

Iterative Views Agreement: An Iterative Low-Rank Based Structured Optimization Method to Multi-View Spectral Clustering

Yang Wang[†], Wenjie Zhang[†], Lin Wu^{‡#}, Xuemin Lin[†], Meng Fang[§], Shirui Pan[‡]

[†]The University of New South Wales, Kensington, Sydney, Australia

[‡]The University of Adelaide, Australia [#]Australian Centre for Robotic Vision

[§]The University of Melbourne, Australia [‡]University of Technology Sydney, Australia

{wangy,zhangw,lxue}@cse.unsw.edu.au; lin.wu@adelaide.edu.au

meng.fang@unimelb.edu.au; shirui.pan@uts.edu.au

Abstract

Multi-view spectral clustering, which aims at yielding an agreement or consensus data objects grouping across multi-views with their graph laplacian matrices, is a fundamental clustering problem. Among the existing methods, Low-Rank Representation (LRR) based method is quite superior in terms of its effectiveness, intuitiveness and robustness to noise corruptions. However, it aggressively tries to learn a common low-dimensional subspace for multi-view data, while inattentively ignoring the local manifold structure in each view, which is critically important to the spectral clustering; worse still, the low-rank minimization is enforced to achieve the data correlation consensus among all views, failing to flexibly preserve the local manifold structure for each view. In this paper, 1) we propose a multi-graph laplacian regularized LRR with each graph laplacian corresponding to one view to characterize its local manifold structure. 2) Instead of directly enforcing the low-rank minimization among all views for correlation consensus, we separately impose low-rank constraint on each view, coupled with a mutual structural consensus constraint, where it is able to not only well preserve the local manifold structure but also serve as a constraint for that from other views, which iteratively makes the views more agreeable. Extensive experiments on real-world multi-view data sets demonstrate its superiority.

1 Introduction

Spectral clustering [Ng *et al.*, 2001; Zelnik-Manor and Perona, 2004], which aims at exploring the local nonlinear manifold (spectral graph) structure inherently embedded in high-dimensional data to partition data into disjoint meaningful groups, is a fundamental clustering problem. Because of its elegance, efficiency and good performance, spectral clustering has become one of the most popular clustering methods. Recently, great attention have shifted from conventional single view/graph to multi-view spectral clustering,

with the motivation of leveraging the complementary information from multi-view data sources where the same data set can be decomposed by different features *e.g.*, an image can be described by its color histogram or shape feature; one document can be represented by page link or document text. As explicitly claimed by numerous pieces of multi-view research [Xu *et al.*, 2015; 2013; Wang *et al.*, 2015c; 2014a], an individual view is unlikely to be sufficiently faithful for effective multi-view learning. Therefore, the integration of multi-view information is both valuable and necessary.

1.1 Motivation

Essentially, the critical issue of multi-view learning is to achieve the agreement/consensus [Gui *et al.*, 2014; Wang *et al.*, 2015b; 2014b; 2016] among all views given the complementary information from multi-views to yield a substantial superior performance in clustering over the single view paradigm. Numerous multi-view based methods are proposed for spectral clustering. [Huang *et al.*, 2012; Bickel and Scheffer., 2004] performs multi-view information incorporation into the clustering process by optimizing certain objective loss function. *Late fusion* strategy [Greene and Cunningham, 2009] designed for multi-view spectral clustering works by first deriving the spectral clustering performance regarding each view, and then combining multiple view-induced results into an optimal one. Such strategy, however, cannot ideally achieve the multi-view agreement, as each view cannot co-regularize with each other during the clustering process.

Canonical Correlation Analysis (CCA) based methods [Blaschko and Lampert., 2008; Chaudhuri *et al.*, 2009] for multi-view spectral clustering are developed by projecting the multi-view data sources into one common lower dimensional subspace, where the spectral clustering is subsequently conducted. One limitation of such method lies in the fact that one common lower-dimensional subspace cannot flexibly characterize the local spectral graph structures from heterogeneous views, resulting into an inferior multi-view spectral clustering. Kumar *et al.* [Kumar *et al.*, 2011] proposed a state-of-the-art co-regularized spectral clustering for multi-view data. They attempted to regularize the eigenvectors of view-dependent graph laplacians and achieve consensus clusters across views. Similarly, a co-training [Blum and

Mitchell, 1998; Wang and Zhou, 2010] framework is proposed for multi-view spectral clustering [Kumar and Daume, 2011], where the similarity matrix from one view is projected into the subspaces spanned by the eigenvectors from other views, then spectral clustering is conducted on such projected similarity matrix. This process is alternately performed until convergence, and the final result is formed by aggregating the clustering results from each individual view.

