NMF-BASED COMPREHENSIVE LATENT FACTOR LEARNING WITH MULTIVIEW DATA

Hua Zheng[1,3], Zhixuan Liang[2], Feng Tian[1], Zhong Ming[3]*

[1]Faculty of Science and Technology, Bournemouth University, UK
[2]Department of Computing, The Hong Kong Polytechnic University, HK, China
[3]Colledge of Computer Science and Software Engineering, Shenzhen University, China
zhenghuamail@126.com, 18094409g@connect.polyu.hk, ftian@bournemouth.ac.uk, mingz@szu.edu.cn

ABSTRACT

Multiview representations reveal the latent information of the data from different perspectives, consistency and complementarity. Unlike most multiview learning approaches, which focus only one perspective, in this paper, we propose a novel unsupervised multiview learning algorithm, called comprehensive latent factor learning (CLFL), which jointly exploits both consistent and complementary information among multiple views. CLFL adopts a non-negative matrix factorization based formulation to learn the latent factors. It learns the weights of different views automatically which makes the representation more accurate. Experiment results on a synthetic and several real datasets demonstrate the effectiveness of our approach.

Index Terms— Comprehensive, multiview learning, latent factor learning, non-negative matrix factorization (NMF).

1. INTRODUCTION

In many real-world applications such as data analytics in video surveillance, image processing and natural language processing, data are collected from diverse domains or obtained from various feature extractors. They exhibit heterogeneous properties which are called multiview data [1]. Conventional machine learning algorithms cannot handle these multiview data well. They tend to concatenate all multiple views into one single view to adapt to the learning setting [1]. However, this may cause over-fitting, if a training sample is small.

Many existing multiview algorithms try to construct a latent subspace shared by multiple views [2–5]. In the latent subspace, distinct views are connected with each other, and the complementary information underlying different views is integrated. The multiview subspace learning (MSL) learns a unified feature representation from the subspace of all views, so all views share one representation. Non-negative matrix factorization (NMF) is one of the most popular and competitive subspace learning method.

Some works have been done in the area of NMF-based multiview learning. Greene et al. [6] proposed an integrated approach for multiview learning. The combination is performed by applying an approach based on matrix factorization to group related clusters produced on individual views. Later, Liu et al. [7] proposed a new framework which formulates a joint matrix factorization process with the constraint that pushes each view into a common consensus. After that, Ou et al. [8] proposed a co-regularized multiview NMF method with a correlation constraint. It exploits the complementary information of multiple views through the co-regularization to accommodate the presence of the noisy views. The authors in [9] took the local geometric structure of each view into consideration, and penalized the disagreement of different views at the same time. And Wang et al. [10] proposed a diverse NMF algorithm which adds a diversity constraint to ensure that data vectors from different views be as diverse as possible, by regularizing the dot product of two vectors from different views close to zero.

All the methods above only focus on one type of perspective among multiple views. For example, Liu [7] focuses on the consistency while Wang [10] focuses on the complementarity. The consistency aims to maximize the agreement among the multiple views, while the complementarity states that each view may contain some knowledge that other views do not have [1]. These two kinds of approaches fail to consider either specific or common factors, which may have a big influence on the accuracy of clustering after using the above NMF-based methods.

To address this limitation, in this paper, we propose a multiview learning algorithm, called comprehensive latent factor learning (CLFL), by jointly exploring both the perspectives of consistency and complementarity for multiview data. Firstly, we study the weight of each view according to the difference of latent factor matrices and the consensus matrix. Those views which have more common information will be given a higher weight. The latent factor for each view is then divided into two parts [11]. One is the common latent factor shared across multiple views, while the other is the specific latent factor to each view. After learning these two parts re-

---

spectively, the common and specific latent factors for the representation were obtained. Then, we integrate them into an optimal representation for clustering. The main contributions of this paper are summarized as follows.

1) Different from most existing multiview latent factor learning methods, CLFL exploits both consistency and complementarity. Therefore, a more comprehensive latent representation can be obtained, representing the original data more accurately.

2) We introduce a single parameter γ to control the distribution of weighting factors for NMF-based multiview latent representation learning, which is easy to adjust and discover the best weights. The weight factor of each view can be assigned automatically depending on the dissimilarity between the representation matrix of each view and the consensus matrix. The weights will be used during the integration of specific latent factors from different views.

