Global and Local Consistent Multi-view Subspace Clustering

Yanbo Fan[2] Ran He[1,2,3] Bao-Gang Hu[2]
[1]Center for Research on Intelligent Perception and Computing, CASIA
[2]National Laboratory of Pattern Recognition, CASIA
[3]Center for Excellence in Brain Science and Intelligence Technology, CAS
{yanbo.fan, rhe, hubg}@nlpr.ia.ac.cn

Abstract

Multi-view clustering aims to cluster data with multiple sources of information. Comparing with single view clustering, it is challenging to make use of the extra information embedded in multiple views. This paper presents a multi-view subspace clustering method (MSC-GL) by simultaneously combining both the global low-rank constraint and local cross topology preserving constraints. The global constraint maximizes the correlation between representational matrices while encouraging each of them to be low rank. The local constraints enable representational matrices under different views to share local structure information. An efficiently iterative algorithm is developed to minimize the proposed joint learning problem, and extensive experiments on five multi-view benchmarks demonstrate that the proposed model outperforms the state-of-the-art multi-view clustering methods.

Keywords: Multi-view clustering; Subspace clustering; Cross topology preserving; Global low-rank representation

1. Introduction

Multi-view data is becoming more common in many pattern recognition fields, here each view often refers to one source of features. For example, web pages can be represented by multiple views based on texts and images. Superior to single-view data, multi-view data can provide productive and complementary information. And multi-view learning often achieves better performance and has good statistic properties [23, 24]. Among multi-view learning, multi-view clustering has received more and more attention due to its ability to handle the large number of unlabeled data [11–13, 16, 23].

One challenging issue in multi-view clustering is how to establish links between different views, and many algorithms have been developed to address it. They can be categorized into three groups, according to the way they deal with the multiple sources of information [20]. The first group finds a common representation for multi-view data and then clusters the data based on this representation. Following this strategy, Kumar *et al*. [13] focus on consistent cluster results through regularizing the difference between view-specific Laplacian embedding. Liu *et al*. [16] learn a common latent structure shared by multiple views via non-negative matrix factorization. Another available strategy is to directly incorporate the information of different views during the training process [12, 26]. For instance, a co-training based approach is proposed to seek cluster results that agree across different views in [12]. The third group uses a late fusion strategy, and the final result is derived from integrating each individual clustering result [2, 8].

Although these proposed methods can improve the clustering performance for multi-view data to some extent, making better use of the interactions and correlations among different views is still a challenging problem. Inspired by the recent advance in the low-rank representation [9, 10, 14] and graph regularization [11], we propose a multi-view subspace clustering model (*MSC-GL*) which incorporates both the global and local structure information of multi-view data. Firstly, we impose a global low-rank constraint on the representational matrices. It can maximize the correlation between them as well as enhance the low-rank and block diagonal properties of each matrix. In addition, different from the existing methods like [11], we propose to use local cross topology preserving constraints so as to share the local topology information of different views. Furthermore, we develop an efficiently iterative algorithm to solve the proposed jointly learning model. Experimental results on several widely used multi-view databases demonstrate the effectiveness of our model.

2. Subspace Clustering

To better illustrate the main idea of our method, we briefly review the basic idea of subspace clustering in this section. The objective of subspace clustering is to cluster the samples according to their underlying subspaces. Many algorithms have been developed to address it [21], and we

978-1-4799-6100-9/15/$31.00 ©2015 IEEE

mainly focus on the recently developed spectral based subspace clustering methods [6, 11–15, 25]. One key step of these approaches is to construct an affinity matrix $\mathbf{S}$ which measures the pairwise similarity of samples. Recently, many works build $\mathbf{S}$ based on data's self-representation $\mathbf{Z}$ [6, 11, 14, 15, 25]. Given a $d \times n$ data matrix $\mathbf{X}$ consisting of n samples $\{\mathbf{x}_p \in R^d\}_{p=1}^n$, $\mathbf{Z}$ can be obtained by solving

$$\begin{aligned} &\min_{\mathbf{Z}} ||\mathbf{E}||_\ell + R(\mathbf{Z}), \\ &s.t \quad \mathbf{X} = \mathbf{D}\mathbf{Z} + \mathbf{E}, \end{aligned} \tag{1}$$

where $\mathbf{D}$ is a dictionary matrix which can be learned or simply set to $\mathbf{X}$, $R(\mathbf{Z})$ represents the constraints on $\mathbf{Z}$ (*e.g.* the structure low-rank constraint [15] or the sparse constraint [6]), $||\cdot||_\ell$ is a proper matrix norm.