The above co-regularized [Kumar *et al.*, 2011] and co-training [Kumar and Daume, 2011] based methods can effectively achieve the clustering consensus under the scenario with noise corruption free in view-dependent feature representations. However, such assumption is hard to be satisfied in practice. To address such stand-out limitation, Low-Rank Representation (LRR) [Xia *et al.*, 2014; Liu *et al.*, 2010; Liu and Yan, 2011; Liu *et al.*, 2013] based approaches have been proposed for multi-view spectral clustering. The basic idea is to decompose data representation from any view into a view-dependent noise corruption term and a common low rank based representation shared by all views, which further leads to common data affinity matrix for clustering. The typical LRR [Xia *et al.*, 2014; Liu *et al.*, 2010] model is formulated below.

$$\begin{aligned} & \min_{Z, E_i} \|Z\|_* + \lambda \sum_{i \in V} \|E_i\|_1 \\ \text{s.t. } & X_i = X_i Z (\text{or } Z \text{ only}) + E_i, i = 1, \dots, V, \end{aligned} \quad (1)$$

where V denotes the number of all views; $X_i \in \mathbb{R}^{d_i \times n}$ denotes the data feature representation for the i^{th} view, n is the number of data objects for each view; d_i is the feature representation dimensions for the i^{th} view. $Z \in \mathbb{R}^{n \times n}$ represents the self-expressive linear sample correlations [Liu *et al.*, 2010] shared by all views with the assumption that the similar samples can be linearly reconstructed by each other. $E_i \in \mathbb{R}^{d_i \times n}$ models the possible noise corruptions in the feature representations for the i^{th} view. $\|E_i\|_1$ is the ℓ_1 norm of E_i representing the summation of the absolute value of all entries from E_i ; λ is the balance parameter.

Despite the effectiveness of LRR for multi-view spectral clustering, they still arguably face the following fundamental limitations:

- LRR attempts to learn a common lowest-rank representation revealing a low-dimensional subspace structure, but inattentively ignore the distinct manifold structures in each view, which turns out to be critically important to multi-view spectral clustering.
- Low-rank constraint is imposed to enforce all views to share the consensus Z in Eq.(1) [Xia *et al.*, 2014], however, such enforced common representation may not flexibly preserve the local manifold structure from heterogeneous views, resulting into a non-ideal multi-view clustering performance.

1.2 Our contributions

To address those stand-out limitations, our method delivers the following novel features:

- To characterize the non-linear spectral graph structure from each view, inspired by [Yin *et al.*, 2016], we propose to couple LRR with multi-graph regularization, where each graph laplacian regularization can characterize the view-dependent non-linear local data similarity.
- To achieve the view agreement while preserving the data correlations within each view, we present an iterative view agreement process in optimizing our objective function. During each iteration, the low-rank representation yielded from each view serves as the constraint to regulate the representation learning from other views. This process iteratively boosts these representations to be more agreeable.
- To model the above intuitions, we figure out a novel objective function and the Linearized Alternating Direction Method with Adaptive Penalty (LADMAP) [Lin *et al.*, 2011] method is deployed to solve it.

2 Iterative Low-Rank based Structured Optimization Method to Multi-view Spectral Clustering

It is well known that the critical issue for spectral clustering lies in how to effectively model the local nonlinear manifold structure [Zelnik-Manor and Perona, 2004]. Hence, for each view, we aim at preserving such nonlinear manifold structure of original high-dimensional data set within the space spanned by the low-rank sparse representations Z_i for the i^{th} view. This can be formulated as:

$$\begin{aligned} & \frac{1}{2} \sum_{j,k}^n \|z_j^i - z_k^i\|^2 W_{jk}^i \\ & = \sum_{j=1}^N (z_j^i)^T z_j^i D_{jj}^i - \sum_{j,k}^N (z_k^i)^T z_j^i W_{jk}^i \\ & = \text{Tr}(Z_i^T D_i Z_i) - \text{Tr}(Z_i^T W_i Z_i) = \text{Tr}(Z_i^T L_i Z_i), \end{aligned} \quad (2)$$

where $z_k^i \in \mathbb{R}^n$ is the k^{th} row of $Z_i \in \mathbb{R}^{n \times n}$ representing the linear correlation representation between x_k and x_j ($j \neq k$) in the i^{th} view; W_{jk}^i is the $(j, k)^{\text{th}}$ entry of the similarity matrix W_i , which encodes the similarity between x_j and x_k from the original high dimensional space for the i^{th} view; $W_i \in \mathbb{R}^{n \times n}$ is the similarity matrix for all the data objects from X_i ; D_i is a diagonal matrix with its k^{th} diagonal entry to be the summation of the k^{th} row of W_i , and $L_i = D_i - W_i$ is the graph laplacian matrix for the i^{th} view; thus Eq.(2) is always dubbed *graph laplacian regularizer*. In this paper, we choose Gaussian kernel to calculate W_{jk}^i as

$$W_{jk}^i = e^{-\frac{\|x_j^i - x_k^i\|_2^2}{2\sigma^2}}, \quad (3)$$

where σ is the bandwidth parameter and $\|\cdot\|_2$ denotes the ℓ_2 norm; Eq.(3) holds if x_k^i is within the s nearest neighbors of x_j^i or vice versa, and it is 0 otherwise. W_{jj}^i ($\forall j$) is set to 0 to avoid self-loop. Eq.(2) explicitly requires Z_i to well characterize the local manifold structure inherently embedded in

original high-dimensional X_i for the i^{th} view, which is of importance to spectral clustering.

