\|V_i - V^*\|$is minimized. The V_i from multi views might not be similar at the same scale one needs to adopt a normalization policy. Each coefficient matrix V is normalized using matrix Q (where, Q is a diagonal matrix, $Q_{k,k} = \sum_i U_{i,k}$. This gives us the following multi- view NMF-based clustering problem,

$$Min_{u_{i,V_i,V^*}} \sum_{i=1}^{v} (\|X_i - U_iV_i^T\|_F^2 + \mu_i\|V_iQ_i - V^*\|_F^2) \quad (6)$$

S.t $U_i \geq 0$, $V_i \geq 0$, $\forall_i \quad s,t\ 1 \leq i \leq v$

This algorithm to learning a joint representation (V^*) of the data totally ignores the intrinsic geometrical structure of each single view. The existing work shown that respecting the geometrical/low-dimensional manifold information can improve clustering quality. The propose system mainly consider multi-view pint in multi dimension datasets.

$$Penalty_{GR} = \frac{1}{2}\sum_{j,l=1}^{N_i} \|(V_i)_j - (V_i)_l\|^2 \times W_{jl}$$
$$= Tr(V_i^TD_iV_i) - Tr(V_i^TW_iV_i)$$
$$= Tr(V_i^TL_iV_i) \quad (7)$$

To enable this, introduce an added graph regularization penalty. Given a similarity matrix W^2one can define a smoothness consequence.

V. METHODOLOGY

A. DATASET

- ORL: This is image dataset contains a set of 400 face images. Here concept two views one based on raw pixel values and the other comprising of HOG features.
- 3Sources: This three-view text dataset is collected from three online news sources. In total there are 948 news articles covering 416 distinct news stories. Of these stories, 169 were reported in all three sources. For our multi-view experiments the dataset containing 169 articles was used.
- BBC Sports: This text dataset is a collection of sports news articles from the BBC Sport web site. For our Multiview experiments choose the 3-view dataset which containing 282 reports.
- Digit: This image dataset is from the UCI repository and consists of 2000 hand-written digits (0-9). This is a 5-view dataset. Similar to two-view experiments it considers the following two views: 216 profile correlations, 240-pixel averages in 2 x 3 windows.
- Cora: This dataset consists of 2708 scientific publications. It considers the following two views for experiments: number of citations between documents and the term-document matrix.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 5, Issue 1 , January 2018

B. PERFORMANCE EVALUATION

- To measure the clustering performance of the proposed algorithms we use the commonly adopted metrics, the accuracy, the Normalized Mutual Information and the Adjusted Rand Index. The clustering accuracy noted (Acc) discovers the one-to-one relationship between two partitions and measures the extent to which each cluster contains data points from the corresponding class. It is defined as follows:
- $Acc = \frac{1}{n}\sum_{i=1}^{n} \delta\,(C_i, map(\mathcal{P}_i))$ (13)
- where n is the total number of samples, Pi is the ith obtained cluster and Ci is the true ith class provided by the data set. (x; y) is the delta function that equals one if x = y and equals zero otherwise, and map(Pi) is the permutation mapping function that maps the obtained label Pi to the equivalent label from the data set.
- The second measure employed is the Normalized Mutual Information (NMI); it is estimated by
- $NMI = \frac{\sum_{k,l} \frac{nkl}{n}\ log\frac{nnkl}{n_k n_l}}{\sqrt{\sum_k n_k \log\frac{n_k}{n}\left(\sum_l n_l \log\frac{n_l}{n}\right)}}$ (14)
- wheren_k denotes the number of data contained in cluster $C_k(1 \le k \le K), n_l$ is the number of data belonging to the class $L_l(1 \le l \le K)$and n_{kl} denotes the number of data that are in the intersection between cluster $C_k$and class L_l.

C. COMPARISON GRAPHS

Table 4.1: Average computation time vs different algorithms

Datasets	**Semi-NMF- PCA**	**F-Semi-NMF-PCA**	**RF-Semi-NMF PCA**	**NMF- multi-view graph clustering**
Coil	1293	1047	1323	967
Orl	1696	1,584	1,718	1232
Webkb	1430	1235	1549	1104

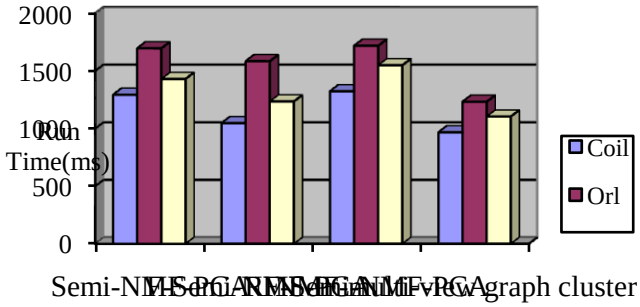


Figure 4.1 Comparison of run time Vs different datasets

Table 4.2 Results obtained by the compared methods on different data sets in terms Acc, NMI and ARI.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 1 , January 2018

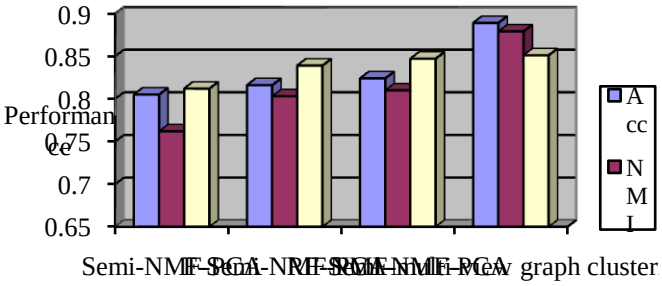


Figure 4.2 Results obtained by the compared methods on Coil data sets in terms Acc, NMI and ARI.