3) As a result of learning the weight factor of each view and combining consistency and complementarity together, CLFL is more adaptable to the number of views than other methods, because we use all views of a dataset to get a competitive result, unlike other methods which only choose two or three most important views for clustering. Also, we use another parameter β to control and adjust the proportion of specific and common latent factors. This enables us to achieve the optimal representation for different datasets.

2. BRIEF REVIEW OF NMF

In this section, we briefly introduce non-negative matrix factorization (NMF) [12]. Given an input non-negative data matrix $X = [x_1, x_2, ..., x_N] \in \mathbb{R}^{M \times N}$, each column of $\mathbf{X}$ is an instance vector. NMF aims to find two non-negative matrices $\mathbf{W} \in \mathbb{R}^{M \times K}$ and $\mathbf{H} \in \mathbb{R}^{N \times K}$ whose product can well approximate the original matrix $\mathbf{X}$. In particular, $\mathbf{H}$ can be considered as the new representation of data in terms of the basis $\mathbf{W}$. The cost function of standard NMF is defined as

$$\min \left\| \mathbf{X} - \mathbf{W}\mathbf{H}^T \right\|_F^2 \quad s.t.\mathbf{W}, \mathbf{H} \geq 0 \tag{1}$$

This standard NMF can be extended to multiview setting by adding the cost function of each single view together. Suppose a dataset has V views, the multiview learning objective function becomes [13]:

$$\sum_{v=1}^{V} \min \left\| \mathbf{X}^{(v)} - \mathbf{W}^{(v)}\mathbf{H}^{(v)T} \right\|_F^2 \quad s.t.\mathbf{W}, \mathbf{H} \geq 0 \tag{2}$$

3. COMPREHENSIVE LATENT FACTOR LEARNING (CLFL)

MultiNMF [7] has become the most famous and effective method for latent factor learning, which learns a joint view representation. However, it fails to discover the view-specific latent factors. We aim to address this issue by proposing a new method CLFL which exploits both specific and common latent factors.

3.1. The Objective Function of CLFL

We define the objective function as:

$$\sum_{v=1}^{V} \left\| \mathbf{X}^{(v)} - \mathbf{W}^{(v)}\mathbf{H}^{(v)T} \right\|_F^2 + \sum_{v=1}^{V} (\alpha^{(v)})^\gamma \left\| \mathbf{H}^{(v)}\mathbf{Q}^{(v)} - \mathbf{H}^* \right\|_F^2$$

$$s.t. \mathbf{W}^{(v)} \geq 0, \mathbf{H}^{(v)} \geq 0, \mathbf{H}^* \geq 0, \sum_{v=1}^{V} \alpha^{(v)} = 1$$

$$\mathbf{Q}^{(v)} = Diag(\sum_{v=1}^{V} \mathbf{W}_{i,1}^{(v)}, \sum_{v=1}^{V} \mathbf{W}_{i,2}^{(v)}, ..., \sum_{v=1}^{V} \mathbf{W}_{i,K}^{(v)})$$

$$\tag{3}$$

$\mathbf{H}^*$ is referring to the common latent factor of all views. We use a single parameter γ to control the distribution of weight factors $\alpha^{(v)}$ in all V views, such that the important views will be assigned bigger weights [14]. Also, in order to make different $\mathbf{H}^{(v)}$ comparable, we constrain $\left\| \mathbf{W}_{.,k}^{(v)} \right\|_1 = 1$ by introducing auxiliary variables $\mathbf{Q}^{(v)}$ [7] to simplify the computation.

3.2. Optimization

The joint optimization function in (3) is not convex over all variables simultaneously. Thus, we propose an iterative optimization algorithm as done in [15]. For each view we have:

$$J = \left\| \mathbf{X}^{(v)} - \mathbf{W}^{(v)}\mathbf{H}^{(v)T} \right\|_F^2 + (\alpha^{(v)})^\gamma \left\| \mathbf{H}^{(v)}\mathbf{Q}^{(v)} - \mathbf{H}^* \right\|_F^2 \tag{4}$$

To solve this optimization problem, we propose an iterative update procedure. Specifically, the following four steps are repeated until the objective function converges:

3.2.1. Fixing $\mathbf{H}^*$, $\mathbf{H}^{(v)}$, $\alpha^{(v)}$, compute $\mathbf{W}^{(v)}$.

Let $\mathbf{\Psi}^{(v)} = [\psi_{i,k}^{(v)}]$ be the Lagrange multiplier matrix for the constraint $w_{i,k}^{(v)} \geq 0$ and L be the Lagrange $L = J + Tr(\mathbf{\Psi}^{(v)}\mathbf{W}^{(v)T})$ where $Tr(.)$ is the trace function. The equation leads to the following updating rules [16]:

$$\mathbf{W}_{i,k}^{(v)} \leftarrow \mathbf{W}_{i,k}^{(v)} \frac{(\mathbf{X}^{(v)}\mathbf{H}^{(v)})_{i,k} + (\alpha^{(v)})^\gamma \sum_{j=1}^{N} \mathbf{H}_{j,k}^{(v)}\mathbf{H}_{j,k}^*}{\mathbf{R}_{i,k}}$$

$$\mathbf{R}_{i,k} = (\mathbf{W}^{(v)}\mathbf{H}^{(v)T}\mathbf{H}^{(v)})_{i,k} + (\alpha^{(v)})^\gamma \sum_{l=1}^{M} \mathbf{W}_{l,k}^{(v)} \sum_{j=1}^{N} \mathbf{H}_{j,k}^{(v)2}$$

$$\tag{5}$$

3.2.2. Fixing $\mathbf{H}^*, \mathbf{W}^{(v)}, \alpha^{(v)}$ compute $\mathbf{H}^{(v)}$

For each view, we first normalize the column vectors of $\mathbf{W}^{(v)}$ using $\mathbf{Q}^{(v)}$, then $\mathbf{W}^{(v)} \leftarrow \mathbf{W}^{(v)}\mathbf{Q}^{(v)^{-1}}$, $\mathbf{H}^{(v)} \leftarrow \mathbf{H}^{(v)}\mathbf{Q}^{(v)}$. Thus, the object function was transformed to,

$$\sum_{v=1}^{V} \left\| \mathbf{X}^{(v)} - \mathbf{W}^{(v)}\mathbf{H}^{(v)T} \right\|_F^2 + \sum_{v=1}^{V} (\alpha^{(v)})^\gamma \left\| \mathbf{H}^{(v)} - \mathbf{H}^* \right\|_F^2$$

$$s.t. \mathbf{W}^{(v)} \geq 0, \mathbf{H}^{(v)} \geq 0, \mathbf{H}^* \geq 0, \sum_{v=1}^{V} \alpha^{(v)} = 1 \tag{6}$$

Following the KKT condition [17] for the non-negativity of $\mathbf{H}^{(v)}$, we have the following equation:

$$\mathbf{H}_{j,k}^{(v)} \leftarrow \mathbf{H}_{j,k}^{(v)} \frac{(\mathbf{X}^{(v)T}\mathbf{W}^{(v)})_{j,k} + (\alpha^{(v)})^\gamma \mathbf{H}_{j,k}^*}{(\mathbf{H}^{(v)}\mathbf{W}^{(v)T}\mathbf{W}^{(v)})_{j,k} + (\alpha^{(v)})^\gamma \mathbf{H}_{j,k}^{(v)}} \tag{7}$$

3.2.3. Fixing $\mathbf{W}^{(v)}, \mathbf{H}^{(v)}$ and $\alpha^{(v)}$ compute $\mathbf{H}^*$

$$\frac{\partial O}{\partial \mathbf{H}^*} = \frac{\partial \sum_{v=1}^{V} \left\| \mathbf{H}^{(v)}\mathbf{Q}^{(v)} - \mathbf{H}^* \right\|_F^2}{\partial \mathbf{H}^*}$$
$$= \sum_{v=1}^{V} (\alpha^{(v)})^\gamma (-2\mathbf{H}^{(v)} + 2\mathbf{H}^*) = 0 \tag{8}$$

Solving Eq. 8 gives us $\mathbf{H}^*$ as

$$\mathbf{H}^* = \frac{\sum_{v=1}^{V} (\alpha^{(v)})^\gamma \mathbf{H}^{(v)}\mathbf{Q}^{(v)}}{\sum_{v=1}^{V} (\alpha^{(v)})^\gamma} \geq 0 \tag{9}$$