Based on the above, we leverage the above graph laplacian regularizer with the low-rank representation. Considering the global clustering structure captured by low-rank representation may prevent us from directly imposing graph Laplacian regularizer for local manifold structure, we propose to impose the sparsity norm ℓ_1 on Z_i , denoted as $\|Z_i\|_1$, which can discriminatively extract the local sparse representative neighborhood of each data object.

As explicitly revealed by most of the multi-view clustering research [Kumar *et al.*, 2011; Kumar and Daume, 2011; Bickel and Scheffer., 2004], it is always anticipated that a data point should be assigned to the same cluster irrespective of views. In other words, the critical issue to ensure ideal multi-view clustering performance is to achieve the clustering agreement among all views. Based on that, we aim to minimizing the difference of such low-rank and sparse representations from different views by proposing a consensus term to coordinate all views to reach clustering agreement.

$$\begin{aligned}
& \min_{Z_i, E_i (i \in V)} \sum_{i \in V} \left(\underbrace{\|Z_i\|_*}_{\text{Low-rank representation}} + \underbrace{\lambda_1 \|E_i\|_1}_{\text{noise and corruption robustness}} \right) \\
& + \underbrace{\lambda_2 \|Z_i\|_1}_{\text{local sparsity modeling}} + \underbrace{\lambda_3 \text{Tr}(Z_i^T L_i Z_i)}_{\text{Graph regularization}} \\
& + \underbrace{\frac{\beta}{2} \sum_{j \in V, j \neq i} \|Z_i - Z_j\|_2^2}_{\text{Views-agreement}} \\
& \text{s.t. } i = 1, \dots, V, \quad X_i = X_i Z_i + E_i, Z_i \geq 0,
\end{aligned} \tag{4}$$

where

- $\|Z_i\|_*$ denotes the low-rank representation revealing the global clustering structure regarding X_i .
- $\|Z_i\|_1$ aims at extracting the local sparse representation of each data object in X_i .
- $\text{Tr}(Z_i^T L_i Z_i)$ characterizes the local manifold structure.
- $\sum_{i, j \in V} \|Z_i - Z_j\|_2^2$ characterizes the agreement among the sparse and low-rank representations from all V views.
- $\|E_i\|_1$ models the possible Laplacian noise contained by X_i , we pose ℓ_1 on $\|E_i\|$ for noise robustness.
- $Z_i \geq 0$ is a non-negative constraint to ensure that each data object is amid its neighbors, through $X_i = X_i Z_i + E_i$, so that the data correlations can be well encoded for the i^{th} view.
- $\lambda_1, \lambda_2, \lambda_3, \beta$ are all trade-off parameters

Eq.(4) is a typical low-rank optimization problem, and a lot of methods are available to solve it. Among these methods, the Alternating Direction Method is the typical solution, which aims at updating each variable alternatively by minimizing the augmented lagrangian function in a Gauss-Seidel fashion. In this paper, we deploy the method of **Linearized Alternating Direction Method with Adaptive Penalty**, dubbed

LADMMap [Lin *et al.*, 2011]. The underlying philosophy of **LADMMap** is to linearly represent the smooth component, which enables Lagrange multipliers to be updated within the feasible approximation error.

Observing that solving all the $Z_i, E_i (i \in V)$ pairs follows the same type of optimization strategy, we only present the optimization strategy for the i^{th} view. To resolve this, we first introduce an auxiliary variable G_i , then solving the Eq.(4) with respect to Z_i, E_i and G_i can be written as follows

$$\begin{aligned}
& \min_{Z_i, E_i, G_i} \|Z_i\|_* + \lambda_1 \|E_i\|_1 + \lambda_2 \|G_i\|_1 \\
& + \lambda_3 \text{Tr}(Z_i^T L_i Z_i) + \frac{\beta}{2} \sum_{j \in V, j \neq i} \|Z_i - Z_j\|_2^2 \tag{5} \\
& \text{s.t. } X_i = X_i Z_i + E_i, G_i = Z_i, G_i \geq 0.
\end{aligned}$$

We then present the augmented lagrangian function of Eq.(5) below

$$\begin{aligned}
& \mathcal{L}(Z_i, E_i, G_i, K_1^i, K_2^i) \\
& = \|Z_i\|_* + \lambda_1 \|E_i\|_1 + \lambda_2 \|G_i\|_1 + \lambda_3 \text{Tr}(Z_i^T L_i Z_i) \\
& + \frac{\beta}{2} \sum_{j \in V, j \neq i} \|Z_i - Z_j\|_2^2 + \langle K_1^i, X_i - X_i Z_i - E_i \rangle + \\
& \langle K_2^i, Z_i - G_i \rangle + \frac{\mu}{2} (\|X_i - X_i Z_i - E_i\|_2^2 + \|Z_i - G_i\|_2^2),
\end{aligned} \tag{6}$$

where $K_1^i \in \mathbb{R}^{d_i \times n}$ and $K_2^i \in \mathbb{R}^{n \times n}$ are Lagrange multipliers, $\langle \cdot, \cdot \rangle$ is the inner product and $\mu > 0$ is a penalty parameter. We update each of the above variables alternatively by minimizing Eq.(6) while with other variables fixed. In what follows, we will provide the details of optimizing Eq.(6) with respect to each variable in next section.