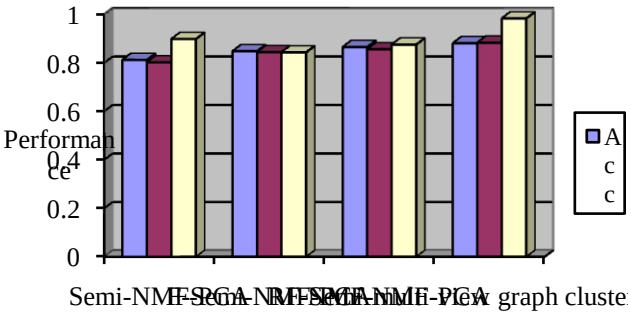


Figure 4.3 Results obtained by the compared methods on Orl data sets in terms Acc, NMI and ARI.

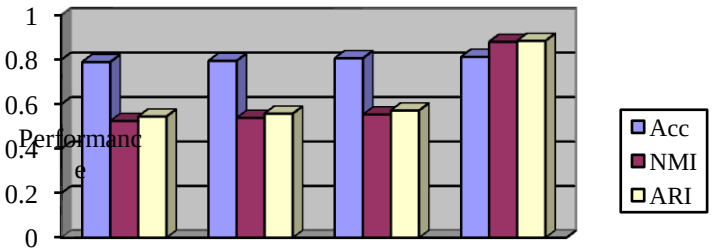


Figure 4.4 Results obtained by the compared methods on Webkb data sets in terms Acc, NMI and ARI

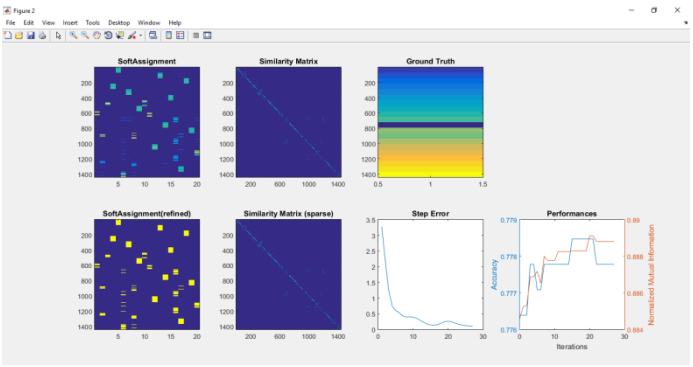
ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

