From Ensemble Clustering to Multi-View Clustering

Zhiqiang Tao[1], **Hongfu Liu**[1], **Sheng Li**[1], **Zhengming Ding**[1] and **Yun Fu**[1,2]
[1]Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA
[2]College of Computer and Information Science, Northeastern University, Boston, MA 02115, USA
{zqtao, hfliu, shengli, allanding, yunfu}@ece.neu.edu

Abstract

Multi-View Clustering (MVC) aims to find the cluster structure shared by multiple views of a particular dataset. Existing MVC methods mainly integrate the raw data from different views, while ignoring the high-level information. Thus, their performance may degrade due to the conflict between heterogeneous features and the noises existing in each individual view. To overcome this problem, we propose a novel Multi-View Ensemble Clustering (MVEC) framework to solve MVC in an Ensemble Clustering (EC) way, which generates Basic Partitions (BPs) for each view individually and seeks for a consensus partition among all the BPs. By this means, we naturally leverage the complementary information of multi-view data in the same partition space. Instead of directly fusing BPs, we employ the low-rank and sparse decomposition to explicitly consider the connection between different views and detect the noises in each view. Moreover, the spectral ensemble clustering task is also involved by our framework with a carefully designed constraint, making MVEC a unified optimization framework to achieve the final consensus partition. Experimental results on six real-world datasets show the efficacy of our approach compared with both MVC and EC methods.

1 Introduction

Multi-View learning benefits from leveraging the complementary information from multiple views of a particular dataset, where these views can be obtained by various sensors or represented with different descriptors. For example, we may capture human activity from RGB video cameras, depth cameras, and on-body sensors [Li *et al.*, 2016]; for vision tasks, images could be encoded by host of hand-crafted and deep features. Some interesting multi-view problems include subspace learning [Ding and Fu, 2014], outlier detection [Li *et al.*, 2015a; Zhao and Fu, 2015], cross-domain learning [Wang *et al.*, 2017], and incomplete multi-view case [Zhao *et al.*, 2016]. In this paper, we focus on the Multi-View Clustering (MVC) problem. Considerable research efforts have been made to solve the MVC problem, such as optimizing certain loss function with concatenated multi-view features [Bickel and Scheffer, 2004; Kumar and III, 2011], conducting traditional clustering on a common low-dimensional latent subspace [Chaudhuri *et al.*, 2009; Xia *et al.*, 2014; Wang *et al.*, 2016; Zhao *et al.*, 2017], and late fusion approaches [Greene and Cunningham, 2009; Bruno and Marchand-Maillet, 2009].

The key problem for MVC is how to integrate the complementary information from multiple views. For instance, [Kumar *et al.*, 2011] enhanced the similarity of eigenvectors learnt from different views and integrated it within a spectral clustering framework; [Liu *et al.*, 2013b] designed a joint matrix factorization method and sought for a compatible solution of multi-view data. However, the noises of each single-view data can seriously degrade the performance of these pioneering works. In light of this, [Xia *et al.*, 2014] learned a consensus affinity graph among multiple views and handled the view-specific noises via a low-rank and sparse decomposition framework. Along this line, [Wang *et al.*, 2016] captured the local manifold structure in each view and achieved the clustering agreement by minimizing the difference across views. All these MVC methods directly integrate the raw multi-view data, which, however, is not an easy task due to the distinct gap among heterogeneous feature spaces. Thus, one promising way (*i.e.* late fusion methods) for solving MVC is to first transform multi-view data into the same partition space [Greene and Cunningham, 2009; Bruno and Marchand-Maillet, 2009], and then obtain the clustering result in an ensemble clustering manner.

Ensemble Clustering (EC) [Strehl and Ghosh, 2003; Fred and Jain, 2005] methods take as input a set of Basic Partitions (BPs) and integrate multiple BPs into a consensus one. Thus, it naturally has the ability of leveraging complementary information from heterogeneous sources [Wu *et al.*, 2015]. Nevertheless, ensemble clustering draws little attention in the field of multi-view clustering, and it neglects to explore the connection among different views. Moreover, most existing EC methods also suffer from the noises in the multi-view BPs, which could be caused by the intra-view or inter-view disagreement among all the BPs [Tao *et al.*, 2016].

To address the above challenges, we propose a novel Multi-View Ensemble Clustering (MVEC) algorithm in this paper. Specifically, we first generate a group of BPs for each single view, and summarize each group as a view-specific co-association matrix [Fred and Jain, 2005], which works as a pairwise affinity matrix upon the categorical data. After that, we employ low-rank and sparse decomposition to seek for the consensus affinity matrix shared by all the views and compensate the disagreement among all the BPs. Meanwhile, the spectral ensemble clustering task [Liu *et al.*, 2015; 2017] is also involved by our MVEC framework with a carefully designed constraint, leading to a unified optimization framework to jointly integrate multi-view information and find the consensus partition.

MVEC is inspired by two previous works [Xia *et al.*, 2014; Tao *et al.*, 2016]. Compared with [Xia *et al.*, 2014], we solve MVC in a partition space rather than using raw features, since it is more reasonable to uncover the cluster structure shared by multi-view co-association matrices. Moreover, different from [Xia *et al.*, 2014] that only learns a low-rank matrix, our approach simultaneously performs the tasks of low-rank representation learning and spectral graph partitioning. On the other side, [Tao *et al.*, 2016] focuses on the single-view case and thus cannot make full use of the multi-view data. In contrast, our MVEC explicitly utilizes the connection between different views and employs the learned consensus partition to iteratively enhance the cluster structure of each view.

The contributions of this work are summarized as follows: (1) A general Multi-View Ensemble Clustering (MVEC) framework is proposed, which effectively exploits the multi-view basic partitions to solve the MVC problem. (2) We jointly learn the view-consensus affinity matrix and find the consensus partition within a unified optimization framework. (3) A novel *self-boost* constraint is designed to iteratively improve the clustering performance.

2 Multi-View Ensemble Clustering

2.1 Problem Formulation

Given a set of n data points with m views (*i.e.*, feature representations or modalities), we denote the dataset of each view as $\mathcal{X}^{(v)} = \{x_1^{(v)}, \dots, x_n^{(v)}\}$, $1 \le v \le m$. For $\forall v$, we assume $\mathcal{X}^{(v)}$ is sampled from K crispy clusters, denoted as $\mathcal{C} = \{C_1, \dots, C_K\}$. Let $\Pi^{(v)} = \{\pi_1^{(v)}, \dots, \pi_r^{(v)}\}$ be a group of r basic partitions (BPs) for $\mathcal{X}^{(v)}$, where each BP $\pi_i^{(v)}$ partitions $\mathcal{X}^{(v)}$ into K_i clusters, *i.e.*, $\pi_i^{(v)} = \{\pi_i^{(v)}(x_1^{(v)}), \dots, \pi_i^{(v)}(x_n^{(v)})\}$ is a set of categorical data, $1 \le \pi_i^{(v)}(x_j^{(v)}) \le K_i$, $1 \le i \le r$, and $1 \le j \le n$. For each view, we obtain $\Pi^{(v)}$ by using the Random Parameter Selection (RPS) strategy [Fred and Jain, 2005; Wu *et al.*, 2015], which performs K-means algorithm on $\mathcal{X}^{(v)}$ r times with varying cluster number from K to $\sqrt{n}$. It has been shown that, RPS can capture various cluster structures existing in the real-world datasets [Fred and Jain, 2005].