3 Optimization Strategy

3.1 Updating Z_i

Minimizing Eq.(6) w.r.t. Z_i is equivalent to minimizing the following

$$\begin{aligned}
\mathcal{L}_1 = & \|Z_i\|_* + \lambda_3 \text{Tr}(Z_i^T L_i Z_i) + \frac{\beta}{2} \sum_{j \in V, j \neq i} \|Z_i - Z_j\|_2^2 \\
& + \frac{\mu}{2} \|X_i - X_i Z_i - E_i\|_2^2 + \frac{1}{\mu} \|K_1^i\|_2^2 + \frac{\mu}{2} \|Z_i - G_i\|_2^2 + \frac{1}{\mu} \|K_2^i\|_2^2
\end{aligned} \tag{7}$$

It cannot yield a closed form throughout Eq.(7). Thanks to **LADMMap**, we can approximately reconstruct the smooth terms of \mathcal{L}_1 via a linear manner. The smooth terms of \mathcal{L}_1 are summarized below

$$\begin{aligned}
& \mathcal{Q}_l(Z_i, E_i, G_i, K_1^i, K_2^i) \\
& = \lambda_3 \text{Tr}(Z_i^T L_i Z_i) + \frac{\beta}{2} \sum_{j \in V, j \neq i} \|Z_i - Z_j\|_2^2 \\
& + \frac{\mu}{2} \|X_i - X_i Z_i - E_i\|_2^2 + \frac{1}{\mu} \|K_1^i\|_2^2 + \frac{\mu}{2} \|Z_i - G_i\|_2^2 + \frac{1}{\mu} \|K_2^i\|_2^2
\end{aligned} \tag{8}$$

Based on Eq.(8), we convert the problem of minimizing Eq.(7) to minimize Eq.(9) below

$$\mathcal{L}_1 = \|Z_i\|_* + \left\langle \frac{\partial \mathcal{Q}_l}{\partial Z_k}, Z_i - Z_i(k) \right\rangle + \frac{\xi}{2} \|Z_i - Z_i(k)\|_2^2, \tag{9}$$

where $\frac{\partial \mathcal{Q}_l}{\partial Z_i(k)}$ denotes the partial gradient of \mathcal{Q}_l w.r.t. Z at $Z_i(k)$, and \mathcal{Q}_l is approximated by the linear representation $\langle \frac{\partial \mathcal{Q}_l}{\partial Z_i(k)}, Z_i - Z_i(k) \rangle$ w.r.t. $Z_i(k)$ together with a proximal term $\frac{\xi}{2} \|Z_i - Z_i(k)\|_2^2$. The above replacement is valid provided that $\xi > 2\lambda_3 e(L_i) + \mu(1 + \|X_i\|_2^2)$, where $e(L_i)$ denotes the largest eigenvalue of L_i . Then the following closed form holds for Eq.(9) for each update.

$$Z_i = \Theta_{\frac{1}{\xi}}(Z_i(k) - \frac{\partial \mathcal{Q}_l}{\partial Z_i(k)} \cdot \frac{1}{\xi}), \quad (10)$$

where $\Theta_\epsilon(A) = US_\epsilon(\Sigma)V^T$ represents the Singular Value Threshold (SVT) operation. USV^T is the singular value decomposition of matrix A , and $S_\epsilon(x) = \text{sign}(x)\max(\|x\| - \epsilon, 0)$ is called the soft threshold operator, $\text{sign}(x)$ is 1 if it is positive and 0 otherwise.

Insights for Iterative Views Agreement.