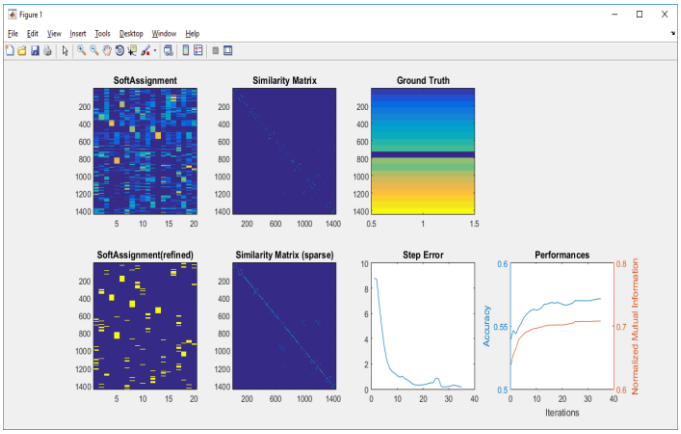
Vol. 5, Issue 1 , January 2018

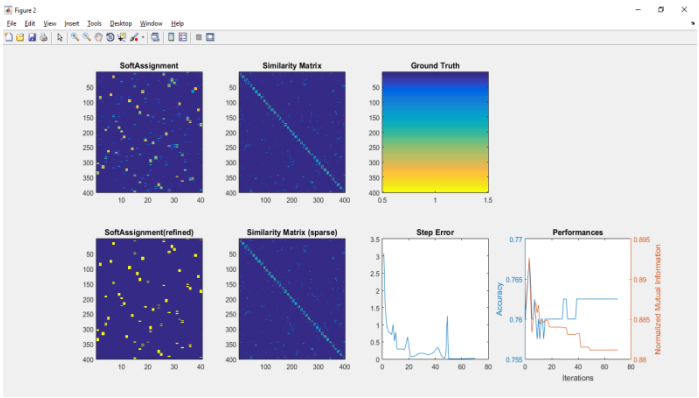
D. SCREENSHOTS

Coil Dataset



Oral Dataset





ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 5, Issue 1 , January 2018

VI. CONCLUSION AND FUTURE WORK

The proposed Multi View Clustering (MVC) algorithm is shown to be effective in states with incomplete or missing information in the high dimension dataset. The propose system to improve the MVC algorithm to support multi-views and view-specific based graph Laplacian regularization. The MVC algorithm is considered specifically for two-view datasets. The first view k-multi-view dimension. The second view k multiple-view algorithm to include view-specific graph Laplacian regularization. The propose algorithm to activity the inherent geometry of the data distribution in each view. The experiments show that the propose multi-view graph clustering (MVGC) algorithm outperforms MVC and other competitive baseline clustering method methods on all the different datasets.

FUTURE WORK

The future work observations show that our algorithm is a good candidate to apply it to image segmentation and text dataset, that will be next task. The applyspectral clustering algorithms using statistical alarm method. The alarm bound discloses that the clustering rate is closely related to the amount of data alarm one can make the clustering rate small by reducing the amount of alarm. The future work shows the that clustering rate converges to zero as the number of representative points produces. These results provide a theoretical foundation for algorithms and also have potentially wider applicability. In particular, a natural direction to pursue in future work is the use of other local data reduction methods (e.g., data squashing and condensation methods) for pre-processing; this bound can be extended to these methods. The future plan to explore other methods for assigning clustering membership to the original data according to the membership of the representative data based on local optimization and edge-swapping methods.

REFERENCES

[1] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, “Clustering with bregman divergences.” Journal of Machine Learning Research, vol. 6, pp. 1705–1749, 2005.
[2] M. Belkin and P. Niyogi. “Laplacian eigenmaps and spectral techniques for embedding and clustering. “In NIPS, 2001.
[3] P. Carmona-Saez, R. D. Pascual-Marqui and A. Pascual-Montano, “Biclustering of gene expression data by non-smooth nonnegative matrix factorization.” BMC Bioinformatics, vol. 7(78), pp. 1–18, 2006.
[4] H. Cho, I. Dhillon, Y. Guan, and S. Sra, “Minimum sum squared residue based co-clustering of gene expression data.” SIAM SDM, pp. 114–125, 2004.
[5] I. S. Dhillon. “Co-clustering documents and words using bipartite spectral graph partitioning.“ ACM SIGKDD, pages 269–274, 2001.
[6] C. Ding, X. He, and H. D. Simon, “On the equivalence of nonnegative matrix factorization and spectral clustering.” SIAM SDM, pp. 606–610, 2005.
[7] C. Ding, T. Li, and W. Peng., “Nonnegative matrix factorization and probabilistic latent semantic indexing: Equivalence,chi-square statistic, and a hybrid method.” AAAI, vol. 42, pp. 137–143, 2006.
[8] C. HQ, Ding, T. Li and M. Jordan. I. “Convex and seminonnegative matrix factorizations. “TPAMI, 32(1), pp. 45–55, 2010.
[9] Q. Gu and J. Zhou, “Co-clustering on manifolds. “ACM SIGKDD, 2009.
[10] J. V. D. Guillamet and B. Schiele, “Introducing a weighted nonnegative matrix factorization for image classification.” Pattern Recognition Letters, vol. 24(14), pp. 2447—2454, 2003.
[11] T. Hofmann, “Unsupervised learning by probabilistic latent semantic analysis.” Machine Learning, vol. 42, pp. 177–196, 2001.
[12] J. Kim and H. Park., “Sparse nonnegative matrix factorization for clustering.” Technical report, Georgia Institute of Technology, 2008.
[13] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization. “NIPS , volume 13, pp. 556–562, 2001.
[14] T. Li and C. Ding., “The relationships among various nonnegative matrix factorization methods for clustering.” IEEE ICDM, pp. 362—371, 2006.
[15] W. Liu and N. Zheng, “Non-negative matrix factorization based methods for object recognition.” Pattern Recognition Letters, vol. 25(8), pp. 893—897, 2004.
[16] F. Shahnaz, M. Berry, P. Pauca, and R. Plemmons, “Document clustering using non-negative matrix factorization.” Information Processing and Management, vol. 42, pp. 373–386, 2006.
[17] A. Strehl and J. Ghosh, “Cluster ensembles - a knowledge reuse framework for combining multiple partitions. “Machine Learning Research, pp. 583–617, 2002.
[18] H. Wang, F. Nie, H. Huang and F. Makedon, “Fast nonnegative matrix tri-factorization for large-scale data co-clustering. “IJCAI, 2011.
[19] R. Zass and A. Shashua, “A unifying approach to hard and probabilistic clustering.” IEEE ICCV, pp. 294–301, 2005.
[20] D. Cai and X. He and J. Han, “SRDA: An efficient algorithm for large-scale discriminant analysis.” TKDE, vol. 20, pp. 1–12, 2008.-272, 2010.