In this paper, we aim to solve multi-view clustering in an ensemble clustering way. By following [Fred and Jain, 2005; Liu *et al.*, 2015], each $\Pi^{(v)}$ is summarized as a co-association matrix $\mathbf{S}^{(v)} \in \mathbb{R}^{n \times n}$:

$$\mathbf{S}^{(v)}(x_p^{(v)}, x_q^{(v)}) = \frac{1}{r} \sum_{i=1}^{r} \delta(\pi_i^{(v)}(x_p^{(v)}), \pi_i^{(v)}(x_q^{(v)})), \tag{1}$$

where $x_p^{(v)}, x_q^{(v)} \in \mathcal{X}^{(v)}$ and $\delta(a, b) = 1$ if $a = b$; 0 otherwise. By this means, we actually compute a pairwise affinity graph for each view upon the partition space, and thus, we can naturally conduct graph partitioning on $\mathbf{S}^{(v)}$ to find a consensus clustering result. In particular, each view could be solved by spectral ensemble clustering [Liu *et al.*, 2015]:

$$\min_{\mathbf{H}} \text{tr}(\mathbf{H}^{\text{T}} \mathbf{L}_s^{(v)} \mathbf{H}) \text{ s.t. } \mathbf{H}^{\text{T}} \mathbf{H} = \mathbf{I}, \tag{2}$$

where $\mathbf{H} \in \mathbb{R}^{n \times K}$ is the scaled partition matrix that represents the cluster membership of all the data points: $\mathbf{H}_{jk} = 1/\sqrt{|C_k|}$ if $x_j \in C_k$; 0 otherwise, and $\mathbf{L}_s^{(v)}$ is the normalized Laplacian matrix [Ng *et al.*, 2001] of the co-association matrix in the v-th view.

However, Eq. (2) only tackles each view individually, yet without utilizing the complementary information between different views. There are still two challenging problems of employing the co-association matrices from multiple views for the clustering aim: *(1) How do we seek a consensus co-association matrix to identify the underlying cluster structure shared by multi-view data? (2) How can we capture the disagreements among multiple BPs of intra-view or inter-view?*

To address the above challenges, we propose to learn a common representation shared by multi-view co-association matrices via low-rank and sparse matrix decomposition [Ye *et al.*, 2012; Xia *et al.*, 2014]. Moreover, we jointly perform the tasks of spectral graph partitioning and low-rank representation learning. Our Multi-View Ensemble Clustering (MVEC) algorithm is formulated as:

$$\begin{aligned} \min_{\mathbf{H}, \mathbf{Z}, \mathbf{E}^{(v)}} \; & \text{tr}(\mathbf{H}^{\text{T}} \mathbf{L}_z \mathbf{H}) + \lambda_1 \|\mathbf{Z}\|_* + \lambda_2 \sum_{v=1}^{m} \|\mathbf{E}^{(v)}\|_1 \\ \text{s.t. } & \mathbf{H}^{\text{T}} \mathbf{H} = \mathbf{I}, \forall v, \mathbf{S}^{(v)} + \mathbf{H}\mathbf{H}^{\text{T}} = \mathbf{S}^{(v)} \mathbf{Z} + \mathbf{E}^{(v)}, \\ & \mathbf{Z} \ge 0, \mathbf{Z}\mathbf{1} = \mathbf{1}, \end{aligned} \tag{3}$$

where $\mathbf{H} \in \mathbb{R}^{n \times K}$ represents the consensus partition, $\mathbf{Z} \in \mathbb{R}^{n \times n}$ is the low-rank representation, $\mathbf{E}^{(v)} \in \mathbb{R}^{n \times n}$ captures noises for the v^{th} view, $\mathbf{1} \in \mathbb{R}^n$ is the vector of all ones, and $\lambda_1, \lambda_2 > 0$ are two balancing parameters. $\mathbf{L}_z = \mathbf{D}_z - \mathbf{Z}$ is the Laplacian matrix [Ng *et al.*, 2001] of $\mathbf{Z}$, and $\mathbf{D}_z$ is a diagonal matrix consisting of the sum of each row in $\mathbf{Z}$. By following [Liu *et al.*, 2013a], we employ the nuclear norm $\|\mathbf{Z}\|_*$ to measure the matrix rank, and the ℓ_1 norm $\|\mathbf{E}^{(v)}\|_1$ to characterize the sparseness.

Considering that, the co-association matrices of all the views share with the same cluster structure, and the rank of $\mathbf{S}^{(v)}$ ideally equals to the cluster number K ($K \ll n$), we seek for a low-rank representation $\mathbf{Z}$ to reveal the membership between data points through all the views. Meanwhile, to alleviate the "conflict" among different views and handle the outliers existing in co-association matrix, we learn a sparse error matrix $\mathbf{E}^{(v)}$ for each single view. Hence, we may take $\mathbf{Z}$ as a consensus pairwise affinity matrix, and find the consensus partition $\mathbf{H}$ on $\mathbf{Z}$ by spectral clustering.

Taking a close look at Eq. (3), we develop a *self-boost* constraint to iteratively enhance the original co-association matrix of each view, *i.e.*, $\mathbf{S}^{(v)} + \mathbf{H}\mathbf{H}^{\text{T}} = \mathbf{S}^{(v)} \mathbf{Z} + \mathbf{E}^{(v)}$, where $\mathbf{H}\mathbf{H}^{\text{T}}$ enjoys a clear cluster structure. By using this carefully designed constraint, we first find a high quality consensus partition $\mathbf{H}$ from $\mathbf{Z}$, and then in return, $\mathbf{H}$ is leveraged to better guide the learning of $\mathbf{Z}$. Besides, we also add the probabilistic simplex constraint ($\mathbf{Z} \geq 0$, $\mathbf{Z1} = \mathbf{1}$) [Duchi *et al.*, 2008] to guarantee the probability property of $\mathbf{Z}$.