After obtaining the optimal self-representation $\mathbf{Z}$, a common way to construct the affinity matrix $\mathbf{S}$ is setting $\mathbf{S} = (|\mathbf{Z}|+|(\mathbf{Z})^T|)/2$, $|\cdot|$ is the absolute operator. Finally a spectral clustering algorithm like [19] is used on $\mathbf{S}$ to achieve the final clustering result.

3. The Proposed *MSC-GL* Model

In this section, we first present the proposed multi-view subspace clustering model (*MSC-GL*) with both the global and local constraints, and then we come up with an efficiently iterative algorithm to optimize it.

3.1. Problem Formulation

Given data $\mathbf{X} = [\mathbf{X}^1; \mathbf{X}^2; \ldots; \mathbf{X}^m] \in R^{d \times n}$ with m views sampled from c classes, $\mathbf{X}^i \in R^{d_i \times n}$ is the feature matrix corresponding to the i-th view. $\mathbf{X}_p^i \in R^{d_i}$ is the p-th column of $\mathbf{X}^i$, and $\mathbf{Z}^i$ is the learned representational matrix corresponding to the i-th view. To make better use of the complementary information across views, we formulate our objective function as

$$\begin{aligned} &\min_{\mathbf{Z}^1,\ldots,\mathbf{Z}^m} \; J = J_1 + \lambda J_2 + \beta J_3 \\ &\quad = \sum_{i=1}^m ||\mathbf{X}^i - \mathbf{X}^i\mathbf{Z}^i||_F^2 + \lambda \sum_{i,j=1;j\neq i}^m tr(\mathbf{Z}^i\mathbf{L}^j(\mathbf{Z}^i)^T) \\ &\quad + \beta||[\mathbf{Z}^1, \mathbf{Z}^2, \ldots, \mathbf{Z}^m||_* \; , \end{aligned} \tag{2}$$

where $||\cdot||_F$ is the Frobenious norm and $||\cdot||_*$ is the trace norm, $tr(\cdot)$ denotes the trace of matrix.

Let $\mathbf{W}^i$ be the similarity matrix of the i-th view, and W_{pq}^i be the similarity between sample $\mathbf{X}_p^i$ and $\mathbf{X}_q^i$ [3, 22]. Assuming that adjacent samples have similar representational coefficients, the local cross topology preserving constraint over $\mathbf{Z}^i$ is formulated as:

$$\sum_{j=1,j\neq i}^m \sum_{p,q}^n W_{pq}^j(\mathbf{Z}_p^i - \mathbf{Z}_q^i)^2 = \sum_{j=1,j\neq i}^m tr(\mathbf{Z}^i\mathbf{L}^j(\mathbf{Z}^i)^T), \tag{3}$$

where $\mathbf{L}^j = \mathbf{D}^j - \mathbf{W}^j$, $\mathbf{D}^j$ is a diagonal matrix with $D_{kk}^j = \sum_l W_{kl}^j$, $\mathbf{Z}_p^i$ and $\mathbf{Z}_q^i$ are the p-th and q-th column of $\mathbf{Z}^i$, respectively. Since Laplacian embedding is commonly considered to be able to preserve the local topology of the original data space [11, 17], $\sum_{j=1,j\neq i}^m tr(\mathbf{Z}^i\mathbf{L}^j(\mathbf{Z}^i)^T)$ is able to imply the local cross topological information, which means that $\mathbf{Z}^i$ obtained via (2) can hold the local topology structure of all other views. Through this type of constraint, multiple views are able to share their local structure information. And we compare it with the mutual independent local topology constraints in the experimental part to better illustrate its superiority. The construction of $\mathbf{W}$ will be discussed further in the experimental part.