We remark that the intuitions for iterative (during each iteration) views-agreement can be captured by expanding $\frac{\partial \mathcal{Q}_l}{\partial Z_i(k)}$ below

$$\begin{aligned} \frac{\partial \mathcal{Q}_l}{\partial Z_i(k)} &= \lambda_3 Z_i(k)(L_i^T + L_i) \\ &+ \mu X_i^T (X_i Z_i(k) - X_i + E_i - \frac{1}{\mu} K_1^i) \\ &+ \mu (Z_i(k) - G_i + \frac{1}{\mu} K_2^i) + \beta \underbrace{\sum_{j \in V, j \neq i} \|Z_i(k) - Z_j\|}_{\sum_{j \in V, j \neq i} Z_i(k) - \sum_{j \in V, j \neq i} Z_j} \end{aligned} \quad (11)$$

We expand the last term $\sum_{j \in V, j \neq i} \|Z_i(k) - Z_j\| = \sum_{j \in V, j \neq i} Z_i(k) - \sum_{j \in V, j \neq i} Z_j$, then we re-write Eq.(11) below

$$\frac{\partial \mathcal{Q}_l}{\partial Z_i(k)} = C_i - \beta \sum_{j \in V, j \neq i} Z_j, \quad (12)$$

where

$$\begin{aligned} C_i &= \lambda_3 Z_i(k)(L_i^T + L_i) + \mu X_i^T (X_i Z_i(k) \\ &- X_i + E_i - \frac{1}{\mu} K_1^i) + \\ &\mu (Z_i(k) - G_i + \frac{1}{\mu} K_2^i) + \beta \sum_{j \in V, j \neq i} Z_j(k) \end{aligned}$$

Substituting Eq.(12) into Eq.(10) yields Eq.(13)

$$\begin{aligned} Z_i &= \Theta_{\frac{1}{\xi}}(Z_i(k) - (C_i - \beta \sum_{j \in V, j \neq i} Z_j) \cdot \frac{1}{\xi}) \\ &= \Theta_{\frac{1}{\xi}}(Z_i(k) - \frac{C_i}{\xi} + \beta \underbrace{\sum_{j \in V, j \neq i} Z_j}_{\text{Influence from other views}} \cdot \frac{1}{\xi}), \end{aligned} \quad (13)$$

where Z_i updating is explicitly influenced from other views *i.e.*, $\sum_{j \in V, j \neq i} Z_j$, which reveals that such low-rank representations *e.g.*, Z_i updating from each view *e.g.*, the i^{th} view are formed by referring to the other views, while served as a constraint to update other views for each iteration so that the complementary information from all views are intuitively leveraged towards a final agreement for clustering.

3.2 Updating E_i

Minimizing Eq.(6) w.r.t. E_i is equivalent to solving the following optimization problem

$$\min_{E_i} \lambda_1 \|E_i\|_1 + \frac{\mu}{2} \|E_i - (X_i - X_i Z_i + \frac{1}{\mu} K_1^i)\|_2^2, \quad (14)$$

where the following closed form solution is hold for E_i according to [Cai *et al.*, 2008]

$$E_i = S_{\frac{\lambda_1}{\mu}}(X_i - X_i Z_i + \frac{1}{\mu} K_1^i) \quad (15)$$

3.3 Updating G_i

Minimizing Eq.(6) w.r.t. G_i is equivalent to solving the following optimization problem

$$\min_{G_i \geq 0} \lambda_2 \|G_i\|_1 + \frac{\mu}{2} \|G_i - (Z_i + \frac{1}{\mu} K_2^i)\|_2^2, \quad (16)$$

where the following closed form solution holds for G_i according to [Cai *et al.*, 2008]

$$G_i = \max\{S_{\frac{\lambda_2}{\mu}}(Z_i + \frac{1}{\mu} K_2^i), 0\} \quad (17)$$

3.4 Updating K_1^i and K_2^i

We update Lagrange multipliers K_1^i via

$$K_1^i = K_1^i + \mu(X_i - X_i Z_i - E_i) \quad (18)$$

and K_2^i via

$$K_2^i = K_2^i + \mu(Z_i - G_i) \quad (19)$$

We remark that μ can be tuned using the adaptive updating strategy as suggested by [Lin *et al.*, 2011] to yield a faster convergence. The optimization strategy alternatively update each variable while fixing others until the convergence condition is met.

Thanks to **LADMAP** [Lin *et al.*, 2011], the above optimization process converges to a globally optimal solution enjoyed. Besides, we may employ the Lanczos method to compute the largest singular values and vectors by only performs multiplication of $Z_i(k) - \frac{\partial \mathcal{Q}_l}{\partial Z_i(k)} \cdot \frac{1}{\xi}$ with vectors, which can be efficiently computed by such successive matrix-vector multiplications.

3.5 Clustering with $Z_i(i = 1, \dots, V)$

Once the converged $Z_i(i = 1, \dots, V)$ are learned for each of the V views, we normalize all column vectors of $Z_i(i = 1, \dots, V)$ while set small entries under given threshold τ to be 0. After that, we can calculate the similarity matrix $W_i(j, k) = \frac{Z_i(j, k) + Z_i(k, j)}{2}$ for the i^{th} view between the j^{th} and k^{th} data objects. The final data similarity matrix for all views are defined as

$$W = \frac{\sum_i^V W_i}{V} \quad (20)$$

The spectral clustering is performed on W calculated via Eq.(20) to yield the final multi-view spectral clustering result.

Table 1: Summary of the multi-view data sets used in our experiments.