2.2 Optimization

A unified optimization framework of three variable groups is provided by Eq. (3), which is convex with respect to each variable by keeping the others fixed. Thus, we can divide it into several subproblems, and solve them in an iterative way. In details, we apply the Augmented Lagrange Multiplier (ALM) algorithm [Lin *et al.*, 2011] to address our MVEC problem. To facilitate the optimization process, we introduce an auxiliary variable $\mathbf{J} \in \mathbb{R}^{n\times n}$ with $\mathbf{Z} = \mathbf{J}$ to make Eq. (3) separable. Then, our problem could be equivalently converted as:

$$
\begin{aligned}
&\min_{\mathbf{H}^{\mathrm{T}}\mathbf{H}=\mathbf{I},\mathbf{J},\mathbf{Z},\mathbf{E}^{(v)}} \mathrm{tr}(\mathbf{H}^{\mathrm{T}}\mathbf{L}_z\mathbf{H}) + \lambda_1\|\mathbf{J}\|_* + \lambda_2\sum_{v=1}^{m}\|\mathbf{E}^{(v)}\|_1 \\
&\text{s.t. } \forall v, \mathbf{S}^{(v)} + \mathbf{H}\mathbf{H}^{\mathrm{T}} = \mathbf{S}^{(v)}\mathbf{Z} + \mathbf{E}^{(v)}, \mathbf{Z} = \mathbf{J}, \\
&\quad \mathbf{Z} \geq 0, \mathbf{Z1} = \mathbf{1}.
\end{aligned} \tag{4}
$$

The augmented Lagrange function of Eq. (4) is written as:

$$
\begin{aligned}
\mathcal{L} =& \mathrm{tr}(\mathbf{H}^{\mathrm{T}}\mathbf{L}_z\mathbf{H}) + \lambda_1\|\mathbf{J}\|_* + \lambda_2\sum_{v=1}^{m}\|\mathbf{E}^{(v)}\|_1 \\
&+ \sum_{v=1}^{m}\Phi(\mathbf{S}^{(v)} + \mathbf{H}\mathbf{H}^{\mathrm{T}} - \mathbf{S}^{(v)}\mathbf{Z} - \mathbf{E}^{(v)}, \mathbf{Y}^{(v)}) \\
&+ \Phi(\mathbf{Z} - \mathbf{J}, \mathbf{\Lambda}) + \langle \mathbf{Z1} - \mathbf{1}, \mathbf{w}\rangle + \frac{\mu}{2}\|\mathbf{Z1} - \mathbf{1}\|_2^2,
\end{aligned} \tag{5}
$$

where $\mathbf{Z} \geq 0$, $\forall v$ $\mathbf{Y}^{(v)} \in \mathbb{R}^{n\times n}$, $\mathbf{\Lambda} \in \mathbb{R}^{n\times n}$, and $\mathbf{w} \in \mathbb{R}^n$ are Lagrange multipliers, $\Phi(\mathbf{A}, \mathbf{B}) \equiv \langle \mathbf{A}, \mathbf{B}\rangle + \frac{\mu}{2}\|\mathbf{A}\|_{\mathrm{F}}^2$, and $\mu > 0$ is a penalty parameter. In the next, we will give the details of iteratively solving $\mathbf{J}$, $\mathbf{Z}$, $\mathbf{E}^{(v)}$ and $\mathbf{H}$ in sequence.

Subproblem of J. Solving $\mathcal{L}$ w.r.t $\mathbf{J}$ is equivalent to:

$$
\min_{\mathbf{J}} \frac{\lambda_1}{\mu_{(t)}}\|\mathbf{J}\|_* + \frac{1}{2}\|\mathbf{J} - (\mathbf{Z}_{(t)} + \frac{\mathbf{\Lambda}_{(t)}}{\mu_{(t)}})\|_{\mathrm{F}}^2. \tag{6}
$$

Following the previous work [Liu *et al.*, 2013a], Eq. (6) could be effectively solved by a closed-form solution:

$$
\mathbf{J}_{(t+1)} = \mathcal{S}_{\frac{\lambda_1}{\mu_{(t)}}}(\mathbf{Z}_{(t)} + \frac{\mathbf{\Lambda}_{(t)}}{\mu_{(t)}}), \tag{7}
$$

where $\mathcal{S}(\cdot)$ is the Singular Value Threshold (SVT) operator [Cai *et al.*, 2010].

Subproblem of Z. Generally, we obtain the solution of $\mathbf{Z}_{(t+1)}$ by taking derivate of $\mathcal{L}$ w.r.t. $\mathbf{Z}$ and setting it as zero. However, it is not straightforward to solve $\mathrm{tr}(\mathbf{H}^{\mathrm{T}}\mathbf{L}_z\mathbf{H})$ in a matrix form, since $\mathbf{L}_z = \mathbf{D}_z - \mathbf{Z}$ and $\mathbf{D}_z$ is the degree matrix of $\mathbf{Z}$. Thus, to simplify this term, we introduce an auxiliary matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ as:

$$
\mathbf{P} = \left[\mathbf{P}_1 \ldots \mathbf{P}_j \ldots \mathbf{P}_n\right], \ \mathbf{P}_j = \begin{bmatrix} \|\mathbf{H}_1 - \mathbf{H}_j\|_2^2 \\ \vdots \\ \|\mathbf{H}_n - \mathbf{H}_j\|_2^2 \end{bmatrix}, \tag{8}
$$

where $\mathbf{H}_j$ is the j^{th} row vector in $\mathbf{H}$. By utilizing the property of Laplacian matrix and Eq. (8), we have the following simple deduction:

$$
\mathrm{tr}(\mathbf{H}^{\mathrm{T}}\mathbf{L}_z\mathbf{H}) = \frac{1}{2}\sum_{i,j}^{n}\|\mathbf{H}_i - \mathbf{H}_j\|^2\mathbf{Z}_{ij} = \frac{1}{2}\mathrm{tr}(\mathbf{P}^{\mathrm{T}}\mathbf{Z}).
$$

***Algorithm 1.** Multi-View Ensemble Clustering by ALM*

Input: Basic partitions sets of m views $\Pi^{(1)}, \ldots, \Pi^{(m)}$, cluster number K, two parameters $\lambda_1, \lambda_2 > 0$.

Initial: $\mathbf{J}_{(0)} = \mathbf{Z}_{(0)} = \mathbf{\Lambda}_{(0)} = \mathbf{0} \in \mathbb{R}^{n\times n}$,
$\mathbf{E}_{(0)}^{(v)} = \mathbf{Y}_{(0)}^{(v)} = \mathbf{0} \in \mathbb{R}^{n\times n}$, $v = 1, \ldots, m$,
$\mathbf{H}_{(0)} = \mathbf{0} \in \mathbb{R}^{n\times K}$, $\mathbf{w}_{(0)} = \mathbf{0} \in \mathbb{R}^{n\times 1}$, $\rho > 1$
$\mu_{(0)} = 10^{-3}$, $\mu_{\max} = 10^{10}$, $\epsilon = 10^{-4}$, $t = 0$.