The third term $||[\mathbf{Z}^1, \mathbf{Z}^2, \ldots, \mathbf{Z}^m]||_*$ reflects the global low-rank constraint. With the assumption that similar samples can be linearly reconstructed by each other, each representational matrix $\mathbf{Z}^i$ is expected to be low-rank and block diagonal. With this global low-rank constraint, we can maximize the correlation between representational matrices while encourage the low-rank and block diagonal properties of each of them.

3.2. Optimization

It is not straight forward to optimize (2) due to the trace norm. Hence we reformulate it with a variational formulation for trace norm [7] and develop an iterative minimization algorithm.

Lemma 1 [7]. *Let $\mathbf{Q} \in R^{n\times m}$. The trace norm of $\mathbf{Q}$ is equal to:*

$$||\mathbf{Q}||_* = \frac{1}{2}\inf_{\mathbf{S}\geq 0} tr(\mathbf{Q}^T\mathbf{S}^{-1}\mathbf{Q}) + tr(\mathbf{S}), \tag{4}$$

where the infimum is attained for $\mathbf{S} = (\mathbf{Q}\mathbf{Q}^T)^{\frac{1}{2}}$.
Based on this lemma, J_3 can be reformulated as

$$\begin{aligned} J_3 &= ||[\mathbf{Z}^1, \mathbf{Z}^2, \ldots, \mathbf{Z}^m]||_* \\ &= \frac{1}{2}\inf_{\mathbf{S}\geq 0} tr([\mathbf{Z}^1, \ldots, \mathbf{Z}^m]^T\mathbf{S}^{-1}[\mathbf{Z}^1, \ldots, \mathbf{Z}^m]) + tr(\mathbf{S}) \\ &= \frac{1}{2}\inf_{\mathbf{S}\geq 0}\sum_{i=1}^m tr((\mathbf{Z}^i)^T\mathbf{S}^{-1}\mathbf{Z}^i) + tr(\mathbf{S}). \end{aligned} \tag{5}$$

And the objective function (2) is equal to:

$$\begin{aligned} &\min_{\mathbf{Z}^1,\ldots,\mathbf{Z}^m}\min_{\mathbf{S}\geq 0} J = \\ &\quad \sum_{i=1}^m ||\mathbf{X}^i - \mathbf{X}^i\mathbf{Z}^i||_F^2 + \lambda\sum_{i,j=1;j\neq i}^m tr(\mathbf{Z}^i\mathbf{L}^j(\mathbf{Z}^i)^T) \\ &\quad + \frac{\beta}{2}\sum_{i=1}^m tr((\mathbf{Z}^i)^T\mathbf{S}^{-1}\mathbf{Z}^i) + \frac{\beta}{2}tr(\mathbf{S}) \end{aligned} \tag{6}$$

Methods	Accuracy(%)				
	Movies617	VOC	Wiki	Animal	3-Sources
Spectral-S	25.70 ± 1.13	80.88 ± 3.12	51.24 ± 3.60	27.21 ± 1.50	52.93 ± 3.59
Spectral-M	26.64 ± 1.14	84.70 ± 0.92	52.80 ± 2.83	28.23 ± 1.71	55.71 ± 5.35
Co-Pairwise	27.89 ± 1.64	69.95 ± 0.00	54.96 ± 1.53	31.65 ± 1.59	58.37 ± 3.28
Co-Centroid	28.57 ± 1.17	79.40 ± 0.57	56.52 ± 1.91	31.06 ± 2.02	58.93 ± 3.07
Co-Training	30.74 ± 1.28	84.88 ± 0.00	55.63 ± 1.56	30.35 ± 1.48	58.37 ± 3.47
Multi-NMF	26.99 ± 1.19	84.80 ± 3.80	56.70 ± 1.81	30.56 ± 1.02	68.40 ± 0.06
Multi-CF	29.60 ± 1.10	92.90 ± 0.00	57.52 ± 1.91	32.11 ± 1.86	69.23 ± 3.54
MSC-GS	32.78 ± 1.38	94.54 ± 0.00	58.95 ± 1.97	32.49 ± 1.01	71.36 ± 5.83
MSC-GL	**34.51 ± 1.17**	**96.17 ± 0.00**	**60.51 ± 2.02**	**33.62 ± 1.38**	**75.65 ± 6.60**

Table 1. Clustering results in terms of accuracy on five benchmark databases.