3.2.4. Fixing $\mathbf{W}^{(v)}, \mathbf{H}^{(v)}$ and $\mathbf{H}^*$ compute $\alpha^{(v)}$

By setting $\mathbf{G}^{(v)} = \left\| \mathbf{H}^{(v)}\mathbf{Q}^{(v)} - \mathbf{H}^* \right\|_F^2$, we have

$$\alpha^{(v)} = \frac{(\gamma \mathbf{G}^{(v)})^{\frac{1}{1-\gamma}}}{\sum_{v=1}^{V} (\gamma \mathbf{G}^{(v)})^{\frac{1}{1-\gamma}}} \tag{10}$$

With these four steps above, we alternatively update $\mathbf{W}^{(v)}$, $\mathbf{H}^{(v)}, \mathbf{H}^*$ as well as $\alpha^{(v)}$ and repeat the process interactively until the objective function is converged.

3.2.5. Calculate the view-specific latent factors $H_s^{(v)}$

To compute the specific latent factor of each view $\mathbf{H}_s^{(v)}$, we apply the approach proposed in [7] to obtain $\mathbf{H}^{(v)}$ and $\mathbf{H}^*$ as follows:

$$\mathbf{H}_s^{(v)} = \mathbf{H}^{(v)} - \mathbf{H}^* \tag{11}$$

Then all the specific latent factors are integrated by using the weights $1 - \alpha^{(v)}$ which were learned previously. The whole specific factor $\mathbf{H}_s$ can be obtained.

$$\mathbf{H}_s = \sum_{v=1}^{V} (1 - \alpha^{(v)})\mathbf{H}_s^{(v)} \tag{12}$$

By integrating the common latent factor and specific latent factor together with a single parameter β, a comprehensive latent factor representation was obtained as

$$\mathbf{H}_F = \beta \mathbf{H}^* + (1 - \beta)\mathbf{H}_s \tag{13}$$

Obviously, the parameter β controls the proportion of these two parts. This makes our method be able to get the optimal representation in all datasets.

Algorithm 1 Algorithm of CLFL

Input: Non-negative Matrix $\{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, ..., \mathbf{X}^{(V)}\}$, parameter γ.
Output: The weights $\alpha^{(v)}$.
1: Normalize each view $\mathbf{X}^{(v)}$ such that $\left\| \mathbf{X}^{(v)} \right\|_1 = 1$
2: Initialize $\mathbf{W}^{(v)}, \mathbf{H}^{(v)}, \mathbf{H}^*$ and $\alpha^{(v)}$
3: **repeat**
4:     **for** $v = 1$ to V **do**
5:         **repeat**
6:             Fixing $\mathbf{H}^*, \mathbf{H}^{(v)}, \alpha^{(v)}$, update $\mathbf{W}^{(v)}$ by Eq.5
7:             Fixing $\mathbf{H}^*, \mathbf{W}^{(v)}, \alpha^{(v)}$ update $\mathbf{H}^{(v)}$ by Eq.7
8:             Fixing $\mathbf{W}^{(v)}, \mathbf{H}^{(v)}$ and $\alpha^{(v)}$ update $\mathbf{H}^*$ by Eq.9
9:             Fixing $\mathbf{W}^{(v)}, \mathbf{H}^{(v)}$ and $\mathbf{H}^*$ update $\alpha^{(v)}$ by Eq.10
10:         **until** Eq.4 converges
11:     **end for**
12: **until** Eq.3 converges
13: Obtain Basis Matrices $\{\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, ..., \mathbf{W}^{(V)}\}$, Coefficient Matrices $\{\mathbf{H}^{(1)}, \mathbf{H}^{(2)}, ..., \mathbf{H}^{(V)}\}$, Consensus Matrix $\mathbf{H}^*$ from MultiNMF [7].
14: **for** $v = 1$ to V **do**
15:     Calculate the specific latent factor of each view by Eq.11
16: **end for**
17: Integrate all the specific latent factors using weights $1 - \alpha^{(v)}$ by Eq.12
18: Integrate the common latent factor and view-specific latent factor together with a single parameter β by Eq.13

---

4. EXPERIMENT

In this section, experiments were conducted to demonstrate the effectiveness of the proposed CLFL in discovering the underlying clustering structure shared by multiple views of data.

4.1. Datasets

One synthetic and three real world datasets were used in the experiments. Below is a brief introduction.