Derive $\mathbf{S}^{(v)}$ from $\Pi^{(v)}$ via Eq. (1) for each view;
while *not converged* **do**
  Update $\mathbf{J}_{(t+1)}$ via Eq. (7);
  Update $\mathbf{Z}_{(t+1)}$ via Eq. (11);
  Update $\mathbf{E}_{(t+1)}$ via Eq. (13);
  Update $\mathbf{L}_z$ and $\mathbf{D}_z$ by $\mathbf{Z}_{(t+1)}$;
  Set $\mathbf{H}_{(t+1)}$ as the K smallest eigenvectors of $\mathbf{L}_z$;
  Update the Lagrangian multipliers via Eq. (15);
  Check the convergence condition:
  $(\max\{\|\mathbf{\Delta}(t+1)^{(1)}\|_\infty, \ldots, \|\mathbf{\Delta}(t+1)^{(m)}\|_\infty\} < \epsilon) \wedge$
  $(\|\mathbf{J}^{(t+1)} - \mathbf{Z}^{(t+1)}\|_\infty < \epsilon) \wedge (\|\mathbf{Z}^{(t+1)}\mathbf{1} - \mathbf{1}\|_\infty < \epsilon)$;
  $\mu_{(t+1)} = \min\{\rho\mu_{(t)}, \mu_{\max}\}$, $t = t + 1$;
end while
Find the final partition π by running K-means on $\mathbf{H}$ or spectral clustering on $\mathbf{Z}$.

Output: The final clustering result π.

Then, we can solve $\mathbf{Z}_{(t+1)}$ via the following equivalent formulation as:

$$
\begin{aligned}
\min_{\mathbf{Z}\geq 0} &\frac{1}{2\mu_{(t)}}\mathrm{tr}(\mathbf{P}_{(t)}^{\mathrm{T}}\mathbf{Z}) + \frac{1}{2}\sum_{v=1}^{m}\|\mathbf{S}^{(v)} + \mathbf{H}_{(t)}\mathbf{H}_{(t)}^{\mathrm{T}} - \mathbf{S}^{(v)}\mathbf{Z} \\
&- \mathbf{E}_{(t)}^{(v)} + \mathbf{Y}_{(t)}^{(v)}/\mu_{(t)}\|_{\mathrm{F}}^2 + \frac{1}{2}\|\mathbf{Z} - \mathbf{J}_{(t)} + \mathbf{\Lambda}_{(t)}/\mu_{(t)}\|_{\mathrm{F}}^2 \\
&+ \frac{1}{2}\|\mathbf{Z1} - \mathbf{1} + \mathbf{w}_{(t)}/\mu_{(t)}\|_2^2.
\end{aligned} \tag{9}
$$

Inspired by [Zhuang *et al.*, 2016; Lin *et al.*, 2011], we linearize the Eq. (9) at $\mathbf{Z}_{(t)}$ as:

$$
\min_{\mathbf{Z}\geq 0} \langle \mathbf{Z} - \mathbf{Z}_{(t)}, \mathbf{F}_{(t)}\rangle + \frac{1}{\eta_{(t)}}\|\mathbf{Z} - \mathbf{Z}_{(t)}\|_{\mathrm{F}}^2, \tag{10}
$$

where $\eta_{(t)} = \|\mathbf{P}_{(t)}\|_2^2 + \sum_v \|\mathbf{S}^{(v)}\|_2^2 + 1 + \|\mathbf{1}\|_2^2$, and $\mathbf{F}_{(t)}$ equals:

$$
\begin{aligned}
&\frac{\mathbf{P}_{(t)}}{2\mu_{(t)}} + \sum_{v=1}^{m}\mathbf{S}^{(v)\mathrm{T}}(\mathbf{S}^{(v)}\mathbf{Z}_{(t)} + \mathbf{E}_{(t)}^{(v)} - \mathbf{S}^{(v)} - \mathbf{H}_{(t)}\mathbf{H}_{(t)}^{\mathrm{T}} - \\
&\frac{\mathbf{Y}_{(t)}^{(v)}}{\mu_{(t)}}) + (\mathbf{Z}_{(t)} - \mathbf{J}_{(t+1)} + \frac{\mathbf{\Lambda}_{(t)}}{\mu_{(t)}}) + (\mathbf{Z}_{(t)}\mathbf{1} - \mathbf{1} + \frac{\mathbf{w}_{(t)}}{\mu_{(t)}})\mathbf{1}^{\mathrm{T}}.
\end{aligned}
$$

Then, the solution of $\mathbf{Z}_{(t+1)}$ is given by:

$$
\begin{aligned}
\mathbf{Z}_{(t+1)} &= \operatorname*{argmin}_{\mathbf{Z}\geq 0}\|\mathbf{Z} - (\mathbf{Z}_{(t)} - \frac{1}{\eta_{(t)}}\mathbf{F}_{(t)})\|_{\mathrm{F}}^2 \\
&= (\mathbf{Z}_{(t)} - \eta_{(t)}^{-1}\mathbf{F}_{(t)})_+.
\end{aligned} \tag{11}
$$

Subproblem of $\mathbf{E}^{(v)}$. For each view v, we update $\mathbf{E}_{(t+1)}^{(v)}$ by:

$$
\begin{aligned}
\min_{\mathbf{E}^{(v)}} &\frac{\lambda_2}{\mu_{(t)}}\|\mathbf{E}^{(v)}\|_1 + \frac{1}{2}\|\mathbf{E}^{(v)} - (\mathbf{S}^{(v)} + \mathbf{H}_{(t)}\mathbf{H}_{(t)}^{\mathrm{T}} \\
&- \mathbf{S}^{(v)}\mathbf{Z}_{(t+1)} + \mathbf{Y}_{(t)}^{(v)}/\mu_{(t)})\|_{\mathrm{F}}^2.
\end{aligned} \tag{12}
$$

Following [Lin *et al.*, 2011], we have:

$$
\mathbf{E}_{(t+1)}^{(v)} = \mathcal{D}_{\frac{\lambda_2}{\mu_{(t)}}}(\mathbf{S}^{(v)} + \mathbf{H}_{(t)}\mathbf{H}_{(t)}^{\mathrm{T}} - \mathbf{S}^{(v)}\mathbf{Z}_{(t+1)} + \mathbf{Y}_{(t)}^{(v)}/\mu_{(t)}), \tag{13}
$$

where $\mathcal{D}(\cdot)$ denotes the element-wise soft-thresholding shrinkage operator [Lin *et al.*, 2011].

Subproblem of H. Clearly, there are two parts w.r.t $\mathbf{H}$ in Eq. (5), which are corresponding to spectral ensemble clustering and our *self-boost* constraint, respectively. Recall that, we devise this constraint to involve an interaction between learning $\mathbf{Z}$ and finding $\mathbf{H}$. As shown by Eq. (11), the term of $\mathbf{H}\mathbf{H}^{\mathrm{T}}$ can effectively hold the cluster structure of $\mathbf{Z}_{(t+1)}$. However, when computing $\mathbf{H}$ from $\mathbf{Z}$, considering this term will take some unnecessary "noises" induced by $\mathbf{E}^{(v)}$ and $\mathbf{Y}^{(v)}$ into the clustering process. Hence, we omit the part of $\mathcal{L}$ that contains $\mathbf{H}\mathbf{H}^{\mathrm{T}}$, and recast the subproblem of $\mathbf{H}$ as:

$$\mathbf{H}_{(t+1)} = \underset{\mathbf{H}^{\mathrm{T}}\mathbf{H}=\mathbf{I}}{\operatorname{argmin}} \operatorname{tr}(\mathbf{H}^{\mathrm{T}}\mathbf{L}_z\mathbf{H}), \tag{14}$$

where $\mathbf{L}_z$ is updated by $\mathbf{Z}_{(t+1)}$. As suggested by [Zha *et al.*, 2002; Dhillon *et al.*, 2004], Eq. (14) is solved by directly setting $\mathbf{H}_{(t+1)}$ as the K smallest eigenvectors of $\mathbf{L}_z$.