Features	UCI	AwA	NUS
1	FC (76)	CQ (2688)	CH(65)
2	PC (216)	LSS (2000)	CM(226)
3	-	PHOG (252)	CORR(145)
4	-	SIFT(2000)	EDH(74)
5	-	RGSIFT(2000)	WT(129)
6	-	SURF(2000)	-
# of data	2000	4000	26315
# of classes	10	50	31

4 Experiments

We evaluate our method on the following data sets:

- **UCI handwritten Digit set**¹: It consists of features of hand-written digits (0-9). The dataset is represented by 6 features and contains 2000 samples with 200 in each category. Analogous to [Lin *et al.*, 2011], we choose 76 Fourier coefficients (FC) of the character shapes and the 216 profile correlations (PC) as two views.
- **Animal with Attribute (AwA)**²: It consists of 50 kinds of animals described by 6 features (views): Color histogram (CQ, 2688-dim), local self-similarity (LSS, 2000-dim), pyramid HOG (PHOG, 252-dim), SIFT (2000-dim), Color SIFT (RGSIFT, 2000-dim), and SURF (2000-dim). We randomly sample 80 images for each category and get 4000 images in total.
- **NUS-WIDE-Object (NUS)** [Chua *et al.*, 2009]: The data set consists of 30000 images from 31 categories. We construct 5 views using 5 features as provided by the website ³: 65-dimensional color histogram (CH), 226-dimensional color moments (CM), 145-dimensional color correlation (CORR), 74-dimensional edge estimation (EDH), and 129-dimensional wavelet texture (WT).

These data sets are summarized in Table 1.

4.1 Baselines

We compare our approach with the following state-of-the-art baselines:

- **MFMSC**: Using the concatenation of multiple features to perform spectral clustering.
- **Multi-view affinity aggregation for multi-view spectral clustering (MAASC)** [Huang *et al.*, 2012].
- **Canonical Correlation Analysis (CCA) based multi-view spectral clustering (CCAMSC)** [Chaudhuri *et al.*, 2009]: Projecting multi-view data into a common subspace, then perform spectral clustering.
- **Co-regularized multi-view spectral clustering (CoMVSC)** [Kumar *et al.*, 2011]: It regularizes the eigenvectors of view-dependent graph laplacians and achieve consensus clusters across views.

¹<http://archive.ics.uci.edu/ml/datasets/Multiple+Features>

²<http://attributes.kyb.tuebingen.mpg.de>

³ims.comp.nus.edu.sg/research/NUS-WIDE.html

- **Co-training** [Kumar and Daume, 2011]: Alternately modify one view’s eigenspace of graph laplacian by referring to the other views’ graph laplacian and their corresponding eigenvectors, upon which, the spectral clustering is conducted. Such process is performed until convergence.
- **Robust Low-Rank Representation method (RLRR)** [Xia *et al.*, 2014], as formulated in Eq.(1).

4.2 Experimental Settings and Parameters Study

For fair comparison, we implement these competitors by following their experimental setting and the parameter tuning steps in their papers. The Gaussian kernel is used throughout experiments on all data sets and σ in Eq.(3) is learned by self-tuning method [Zelnik-Manor and Perona, 2004], and $s = 20$ to construct s -nearest neighbors for each data object to calculate Eq.(3). To measure the clustering results, we use two standard metrics: clustering accuracy (ACC) (Ratio for the number of data objects having same clustering label and ground truth label against total data objects), and normalized mutual information (NMI). Please refer to [Chen *et al.*, 2011] for details of these two clustering metrics. All experiments are repeated 10 times, and we report their averaged mean value.

Feature noise modeling for robustness: Following [Siyahjani *et al.*, 2015], for each view-specific feature representation, 20% feature elements are corrupted with uniform distribution over the range $[-5, 5]$, which is consistent to the practical setting while matching with **RLRR** and our method.

We set $\lambda_1 = 2$, $\lambda_2 = 0.08$ in Eq.(4). To validate the effectiveness of Multi-graph regularization and iterative views agreement, we test the value of ACC and NMI over a range of λ_3 and β in Eq.(4) in the next subsection.

4.3 Validating Multi-graph regularization and Iterative views agreement

We test λ_3 for multi-graph regularization term and β for iterative views agreement within the interval $[0.001, 10]$ over the AwA data set and adopt such setting for other data sets. Specifically, we test each value of one parameter while fixing the value of the other parameter, the results are then illustrated in Fig. 1.

From both Fig.1 (a) and (b), the following observations can be identified:

- when fixing the value of λ_3 , increasing the value of β can basically improve the ACC and NMI value of our method. The similar observation can be identified vice versa; that is, fixing the value of β , meanwhile increasing the λ_3 can always lead to the clustering improvement in terms of both ACC and NMI.
- Both the above clustering measures ACC and NMI will unsurprisingly increase when both λ_3 and β increases until reach the optimal pair-combinations, then slightly decrease.

Upon the above observations, we choose a balance pair values: $\lambda_3 = 0.5$ and $\beta = 0.1$ for our method.