- **Synthetic dataset:** It is generated from two clusters. Each cluster is composed of three Gaussian components that means an instance is represented as three views. For each cluster, five hundreds instances are randomly sampled. More details can be found in [18].

- **3-Sources Text Dataset:** It is collected from three online news sources: BBC, Reuters, and The Guardian.

- **Reuters Multilingual dataset:** It contains feature characteristics of documents that are translated into 5 languages over 6 categories.

- **UCI Handwritten Digit dataset:** This handwritten digits (0-9) data is from the UCI repository.

4.2. Methods to Compare

We compared the proposed approach with several latent factor learning methods, including Single View (**BSV** and **WSV**) [13], **ConNMF** [13], **ColNMF** [19], and **Multi-NMF** [7]. Three metrics, Clustering Accuracy (**AC**), Normalized Mutual Information (**NMI**) and Purity are used to evaluate their performance.

4.3. Clustering Results

In our experiments, we run 20 times and obtain the average performance. The clustering results of different algorithms on four datasets are showed in Table 1. As we can see, CLFL outperforms all the other five algorithms in four different datasets in three metrics. Compared with MultiNMF, CLFL improves performances average 5% on all datasets in terms of AC, NMI, and Purity, which proves the effectiveness of Our comprehensive learning method adding specific latent factor to the common latent factor. Specifically, the weights $\alpha^{(v)}$ play a crucial role to integrate the specific latent factors. In fact, we found that, without the weights $\alpha^{(v)}$ the clustering results even worse than MultiNMF. Also, Fig 1 shows how the performance of CLFL on the digit and 3-source datasets varies with the parameters β. We can see that, in the digit, when β is set nearly 0.6 it achieves the best performance and it will be around 0.7 for the 3-source. So, there is an optimal point between the specific latent factor and the common shared factor. Since it differs from dataset to dataset, we use the parameter β to get the optimal representation in different datasets. Usually, the common shared factor holds a larger proportion around 0.6 to 0.8.
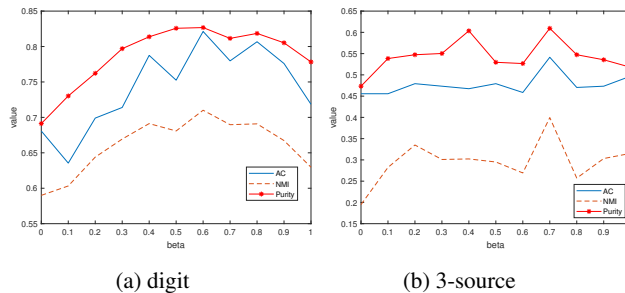
Table 1: AC, NMI and Purity of different methods

Methods	Metrics	Synthetic	3-Source	Reuters	digit
WSV	AC	0.5375	0.3876	0.1894	0.5385
	NMI	0.5023	0.2079	0.1618	0.5398
	Purity	0.6630	0.4586	0.1889	0.5570
BSV	AC	0.5815	0.4970	0.2711	0.6630
	NMI	0.5978	0.2937	0.1064	0.6252
	Purity	0.7640	0.5266	0.2806	0.6700
ConNMF	AC	0.6740	0.4290	0.1939	0.6865
	NMI	0.6535	0.2029	0.1471	0.6507
	Purity	0.7290	0.4734	0.2100	0.7340
ColNMF	AC	0.6550	0.4408	0.2711	0.6005
	NMI	0.5641	0.2163	0.1056	0.5376
	Purity	0.6520	0.5118	0.2917	0.7265
MultiNMF	AC	0.8465	0.6154	0.4483	0.8140
	NMI	0.7455	0.4300	0.3128	0.7336
	Purity	0.8515	0.6272	0.4500	0.8210
CLFL	AC	**0.8920**	**0.6450**	**0.4733**	**0.8740**
	NMI	**0.8008**	**0.4407**	**0.3170**	**0.7722**
	Purity	**0.8895**	**0.6391**	**0.4692**	**0.8780**

4.4. Discussion of the Parameter γ

We use one parameter γ to control the distribution of weight factors for different views [20]. From Eq.(10), we can see that when $\gamma \to \infty$, the weights for all views are equal. When $\gamma \to 1$, the weight factor 1 will be assigned to the view whose $\mathbf{G}^{(v)}$ value is the smallest while 0 is assigned to the weights of the other views. In the experiment, we choose γ from [0.001, 0.01] to give the best $\alpha^{(v)}$ for each view. More theoretical details about how to choose the parameter will be studied in our future work.