Multipliers. Totally, we have $m+2$ multipliers, which can be updated by:

$$\begin{aligned}
&\mathbf{\Delta}^{(v)}_{(t+1)} = \mathbf{S}^{(v)} + \mathbf{H}_{(t+1)}\mathbf{H}^{\mathrm{T}}_{(t+1)} - \mathbf{S}^{(v)}\mathbf{Z}_{(t+1)} - \mathbf{E}^{(v)}_{(t+1)},\\
&\mathbf{Y}^{(v)}_{(t+1)} = \mathbf{Y}^{(v)}_{(t)} + \mu_{(t)}\Delta^{(v)}, \forall v = 1, \dots, m.\\
&\mathbf{\Lambda}_{(t+1)} = \mathbf{\Lambda}_{(t)} + \mu_{(t)}(\mathbf{Z}_{(t+1)} - \mathbf{J}_{(t+1)}),\\
&\mathbf{w}_{(t+1)} = \mathbf{w}_{(t)} + \mu_{(t)}(\mathbf{Z}_{(t+1)}\mathbf{1} - \mathbf{1}),
\end{aligned} \tag{15}$$

where $\mathbf{\Delta}^{(v)}$ is introduced for the conciseness. The entire solution for MVEC is summarized in Algorithm 1.

2.3 Discussion

Convergence Analysis. As shown by Algorithm 1, the proposed MVEC can be divided into four subproblems, each of which is convex w.r.t one variable and solved with a closed-form solution. Thus, by finding the optimal solution for each subproblem alternatively, our algorithm at least converges to a local minimum solution.

Complexity Analysis. The major computation of Algorithm 1 lies at step 3-5 and step 7. In details, step 3 costs $\mathcal{O}(n^3)$ due to the SVD operation. Step 4 involves several matrix multiplications, leading to a complexity of $\mathcal{O}(ln^3)$, where l is the number of multiplications. For each view, Eq. (13) takes $\mathcal{O}(n^2)$, thus the complexity of step 5 is $\mathcal{O}(mn^2)$. The eigenvalue decomposition in step 7 costs $\mathcal{O}(n^3)$. Hence, the total cost of Algorithm 1 is $\mathcal{O}(T(mn^2 + (l+2)n^3))$, where T is the iteration number. To make our algorithm scalable for large-scale datasets, several off-the-shell acceleration methods could be used, such as *divide-and-conquer* [Talwalkar *et al.*, 2013] and the skinny SVD based ones [Zhang *et al.*, 2014; Xiao *et al.*, 2015].

3 Experiment

3.1 Experimental Setting

Datasets. Six real-world datasets are used in the experiment, which cover three text-type ones, *i.e.*, the *3-Source*[1] dataset, the *4-Areas*[2] dataset, and the *BBCSport* dataset provided by [Xia *et al.*, 2014]; and three image-type ones, *i.e.*, a 20-class subset [Li *et al.*, 2015b] of the *Caltech101*[3] image dataset, the UCI *Digit*[4] dataset, and the *Notting-Hill* dataset [Zhang *et al.*, 2009] provided by [Cao *et al.*, 2015]. We summarize these datasets in Table 1.

Table 1: Dataset Details

Dataset	#Instance	#View	#Class	Type
3-Sources	169	3	6	text
4-Areas	4236	2	4	text
BBCSport	544	2	5	text
Caltech101-20	2386	6	20	image
Digit	2000	2	10	image
Notting-Hill	550	3	5	image

Compared Methods. *Multi-View Clustering Methods*. By following [Xia *et al.*, 2014], we generate affinity matrix with Gaussian kernels, and conduct spectral clustering [Ng *et al.*, 2001] under different setting to obtain three baseline methods: (1) $\text{Spectral}_{\text{BSV}}$ returns the clustering result of the Best Single View (BSV); (2) $\text{Spectral}_{\text{CON}}$ concatenates the feature of each individual view and performs spectral clustering on the affinity graph built with the concatenated features; (3) $\text{Spectral}_{\text{SUM}}$ sums the affinity matrices of all the views and does spectral clustering on the averaged one. Besides, several effective multi-view clustering methods are also employed in the experiment, including Co-Regularized Spectral Clustering (CRSC) [Kumar *et al.*, 2011], Multi-View Nonnegative Matrix Factorization (MultiNMF) [Liu *et al.*, 2013b] and Robust Multi-View Spectral Clustering (RMVSC) [Xia *et al.*, 2014]. *Ensemble Clustering Methods*. Since our approach actually takes as input existing basic partitions (BPs), we also compare the proposed MVEC with three state-of-the-art ensemble clustering algorithms, including K-means based Consensus Clustering (KCC) [Wu *et al.*, 2015], Spectral Ensemble Clustering (SEC) [Liu *et al.*, 2015; 2017], and Robust SEC (RSEC) [Tao *et al.*, 2016]. SEC_{BSV} (RSEC_{BSV}) represents the best ensemble clustering result of each single view, whilst SEC_{SUM} (RSEC_{SUM}) denotes the result with the averaged co-association matrices of all the views. Note that, since KCC directly finds consensus clustering among multiple BPs, we only report the KCC_{BSV}.

Clustering Tools. Following [Wu *et al.*, 2015; Liu *et al.*, 2016; Tao *et al.*, 2017], we generate a set of $r = 100$ basic partitions for each individual view by using the Random Parameter Selection (RPS) strategy, which performs K-means algorithm with cosine distance and various cluster numbers. We feed these basic partitions sets as the default input to all the EC methods. For traditional clustering methods, we directly take the raw data as input and follow their preprocessing steps. Two widely-used clustering validation criteria are used to evaluate the clustering performance of all the methods, which are *Normalized Mutual Information* (*NMI*) and *Normalized Rand Index* (*Rn*) [Wu *et al.*, 2009]. Both these two metrics are positive measures and ranged from 0 to 1, where *NMI* will drop to zero for the random partition and *Rn* might be negative to the extremely poor clustering result.