Table 2: Clustering results in terms of ACC on three benchmark data sets.

ACC (%)	UCI digits	AwA	NUS
MFMSC	43.81	17.13	22.81
MAASC	51.74	19.44	25.13
CCAMSC	73.24	24.04	27.56
CoMVSC	80.27	29.93	33.63
Co-training	79.22	29.06	34.25
RLRR	83.67	31.49	35.27
Ours	86.39	37.22	41.02

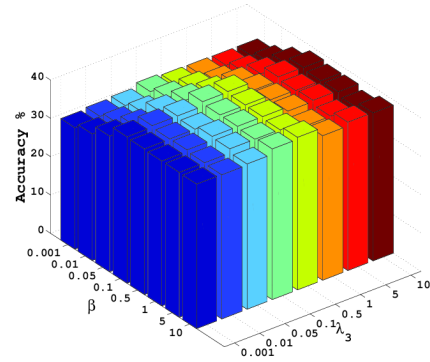
Table 3: Clustering results in terms of NMI on three benchmark data sets.

NMI (%)	UCI digits	AwA	NUS
MFMSC	41.57	11.48	12.21
MAASC	47.85	12.93	11.86
CCAMSC	56.51	15.62	14.56
CoMVSC	63.82	17.30	7.07
Co-training	62.07	18.05	8.10
RLRR	81.20	25.57	18.29
Ours	85.45	31.74	20.61

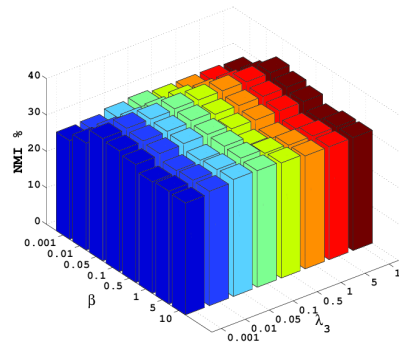
4.4 Experimental Results and Analysis

We report the compared clustering results in terms of ACC and NMI in Table 2 and Table 3, upon which, the following observations can be drawn:

- Nearly most of the clustering performance in terms of both ACC and NMI are better than **RLRR**, which further demonstrates the effectiveness of our multi-graph regularization and iterative views agreement scheme combining with LRR scheme for multi-view spectral clustering.
- First, comparing with view-fusion methods like **MFMSC** and **MAASC**, our method improves the clustering performance by a notable margin on all data sets. Specifically, it highly improves the clustering performance in terms of ACC from 43.81% (**MFMSC**), 51.74% (**MAASC**) to 86.39% on UCI digits data set. Such notable improvement can be also observed on AwA and NUS data sets.
- Second, **CCAMSC** that learns a common low-dimensional subspace among multi-view data is less effective in clustering due to its incapability of encoding local graph structures from heterogeneous views within only a common subspace. In contrast, our method can well address such problem with a novel iterative views-agreement scheme, which is notably evidenced in terms of both ACC and NMI.
- Comparing with co-regularized paradigms (**CoMVSC**, and **Co-training**), our method works more effectively in the presence of noise corruptions. For example, in NUS data set, it improves the clustering accuracy from 33.63% (**CoMVSC**), 34.25% (**Co-training**) to 41.02%. Although **RLRR** is also effective to deal with practical



(a)



(b)

Figure 1: Parameters study of λ_3 and β for our multi-graph regularization and iterative views agreement scheme on AwA dataset. (a) ACC against parameters λ_3 and β . (b) NMI against parameters λ_3 and β .

noise-corrupted multi-view data. However, as aforementioned, learning only one common low-rank correlation representation shared by all views is failed to flexibly capture all the local nonlinear manifold structures from all views, which is crucial to spectral clustering, while our technique can deliver a better performance.

5 Conclusions

In this paper, we propose an iterative structured low-rank optimization method to multi-view spectral clustering. Unlike existing methods, Our method can well encode the local data manifold structure from each view-dependent feature space, and achieve the multi-view agreement via an iterative fashion, while better preserve the flexible nonlinear manifold structure from all views. The superiorities are validated by extensive experiments over real-world multi-view data sets.

One future direction is to adapt the proposed iterative fashion technique to cross-view based research [Wu *et al.*, 2013; Wang *et al.*, 2015a] by dealing with multiple data source yet corresponding to the same latent semantics. We aim to develop the novel iterative technique to learn the projections for multiple data sources into the common latent space to well characterize the shared latent semantics.