5. CONCLUSION

In this paper, we have proposed a novel latent factor learning algorithm called CLFL. It discovers a comprehensive latent representation for multiview data, by exploiting the consistent and complementary information among different views, simultaneously. Also, CLFL learns the weights of all different views automatically. The specific latent factor and the common shared factor are integrated with a single parameter to control their weights for optimal representation of datasets. The clustering experimental results on four different datasets have demonstrated the effectiveness of our approach. In the future work, we will study the theory of choosing parameters and investigate to take local geometrical information into consideration to learn a better representation.

(a) digit   (b) 3-source

Fig. 1: Performance of CLFL with parameters β

6. REFERENCES

[1] Chang Xu, Dacheng Tao, and Chao Xu, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.

[2] Aaron Shon, Keith Grochow, Aaron Hertzmann, and Rajesh P Rao, "Learning shared latent structure for image synthesis and robotic imitation," in *Advances in neural information processing systems*, 2006, pp. 1233–1240.

[3] Ajit P Singh and Geoffrey J Gordon, "Relational learning via collective matrix factorization," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008, pp. 650–658.

[4] Haifeng Liu, Zhaohui Wu, Xuelong Li, Deng Cai, and Thomas S Huang, "Constrained nonnegative matrix factorization for image representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1299–1311, 2012.

[5] Zechao Li, Jing Liu, Xiaobin Zhu, Tinglin Liu, and Hanqing Lu, "Image annotation using multi-correlation probabilistic matrix factorization," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 1187–1190.

[6] Derek Greene and Pádraig Cunningham, "A matrix factorization approach for integrating multiple data views," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2009, pp. 423–438.

[7] Jialu Liu, Chi Wang, Jing Gao, and Jiawei Han, "Multiview clustering via joint nonnegative matrix factorization," in *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM, 2013, pp. 252–260.

[8] Weihua Ou, Fei Long, Yi Tan, Shujian Yu, and Pengpeng Wang, "Co-regularized multiview nonnegative matrix factorization with correlation constraint for representation learning," *Multimedia Tools and Applications*, vol. 77, no. 10, pp. 12955–12978, 2018.

[9] Weihua Ou, Shujian Yu, Gai Li, Jian Lu, Kesheng Zhang, and Gang Xie, "Multi-view non-negative matrix factorization by patch alignment framework with view consistency," *Neurocomputing*, vol. 204, pp. 116–124, 2016.

[10] Jing Wang, Feng Tian, Hongchuan Yu, Chang Hong Liu, Kun Zhan, and Xiao Wang, "Diverse non-negative matrix factorization for multiview data representation," *IEEE transactions on cybernetics*, vol. 48, no. 9, pp. 2620–2632, 2018.

[11] Jing Liu, Yu Jiang, Zechao Li, Zhi-Hua Zhou, Hanqing Lu, et al., "Partially shared latent factor learning with multiview data.," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 26, no. 6, pp. 1233–1246, 2015.

[12] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788, 1999.

[13] Wei Xu, Xin Liu, and Yihong Gong, "Document clustering based on non-negative matrix factorization," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 2003, pp. 267–273.

[14] Jing Wang, Xiao Wang, Feng Tian, Chang Hong Liu, Hongchuan Yu, and Yanbei Liu, "Adaptive multi-view semi-supervised nonnegative matrix factorization," in *International conference on neural information processing*. Springer, 2016, pp. 435–444.

[15] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2011.

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

[17] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[18] Jeong Woo Son, Junkey Jeon, Alex Lee, and Sun-Joong Kim, "Spectral clustering with brainstorming process for multi-view data.," in *AAAI*, 2017, pp. 2548–2554.

[19] Zeynep Akata, Christian Thurau, and Christian Bauckhage, "Non-negative matrix factorization in multimodality data for segmentation and label prediction," in *16th Computer vision winter workshop*, 2011.

[20] Xiao Cai, Feiping Nie, and Heng Huang, "Multi-view k-means clustering on big data.," in *IJCAI*, 2013, pp. 2598–2604.