[1] http://mlg.ucd.ie/datasets

[2] http://web.cs.ucla.edu/~yzsun/data/four_area.zip

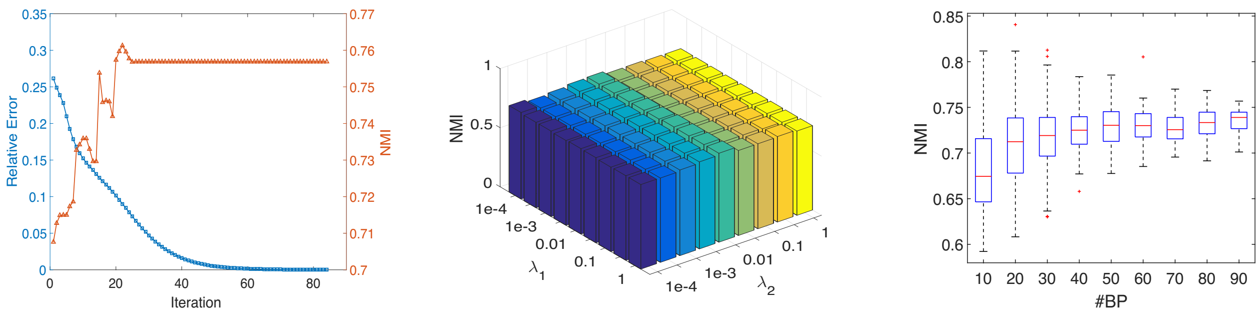
[3] http://www.vision.caltech.edu/Image_Datasets/Caltech101/

[4] http://archive.ics.uci.edu/ml/datasets.html

Table 2: Clustering performance on six real-world datasets by *NMI* (%)

Datasets	*3-Sources*$_*$	*4-Areas*$_*$	*BBCSport*$_*$	*Caltech101-20*	*Digit*	*Notting-Hill*$_*$	score
Spectral$_{\text{BSV}}$	47.14±2.45	43.27±8.53	71.76±0.32	59.65±1.35	63.92±2.40	71.58±2.90	4.49
Spectral$_{\text{CON}}$	51.66±2.22	0.87±0.09	54.77±1.52	38.19±0.84	65.09±2.37	64.28±5.66	3.41
Spectral$_{\text{SUM}}$	46.40±4.09	37.06±10.45	60.56±2.48	52.65±1.58	76.60±3.07	77.59±4.68	4.38
CRSC	51.55±2.93	32.97±2.39	63.71±2.48	56.29±1.36	72.23±3.05	75.20±4.74	4.38
MultiNMF	39.69±3.95	9.55±1.92	22.60±3.43	59.23±1.43	72.81±2.55	78.53±2.90	3.60
RMVSC	37.66±4.20	50.81±0.04	80.26±2.91	54.52±1.47	76.71±1.95	67.17±4.93	4.57
KCC$_{\text{BSV}}$	60.42±2.81	49.03±6.18	84.25±2.33	60.67±1.10	74.59±2.78	77.08±7.10	5.08
SEC$_{\text{BSV}}$	55.23±1.87	55.22±5.09	*87.26±2.08*	61.91±1.07	76.73±1.09	73.80±1.42	5.13
SEC$_{\text{SUM}}$	59.21±3.50	*74.06±0.00*	86.86±4.47	59.56±1.28	66.90±2.89	69.73±0.04	*5.22*
RSEC$_{\text{BSV}}$	*63.60±0.00*	43.09±4.42	64.95±0.00	*62.13±1.27*	**88.20±1.68**	78.61±0.00	5.03
RSEC$_{\text{SUM}}$	41.65±0.00	46.67±0.00	62.24±0.00	57.92±1.24	85.72±1.68	*80.50±0.20*	4.69
MVEC	**75.69±0.00**	**77.66±0.00**	**90.83±0.00**	**62.87±0.94**	*86.66±0.99*	**84.17±0.91**	**5.98**

The top *NMI* value is highlighted by red bold font and the second best by blue italic.
* indicates the difference between MVEC and runner-up is statistically significant on this dataset.



(a) Convergence Analysis (b) Parameter Analysis (c) Exploration to BPs number

Figure 1: Discussions to MVEC on the *3-Sources* dataset. (a) Convergence (blue line) and *NMI* (orange line) curves with respect to iteration. (b) Parameter analysis to λ_1 and λ_2. (c) Exploration to the number of Basic Partitions (BPs).

For all the compared methods, we use the true cluster number for fair comparison, and run the codes provided by authors with the parameters as suggested in their work. We test each method 20 times, and report the average result as well as the *standard deviation* (*std*). To validate the statistic significance of comparison results, the p-value is also calculated with t-test. In the experiment, we set $\lambda_1 = 1$ and $\lambda_2 = 0.01$ as the default setting for our MVEC method.

3.2 Clustering Performance

Table 2 and Table 3 summarize the clustering results of the proposed MVEC and other methods in terms of *NMI* and *Rn*, respectively. As can be seen, our approach generally performs best on all the datasets by both metrics. In details, we improve around 12% *NMI* (8% *Rn*) and 4% *NMI* (6% *Rn*) over the runner-up on *3-Sources* and *4-Areas*, which clearly shows the effectiveness of our approach. To further evaluate the performance, we compute a measurement score by following [Wu *et al.*, 2015]: score$(A_i) = \sum_j \frac{f(A_i, D_j)}{\max_i f(A_i, D_j)}$, where $f(A_i, D_j)$ indicates the *NMI* or *Rn* value of A_i method on the D_j dataset. This score gives an overall evaluation on all the datasets, which shows our approach outperforms the other methods substantially. Besides, according to p-value, our model outperforms the second best method with a statistically significant level in most cases.

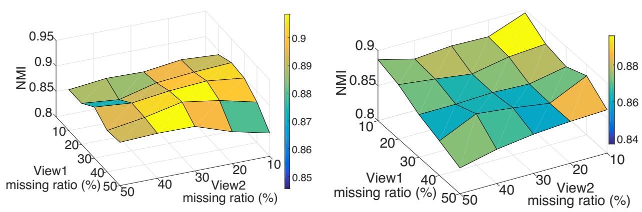
Ensemble Clustering vs Multi-View Clustering. Based on the input data, we may divide all the methods in Table 2 and Table 3 as two categories, such as ensemble clustering (EC) methods (*e.g.*, KCC$_{\text{BSV}}$, SEC$_{\text{BSV}}$ and KCC$_{\text{SUM}}$) that employ basic partitions, and multi-view clustering methods (*e.g.*, Spectral$_{\text{SUM}}$, CRSC and RMVC) that directly use the multi-view data. As can be seen, EC methods generally outperform the multi-view clustering ones, and even the EC method of single-view performs much better than the traditional ones. This demonstrates the significant superiority and great potentiality of using ensemble clustering methods to solve the multi-view clustering problem.