Methods	Normalized Mutual Information(%)				
	Movies617	VOC	Wiki	Animal	3-Sources
Spectral-S	25.47 ± 0.85	61.96 ± 3.43	54.25 ± 2.96	15.70 ± 0.65	53.38 ± 2.12
Spectral-M	27.19 ± 0.70	66.25 ± 1.41	55.81 ± 1.68	17.50 ± 1.12	53.83 ± 3.97
Co-Pairwise	28.04 ± 0.73	45.10 ± 0.00	53.82 ± 0.83	19.90 ± 1.51	62.25 ± 2.76
Co-Centroid	28.02 ± 0.72	55.12 ± 0.80	56.40 ± 0.52	18.50 ± 1.48	62.65 ± 2.51
Co-Training	30.74 ± 1.28	63.17 ± 0.00	56.23 ± 1.19	18.98 ± 0.73	63.15 ± 1.79
Multi-NMF	27.45 ± 0.55	67.18 ± 4.03	55.70 ± 0.95	18.77 ± 0.71	60.20 ± 0.06
Multi-CF	30.09 ± 1.32	79.23 ± 0.00	57.44 ± 0.66	21.25 ± 1.76	67.91 ± 4.41
MSC-GS	31.86 ± 0.98	79.72 ± 0.00	57.12 ± 1.23	19.42 ± 0.92	68.71 ± 4.47
MSC-GL	**33.33 ± 0.64**	**82.47 ± 0.00**	**57.80 ± 0.39**	**22.03 ± 1.33**	**70.23 ± 3.72**

Table 2. Clustering results in terms of normalized mutual information on five benchmark databases.

Thus we obtain an efficient algorithm through iterative optimizing over $\mathbf{S}$ and each $\mathbf{Z}^i$.

Step.1 For problem (6), by fixing each $\mathbf{Z}^i$, it is easy to solve $\mathbf{S}$ according to lemma.1.

$$\mathbf{S} = (\mathbf{Z}\mathbf{Z}^T + \mu\mathbf{I})^{\frac{1}{2}}, \mathbf{Z} = [\mathbf{Z}^1, \mathbf{Z}^2, \ldots, \mathbf{Z}^m]. \tag{7}$$

Here the small term $\mu\mathbf{I}$ ensures $\mathbf{S}$ is invertible and thus the infimum can be attained [7].

Step.2 Given $\mathbf{S}$, optimizing (6) over $\mathbf{Z}^i$ is equal to

$$\begin{aligned}\min_{\mathbf{Z}^i} \quad & ||\mathbf{X}^i - \mathbf{X}^i\mathbf{Z}^i||_F^2 + \lambda \sum_{j=1, j\neq i}^{m} tr(\mathbf{Z}^i\mathbf{L}^j(\mathbf{Z}^i)^T) \\ & + \frac{\beta}{2} tr((\mathbf{Z}^i)^T\mathbf{S}^{-1}\mathbf{Z}^i).\end{aligned} \tag{8}$$

Taking its partial derivatives with respect to $\mathbf{Z}^i$ and setting it to zero, we obtain the following equation

$$((\mathbf{X}^i)^T\mathbf{X}^i + \frac{\beta}{2}\mathbf{S}^{-1})\mathbf{Z}^i + \mathbf{Z}^i(\lambda \sum_{j=1, j\neq i}^{m} \mathbf{L}^j) = (\mathbf{X}^i)^T\mathbf{X}^i. \tag{9}$$

This is a standard Sylvester equation and has a unique solution [1]. As we can see, the computation of $\mathbf{Z}^i$ is independent of each other, so it can be paralleled easily. After getting the optimal value of each $\mathbf{Z}^i$, we simply run spectral clustering algorithm like [19] on them and adopt the best result as the final clustering result. The whole procedure is summarized in Algorithm 1.