References

- [Bickel and Scheffer., 2004] S. Bickel and T. Scheffer. Multi-view clustering. In *IEEE ICDM*, 2004.
- [Blaschko and Lampert., 2008] M. Blaschko and C. Lampert. Correlational spectral clustering. In *CVPR*, 2008.
- [Blum and Mitchell, 1998] Avrim Blum and Tom M. Mitchell. Combining labeled and unlabeled data with co-training. In *COLT*, 1998.
- [Cai *et al.*, 2008] Jian-Feng Cai, Emmanuel J. Candes, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization.*, 20(4):1956–1982, 2008.
- [Chaudhuri *et al.*, 2009] K. Chaudhuri, S. Kakade, K. Livescu, and K. Sridharan. Multi-view clustering via canonical correlation analysis. In *ICML*, 2009.
- [Chen *et al.*, 2011] Wen-Yen Chen, Yangqiu Song, Hongjie Bai, Chih-Jen Lin, and Edward Y. Chang. Parallel spectral clustering in distributed systems. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(3):568–586, 2011.
- [Chua *et al.*, 2009] Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yan-Tao Zheng. Nus-wide: A real-world web image database from national university of singapore. In *ACM CIVR*, 2009.
- [Greene and Cunningham, 2009] D. Greene and P. Cunningham. A matrix factorization approach for integrating multiple data views. In *ECMLPKDD*, 2009.
- [Gui *et al.*, 2014] Jie Gui, Dacheng Tao, Zhenan Sun, Yong Luo, Xinge You, and Yuan Yan Tang. Group sparse multiview patch alignment framework with view consistency for image classification. *IEEE Transactions on Image Processing*, 23(7):3126–3137, 2014.
- [Huang *et al.*, 2012] Hsin-Chien Huang, Yung-Yu Chuang, and Chu-Song Chen. Affinity aggregation for spectral clustering. In *CVPR*, 2012.
- [Kumar and Daume, 2011] Abhishek Kumar and Hal Daume. 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.
- [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 and Yan, 2011] Guangcan Liu and Shuicheng Yan. Latent low-rank representation for subspace segmentation and feature extraction. In *ICCV*, 2011.
- [Liu *et al.*, 2010] Guangcan Liu, Zhouchen Lin, and Yong Yu. Robust subspace segmentation by low-rank representation. In *ICML*, 2010.
- [Liu *et al.*, 2013] 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.
- [Ng *et al.*, 2001] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *NIPS*, 2001.
- [Siyahjani *et al.*, 2015] Farzad Siyahjani, Ranya Almohsen, Sinan Sabri, and Gianfranco Doretto. A supervised low-rank method for learning invariant subspace. In *ICCV*, 2015.
- [Wang and Zhou, 2010] Wei Wang and Zhi-Hua Zhou. A new analysis of co-training. In *ICML*, 2010.
- [Wang *et al.*, 2014a] Yang Wang, Xuemin Lin, Lin Wu, Wenjie Zhang, and Qing Zhang. Exploiting correlation consensus: towards subspace clustering for multi-modal data. In *ACM Multimedia*, 2014.
- [Wang *et al.*, 2014b] Yang Wang, Jian Pei, Xuemin Lin, and Qing Zhang. An iterative fusion approach to graph-based semi-supervised learning from multiple views. In *PAKDD*, 2014.
- [Wang *et al.*, 2015a] Yang Wang, Xuemin Lin, Lin Wu, Wenjie Zhang, and Qing Zhang. Lbmch: Learning bridging mapping for cross-modal hashing. In *ACM SIGIR*, 2015.
- [Wang *et al.*, 2015b] Yang Wang, Xuemin Lin, Lin Wu, Wenjie Zhang, Qing Zhang, and Xiaodi Huang. Robust subspace clustering for multi-view data by exploiting correlation consensus. *IEEE Trans. Image Processing.*, 24(11):3939–3949, 2015.
- [Wang *et al.*, 2015c] Yang Wang, Wenjie Zhang, Lin Wu, Xuemin Lin, and Xiang Zhao. Unsupervised metric fusion over multiview data by graph random walk-based cross-view diffusion. *IEEE Trans. Neural Netw. Learning Syst.*, 2015.
- [Wang *et al.*, 2016] Yang Wang, Xuemin Lin, Lin Wu, Qing Zhang, and Wenjie Zhang. Shifting multi-hypergraphs via collaborative probabilistic voting. *Knowl. Inf. Syst.*, 46(3):515–536, 2016.
- [Wu *et al.*, 2013] Lin Wu, Yang Wang, and John Shepherd. Efficient image and tag co-ranking: a bregman divergence optimization method. *ACM Multimedia*, 2013.
- [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.
- [Xu *et al.*, 2013] Chang Xu, Dacheng Tao, and Chao Xu. A survey on multi-view learning. *Corr abs/1304.5634*, 2013.
- [Xu *et al.*, 2015] Chang Xu, Dacheng Tao, and Chao Xu. Multi-view intact space learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015.
- [Yin *et al.*, 2016] Ming Yin, Junbin Gao, and Zhouchen Lin. Laplacian regularized low-rank representation and its applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(3):504–517, 2016.
- [Zelnik-Manor and Perona, 2004] Lihi Zelnik-Manor and Pietro Perona. Self-tuning spectral clustering. In *NIPS*, 2004.