Single-View vs Multi-View. As shown by Table 2 and 3, traditional multi-view clustering methods perform slightly better than the single-view ones. For example, RMVSC only improves 0.08 (0.04) overall score over Spectral$_{\text{BSV}}$ by *NMI* (*Rn*). The similar observation appears at the ensemble clustering case, where the EC methods integrating BPs from all the views generally outperform the BSV ones with a little improvement. Specifically, SEC$_{\text{SUM}}$ is only 0.09 (0.21) higher than SEC$_{\text{BSV}}$ in Table 2 (3). These two observations indicate that: (i) Traditional multi-view clustering methods sometimes cannot improve the performance compared with the single-view ones, since they may suffer from the conflict among different views; (ii) Existing EC methods neglect to fully exploit the multi-view information. In contrast, our MVEC method not only improves the robustness of co-association matrix to the multi-view BPs, but also explicitly considers the connection between different views. Thus, we achieve better perfor-

Table 3: Clustering performance on six real-world datasets by *Rn* (%)

Datasets	*3-Sources*$_*$	*4-Areas*$_*$	*BBCSport*$_*$	*Caltech101-20*$_*$	*Digit*	*Notting-Hill*$_*$	score
$\text{Spectral}_{\text{BSV}}$	35.53±3.30	30.66±12.59	69.57±0.14	30.79±3.22	53.77±3.80	71.78±5.82	3.91
$\text{Spectral}_{\text{CON}}$	33.77±4.01	-0.02±0.00	46.67±3.71	16.51±1.05	55.30±3.54	62.85±8.11	2.84
$\text{Spectral}_{\text{SUM}}$	34.60±4.92	21.53±14.84	55.19±2.31	26.69±2.46	69.50±5.50	74.17±7.65	3.75
CRSC	35.78±5.12	17.73±4.16	61.92±0.83	29.43±2.43	66.79±5.05	69.61±7.94	3.77
MultiNMF	17.20±5.58	0.37±0.54	13.58±2.75	34.15±4.92	64.59±4.79	*77.43±6.84*	2.92
RMVSC	25.05±4.94	46.21±0.05	78.65±8.40	26.06±2.18	70.73±3.73	58.63±8.10	3.95
KCC_{BSV}	47.24±6.84	33.57±10.23	86.18±2.38	29.46±2.84	60.98±6.25	73.38±11.33	4.39
SEC_{BSV}	37.09±2.89	44.47±7.98	*89.73±3.12*	30.43±1.68	66.33±2.03	65.64±2.03	4.38
SEC_{SUM}	43.24±6.53	*77.35±0.00*	86.96±8.20	29.24±2.70	55.09±3.82	57.81±0.11	*4.59*
RSEC_{BSV}	*54.52±0.00*	25.42±8.85	40.74±0.00	33.65±2.18	**87.35±0.00**	73.20±0.00	4.31
RSEC_{SUM}	16.39±0.00	41.56±0.00	36.93±0.00	*36.54±2.11*	77.00±3.82	76.95±0.75	3.85
MVEC	**62.21±0.00**	**83.25±0.00**	**92.06±0.00**	**43.06±3.63**	*81.55±4.12*	**80.76±1.19**	**5.93**

The top *Rn* value is highlighted by red bold font and the second best by blue italic.
* indicates the difference between MVEC and runner-up is statistically significant on this dataset.



(a) Noisy BPs (b) Noisy Features

Figure 2: The performance of MVEC to noisy data on the *BBCSport* dataset.

mance than multi-view and ensemble clustering methods.

3.3 Discussion

Convergence. To show the convergence property of our method, we compute the relative error of the stop criterion by $\max\{\|\mathbf{\Delta}^{(v)}\|_{\mathrm{F}}/\|\mathbf{S}^{(v)}\|_{\mathrm{F}}\}_{v=1}^{m}$. Fig. 1(a) shows the relative error curve of our approach. As can be seen, MVEC converges steadily within 60 iterations. Moreover, we also plot the *NMI* curve during the optimization process. Along with iterations, the clustering performance of MVEC generally goes up, and achieves stable within several reasonable fluctuations, which indicates our approach has a strong convergence behavior. It is worthy to note that, the similar convergence behavior of MVEC can be observed on the other datasets.

Parameter Analysis. Recall the Eq. (3), λ_1 and λ_2 are two major parameters in our model, where λ_1 controls the rankness of the low-rank representation and λ_2 balances the sparseness of error matrices for all the views. We report the clustering performance of MVEC by ranging λ_1 and λ_2 within the set of $\{10^{-4}, 5\times10^{-4}, 10^{-3}, 5\times10^{-3}, 0.01, 0.05, 0.1, 0.5, 1\}$. As shown by Fig. 1(b), our approach is quite insensitive to these two parameters. Considering that, the error matrices can compensate the disagreement between different views, we suggest to set λ_2 as a relatively small value.

Impact of Basic Partitions Number. We employ $r = 100$ basic partitions (BPs) for each view v (denoted by $\Pi^{(v)}$) in all the above experiments. Here, we will explore the impact of different BPs number to the clustering performance of our approach. In details, for $\forall v$, we randomly sample r BPs from $\Pi^{(v)}$, and conduct MVEC with the sampled BPs sets of all the views. This process is repeated 100 times for a certain BPs number r. Fig. 1(c) shows the boxplots of *NMI* on the *3-Sources* dataset, where r is increased from 10 to 90 with a step size of 10. One may note that, as r increases, the *NMI* generally goes up and its variance tends to be reduced. This is mainly because a larger BPs number can enrich the diversity of basic partitions and also improve the stableness of ensemble clustering [Wu *et al.*, 2015]. In summary, our method is robust to a large BPs number, such as $r \geq 50$.

Impact of Noisy Data. We consider noisy data under two scenarios: (1) noisy basic partitions (BPs) and (2) noisy features, where the first one randomly corrupts BPs for each view according to a certain missing ratio, and the second randomly "dropouts" some feature dimensions for the multi-view data and then generates BPs upon the noisy features. As shown by Fig. 2, for both these two cases, we range the missing ratio from 10% to 50% along each view, and report the clustering performance of MVEC by *NMI*. Generally, our method is quite stable to the noises among basic partitions, due to the robustness inherited from ensemble clustering. On the other side, the performance of MVEC mainly degrades along with the diagonal direction of Fig. 2(b), showing that our method is robust to the noises of an individual view.

4 Conclusion

In this paper, we presented a novel Multi-View Ensemble Clustering (MVEC) framework, which employed co-association matrix to characterize the pairwise affinity of each sing-view data upon multiple basic partitions. A unified optimization framework with the *self-boost* constraint was provided to jointly learn a consensus affinity matrix shared by multiple views and find the final clustering result. Experiments on six real-world datasets showed the effectiveness of the proposed method compared with both ensemble and multi-view clustering methods.

Acknowledgments

This work is supported in part by the NSF IIS award 1651902, ONR Young Investigator Award N00014-14-1-0484 and U.S. Army Research Office Young Investigator Award W911NF-14-1-0218.

References

[Bickel and Scheffer, 2004] Steffen Bickel and Tobias Scheffer. Multi-view clustering. In *ICDM*, 2004.

[Bruno and Marchand-Maillet, 2009] Eric Bruno and Stphane Marchand-Maillet. Multiview clustering: a late fusion approach using latent models. In *SIGIR*, 2009.

[Cai *et al.*, 2010] Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[Cao *et al.*, 2015] Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang. Diversity-induced multi-view subspace clustering. In *CVPR*, 2015.