Algorithm 1 : Iterative Algorithm for *MSC-GL* model

Input: Multi-view data $\mathbf{X}^i \in R^{d_i \times n}$ $(i = 1, \ldots, m)$, parameters λ and β, number of clusters c, $\mu = 10^{-5}$.

Output: $\mathbf{Z}^i \in R^{n\times n}, i = 1, \ldots, m.$

1: Set $t = 0$, initialize each $\mathbf{Z}^i(t)$ as zero matrices.
2: Calculate Laplace matrices $\mathbf{L}^i \in R^{n\times n}, i = 1, \ldots, m.$
3: **repeat**
4: Set $\mathbf{Z}(t) = [\mathbf{Z}^1(t), \ldots, \mathbf{Z}^m(t)]$,
5: Compute $\mathbf{V}\mathit{Diag}(\mathbf{s})\mathbf{V}^T$ as the eigenvalue decomposition of $(\mathbf{Z}(t)(\mathbf{Z}(t))^T + \mu\mathbf{I})$,
6: Update $\mathbf{S}^{-1}(t) = \mathbf{V}\mathit{Diag}(1./\sqrt{\mathbf{s}})\mathbf{V}^T$, $t = t + 1$,
7: Update $\mathbf{Z}^i(t)(i = 1, \ldots, m)$ according to (9),
8: **until** convergence

4. Experiments

In this section, we compare the proposed model, *MSC-GL*, with the start-of-the-art subspace clustering methods on five widely used benchmark databases.

4.1. Databases

Movies617 dataset[1] It consists of 617 movies with 17 labels extracted from IMDb. The two views are the 1878 keywords representation and the 1398 actors representation.

Pascal VOC2007 dataset [5] It consists of 20 categories with 9963 images in total. By removing images with multiple categories, we get 5649 images left. Furthermore, considering the computational cost of some compared methods, we select the first three categories for evaluation. And

[1]http://membres-lig.imag.fr/grimal/data.html.

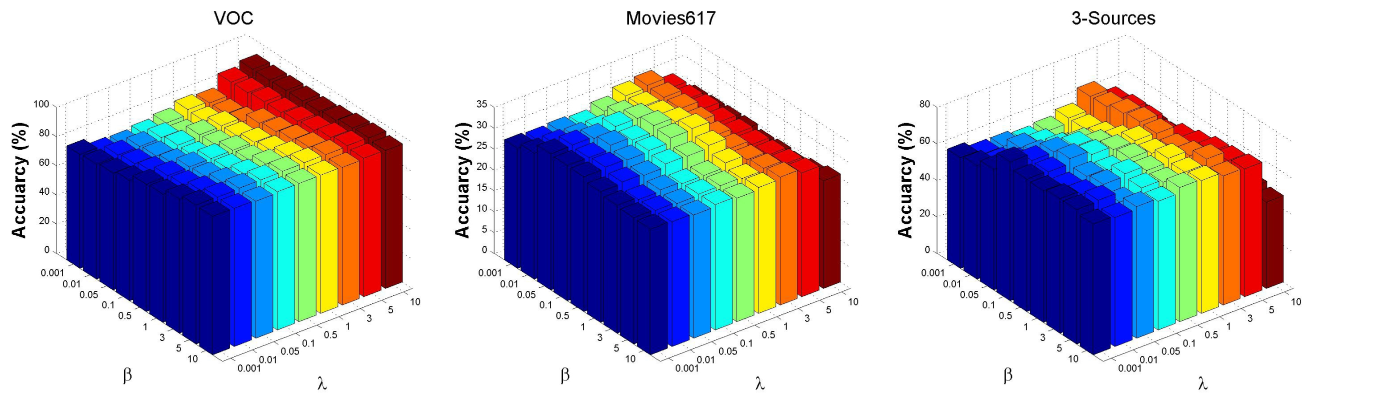

Figure 1. Accuracy of the proposed model vs. λ and β

we use 512-dimensional gist features and 399-dimensional word frequency features as the two view representations.