[Chaudhuri *et al.*, 2009] Kamalika Chaudhuri, Sham M. Kakade, Karen Livescu, and Karthik Sridharan. Multi-view clustering via canonical correlation analysis. In *ICML*, 2009.

[Dhillon *et al.*, 2004] Inderjit S. Dhillon, Yuqiang Guan, and Brian Kulis. Kernel k-means: Spectral clustering and normalized cuts. In *SIGKDD*, 2004.

[Ding and Fu, 2014] Zhengming Ding and Yun Fu. Low-rank common subspace for multi-view learning. In *ICDM*, 2014.

[Duchi *et al.*, 2008] John C. Duchi, Shai Shalev-Shwartz, Yoram Singer, and Tushar Chandra. Efficient projections onto the l_1-ball for learning in high dimensions. In *ICML*, 2008.

[Fred and Jain, 2005] Ana L. N. Fred and Anil K. Jain. Combining multiple clusterings using evidence accumulation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(6):835–850, 2005.

[Greene and Cunningham, 2009] Derek Greene and Pádraig Cunningham. A matrix factorization approach for integrating multiple data views. In *ECML PKDD*, 2009.

[Kumar and III, 2011] Abhishek Kumar and Hal Daum III. A co-training approach for multi-view spectral clustering. In *ICML*, 2011.

[Kumar *et al.*, 2011] Abhishek Kumar, Piyush Rai, and Hal Daume. Co-regularized multi-view spectral clustering. In *NIPS*, 2011.

[Li *et al.*, 2015a] Sheng Li, Ming Shao, and Yun Fu. Multi-view low-rank analysis for outlier detection. In *SDM*, 2015.

[Li *et al.*, 2015b] Yeqing Li, Feiping Nie, Heng Huang, and Junzhou Huang. Large-scale multi-view spectral clustering via bipartite graph. In *AAAI*, 2015.

[Li *et al.*, 2016] Sheng Li, Yaliang Li, and Yun Fu. Multi-view time series classification: A discriminative bilinear projection approach. In *CIKM*, 2016.

[Lin *et al.*, 2011] Zhouchen Lin, Risheng Liu, and Zhixun Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In *NIPS*, 2011.

[Liu *et al.*, 2013a] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(1):171–184, 2013.

[Liu *et al.*, 2013b] Jialu Liu, Chi Wang, Jing Gao, and Jiawei Han. Multi-view clustering via joint nonnegative matrix factorization. In *SDM*, 2013.

[Liu *et al.*, 2015] Hongfu Liu, Tongliang Liu, Junjie Wu, Dacheng Tao, and Yun Fu. Spectral ensemble clustering. In *SIGKDD*, 2015.

[Liu *et al.*, 2016] Hongfu Liu, Ming Shao, Sheng Li, and Yun Fu. Infinite ensemble for image clustering. In *SIGKDD*, 2016.

[Liu *et al.*, 2017] Hongfu Liu, Junjie Wu, Tongliang Liu, Dacheng Tao, and Yun Fu. Spectral ensemble clustering via weighted k-means: Theoretical and practical evidence. *IEEE Trans. Knowl. Data Eng.*, 29(5):1129–1143, 2017.

[Ng *et al.*, 2001] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *NIPS*, 2001.

[Strehl and Ghosh, 2003] Alexander Strehl and Joydeep Ghosh. Cluster ensembles — a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3:583–617, 2003.

[Talwalkar *et al.*, 2013] Ameet Talwalkar, Lester Mackey, Yadong Mu, Shih-Fu Chang, and Michael I. Jordan. Distributed low-rank subspace segmentation. In *ICCV*, 2013.

[Tao *et al.*, 2016] Zhiqiang Tao, Hongfu Liu, Sheng Li, and Yun Fu. Robust spectral ensemble clustering. In *CIKM*, 2016.

[Tao *et al.*, 2017] Zhiqiang Tao, Hongfu Liu, and Yun Fu. Simultaneous clustering and ensemble. In *AAAI*, 2017.

[Wang *et al.*, 2016] Yang Wang, Wenjie Zhang, Lin Wu, Xuemin Lin, Meng Fang, and Shirui Pan. Iterative views agreement: An iterative low-rank based structured optimization method to multi-view spectral clustering. In *IJCAI*, 2016.

[Wang *et al.*, 2017] Shuyang Wang, Zhengming Ding, and Yun Fu. Coupled marginalized auto-encoders for cross-domain multi-view learning. In *IJCAI*, 2017.

[Wu *et al.*, 2009] Junjie Wu, Hui Xiong, and Jian Chen. Adapting the right measures for k-means clustering. In *SIGKDD*, 2009.

[Wu *et al.*, 2015] Junjie Wu, Hongfu Liu, Hui Xiong, Jie Cao, and Jian Chen. K-means-based consensus clustering: A unified view. *IEEE Trans. Knowl. Data Eng.*, 27(1):155–169, 2015.

[Xia *et al.*, 2014] Rongkai Xia, Yan Pan, Lei Du, and Jian Yin. Robust multi-view spectral clustering via low-rank and sparse decomposition. In *AAAI*, 2014.

[Xiao *et al.*, 2015] Shijie Xiao, Wen Li, Dong Xu, and Dacheng Tao. Falrr: A fast low rank representation solver. In *CVPR*, 2015.

[Ye *et al.*, 2012] Guangnan Ye, Dong Liu, I-Hong Jhuo, and Shih-Fu Chang. Robust late fusion with rank minimization. In *CVPR*, 2012.

[Zha *et al.*, 2002] Hongyuan Zha, Xiaofeng He, Chris Ding, Ming Gu, and Horst D. Simon. Spectral relaxation for k-means clustering. In *NIPS*, 2002.

[Zhang *et al.*, 2009] Yifan Zhang, Changsheng Xu, Hanqing Lu, and Yueh-Min Huang. Character identification in feature-length films using global face-name matching. *IEEE Trans. Multimedia*, 11(7):1276–1288, 2009.

[Zhang *et al.*, 2014] Xin Zhang, Fuchun Sun, Guangcan Liu, and Yi Ma. Fast low-rank subspace segmentation. *IEEE Trans. Knowl. Data Eng.*, 26(5):1293–1297, 2014.

[Zhao and Fu, 2015] Handong Zhao and Yun Fu. Dual-regularized multi-view outlier detection. In *IJCAI*, 2015.

[Zhao *et al.*, 2016] Handong Zhao, Hongfu Liu, and Yun Fu. Incomplete multi-modal visual data grouping. In *IJCAI*, 2016.

[Zhao *et al.*, 2017] Handong Zhao, Zhengming Ding, and Yun Fu. Multi-view clustering via deep matrix factorization. In *AAAI*, 2017.

[Zhuang *et al.*, 2016] Liansheng Zhuang, Jingjing Wang, Zhouchen Lin, Allen Y. Yang, Yi Ma, and Nenghai Yu. Locality-preserving low-rank representation for graph construction from nonlinear manifolds. *Neurocomputing*, 175:715–722, 2016.