Wiki Text-image dataset [18] It consists of 2173/693 (training/ testing) image-text pairs from 10 categories. The two views are 10 dimensional latent Dirichlet allocation model based text features and 128 dimensional SIFT histogram image features. We randomly select 60 samples from each category for evaluation.

Animal dataset[2] It consists of 30475 images of 50 animals with six pre-extracted features. We make use of PyramidHOG (PHOG), colorSIFT and SURF features as three views. Besides, we select the first ten categories with each consists of randomly selected 50 samples as a subset for evaluation.

3-Sources Text dataset[3] It is collected from three well-known online news sources: BBC, Reuters and The Guardian. There are 416 distinct stories that manually categorized into six classes. Among them, 169 stories are reported in all the three sources and are used in our experiments as in [16].

4.2. Experimental Settings and Results

To evaluate the performance of the proposed *MSC-GL* model, we compare it with following algorithms. 1) **Spectral-S:** The spectral clustering method in [19] is used to cluster each view's data and the best result is selected. 2) **Spectral-M:** Concatenating the features of all the views and implementing spectral clustering method in [19] on this concatenated representation. 3) **Co-Pairwise, Co-Centroid:** Two co-regularization schemes on the eigenvectors of Laplacian matrices [13]. 4) **Co-Training:** Alternately modifying one view's graph structure using other views' information [12]. 5)**Multi-NMF:** A multi-view clustering method which bases on multi-view non-negative matrix factorization [16]. 6) **Multi-CF:** A multi-view clustering and feature learning framework that bases on structure sparsity [23].

Two commonly used measures, clustering accuracy and normalized mutual information (NMI) [4] are used to measure the clustering performance. As k-means is used in all the experiments, it is run 20 times with random initialization and both mean values and standard derivations are reported. For all databases, we tune the parameters of the compared methods manually to obtain their best results. For the proposed *MSC-GL* model, we construct an undirected k nearest neighbor graph under each view with Gaussian kernel [22] and obtain the Laplacian matrices based on these graphs. k is empirically set to be 5 on all the databases. In addition, we provide the results of the *MSC-GS* model, which uses the local topology information in each view independently for better comparison. It can be easily implemented by substituting $\sum_{i=1}^{m} tr(\mathbf{Z}^i\mathbf{L}^i(\mathbf{Z}^i)^T)$ for $\sum_{i,j=1;j\neq i}^{m} tr(\mathbf{Z}^i\mathbf{L}^j(\mathbf{Z}^i)^T)$ in function (2).

Numerical results are shown in Table.1 and Table.2. It is obvious that *MSC-GL* outperforms all the compared methods, thus the global and local constraints we developed are rational and efficient. More specifically, the proposed local cross topology preserving constraints are superior to the mutual independent topology constraints according to the results of *MSC-GL* and *MSC-GS*. Furthermore, to better illustrate the influence of the parameters, we report the clustering accuracy as a function of λ and β in Figure.1. Due to space limited, only accuracy results on VOC, Movies617 and 3-Sources are reported, and similar results can be revealed on other databases and measures. We can observe that both of the parameters are important and there exists a large parameter space so that our model surpasses other methods.

5. Conclusion

This paper has developed a novel multi-view subspace clustering method that explicitly leverages both global and

[2]http://attributes.kyb.tuebingen.mpg.de/.
[3]http://mlg.ucd.ie/datasets/3sources.html.

local structure information. By introducing the global low-rank constraint and local cross topology preserving constraints, we can make better use of the complementary information provided by multi-view data. The experimental results on several widely used databases have demonstrated that the proposed method significant improves the performance of multi-view clustering.

Acknowledgement

This work is founded by the National Natural Science Foundation of China (Grant No.61473289,61135002) and the Youth Innovation Promotion Association of the Chinese Academy of Sciences (Grant No.2015190).

References

[1] R. H. Bartels and G. Stewart. Solution of the matrix equation. $AX + XB = C$. *Communications of the ACM*, 15(9):820–826, 1972. 3

[2] E. Bruno and S. Marchand-Maillet. Multiview clustering: A late fusion approach using latent models. In *ACM Conference on Research and Development in Information Retrieval*, pages 736–737. ACM, 2009. 1

[3] B. Chen and J. C. Príncipe. Maximum correntropy estimation is a smoothed map estimation. *IEEE Signal Processing Letters*, 19(8):491–494, 2012. 2

[4] W.-Y. Chen, Y. Song, H. Bai, C.-J. Lin, and E. Y. Chang. Parallel spectral clustering in distributed systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):568–586, 2011. 4

[5] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng. Nus-wide: a real-world web image database from national university of singapore. In *ACM Conference on Image and Video Retrieval*, page 48, 2009. 3

[6] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, 2013. 2

[7] E. Grave, G. R. Obozinski, and F. R. Bach. Trace lasso: a trace norm regularization for correlated designs. In *Advances in Neural Information Processing Systems*, pages 2187–2195, 2011. 2, 3

[8] D. Greene and P. Cunningham. A matrix factorization approach for integrating multiple data views. In *Machine Learning and Knowledge Discovery in Databases*, pages 423–438. 2009. 1

[9] R. He, Z. Sun, T. Tan, and W.-S. Zheng. Recovery of corrupted low-rank matrices via half-quadratic based nonconvex minimization. In *Computer Vision and Pattern Recognition*, pages 2889–2896, 2011. 1

[10] R. He, T. Tan, and L. Wang. Robust recovery of corrupted low-rankmatrix by implicit regularizers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(4):770–783, 2014. 1

[11] H. Hu, Z. Lin, J. Feng, and J. Zhou. Smooth representation clustering. In *Computer Vision and Pattern Recognition*, pages 3834–3841, 2014. 1, 2

[12] A. Kumar and H. Daumé. A co-training approach for multi-view spectral clustering. In *International Conference on Machine Learning*, pages 393–400, 2011. 1, 2, 4

[13] A. Kumar, P. Rai, and H. Daume. Co-regularized multi-view spectral clustering. In *Advances in Neural Information Processing Systems*, pages 1413–1421, 2011. 1, 2, 4

[14] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013. 1, 2

[15] G. Liu, Z. Lin, and Y. Yu. Robust subspace segmentation by low-rank representation. In *International Conference on Machine Learning*, pages 663–670, 2010. 2

[16] J. Liu, C. Wang, J. Gao, and J. Han. Multi-view clustering via joint nonnegative matrix factorization. In *SIAM Data Mining Conference*, pages 252–260, 2013. 1, 4

[17] W. Liu, J. Wang, S. Kumar, and S.-F. Chang. Hashing with graphs. In *International Conference on Machine Learning*, pages 1–8, 2011. 2

[18] D. Pracner, N. Tomašev, M. Radovanovic, D. Mladenic, and M. Ivanovic. Wikimage: Correlated image and text datasets. In *International Multiconference on Information Society*, pages 141–144, 2011. 4

[19] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000. 2, 3, 4

[20] S. Sun. A survey of multi-view machine learning. *Neural Computing and Applications*, 23(7-8):2031–2038, 2013. 1

[21] R. Vidal. Subspace clustering. *IEEE Signal Processing Magazine*, 2(28):52–68, 2011. 1

[22] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007. 2, 4

[23] H. Wang, F. Nie, and H. Huang. Multi-view clustering and feature learning via structured sparsity. In *International Conference on Machine Learning*, pages 352–360, 2013. 1, 4

[24] K. Wang, R. He, W. Wang, L. Wang, and T. Tan. Learning coupled feature spaces for cross-modal matching. In *International Conference on Computer Vision*, pages 2088–2095, 2013. 1

[25] Y.-X. Wang, H. Xu, and C. Leng. Provable subspace clustering: When lrr meets ssc. In *Advances in Neural Information Processing Systems*, pages 64–72, 2013. 2

[26] X. Zhao, N. Evans, and J.-L. Dugelay. A subspace co-training framework for multi-view clustering. *Pattern Recognition Letters*, 41:73–82, 2014. 1