Proceedings of Machine Learning Research 95:17-32, 2018 ACML 2018

# Co-regularized Multi-view Subspace Clustering

**Hong Yu** hongyu@dlut.edu.cn
**Tiantian Zhang** ztt_dlut@163.com
**Yahong Lian** lianyahong1@163.com
**Yu Cai** caiyu_ssdut@163.com
*Dalian University of Technology, Dalian, China*

**Editors:** Jun Zhu and Ichiro Takeuchi

## Abstract

For many clustering applications, Multi-view data sets are very common. Multi-view clustering aims to exploit information across views instead of individual views, which is promising to improve clustering performance. Note that a high-dimensional data set usually distributes on certain low-dimensional subspace. Thus, many multi-view subspace clustering algorithms have been developed. However, existing multi-view subspace clustering methods rarely perform clustering on the subspace representation of each view simultaneously as well as keep the indicator consistency among the representations, i.e., the same data point in different views should be assigned to the same cluster. In this paper, we propose a novel multi-view subspace clustering method. In our method, we use the indicator matrix to ensure that we perform clustering on the subspace representation of each view simultaneously. And at the same time, a co-regularized term is added to guarantee the consistency of the indicator matrices. Experiments on several real-world multi-view datasets demonstrate the effectiveness and superiority of our proposed method.

**Keywords:** Co-regularized, Multi-view Clustering, Subspace Clustering

## 1. Introduction

Clustering is a popular unsupervised learning technique in data mining. Traditional clustering aims to identify groups of "similar behavior" in single view data. As the real-world data are always captured from multiple sources or represented by several distinct feature views, multi-view data are very common in many real-world applications. Different views of data describe different features of data. For example, in the field of computer vision, images and videos can be represented by different types of features, such as color and texture information. In natural language processing tasks, the same document or corpus is available in multiple languages. Each language version can be used as a view. Web pages can also be represented using multi-view features based on text and hyperlinks. Each view may provide complementary information that other views do not have. Therefore, clustering with complementary information provided by multiple views would obtain better results than clustering only on a single view. The main obstacle of multi-view clustering is how to integrate the information provided by multiple views to achieve the purpose of improving the clustering performance. Many researches have been conducted to develop effective multi-view algorithms and significant results have been achieved (Kumar et al. (2011); Kumar and III (2011); Liu et al. (2013); Guo (2013); Cao et al. (2015); Nie et al. (2017)). Those methods usually focus on integrating multiple views into a low-dimensional representation and then performing spectral clustering on this low-dimensional representation.

Although a number of existing multi-view clustering methods have achieved considerable performance, these methods generally ignore the priori information of the dataset. For example, in face of clustering, many high-dimensional data are supposed to drawn from multiple low-dimensional subspaces, which means the intrinsic dimension of data is much lower than the original dimension, with the characteristics of sparse or low-rank. Taking these priori information into consideration will hopefully improve the clustering performance. In the past several years, subspace clustering has received extensive attention and a great number of algorithms have been proposed. Such as subspace clustering via least squares regression (LSR)(Lu et al. (2012); Shao et al. (2015)), sparse subspace clustering(SSC) (Elhamifar and Vidal (2013)), and subspace clustering based on low-rank representation (LRR) (Liu et al. (2010a); Li et al. (2016); Zhao et al. (2016)).

Some multi-view subspace clustering algorithms have also been proposed and made great progress. Yin et al. (2015) proposed a multi-view clustering method via pairwise sparse subspace representation, which consists of a weakly semi-supervised link constrained multi-view clustering and a special co-regularization based multi-view clustering method. The latter method tries to find a unified subspace representation. However, in theory, this method doesn't work well. Because although the data block structures in different subspace representations are similar, the magnitude of element values in them can be dramatically different. Gao et al. (2015) proposed a novel multi-view subspace clustering. Instead of computing a common subspace representation to unify subspace representation of each view, they proposed to integrate the clustering results using different subspace representations. Meanwhile, a separated spectral clustering post-processing step is been induced in order to achieve sub-optimal results. Cao et al. (2015) proposed a diversity-induced multi-view subspace clustering method utilizing the Hilbert Schmidt Independence Criterion to explore the complementarity of multi-view representations. In general, these approaches tend to obtain the subspace representation of each view firstly, then get the affinity matrix of the data set, and perform spectral clustering on the affinity matrix to obtain clustering results finally.

In this paper, we propose a novel multi-view subspace clustering method. Instead of separately obtaining subspace representation for each view and then merging them to perform clustering, we perform clustering on the subspace representation of each view simultaneously. And a pairwise co-regularization is developed to keep the consistency across the views. More specifically, we use an indicator matrix to learn the embedded consensus information of each view synthetically. In addition, a pairwise co-regularization constraint on indicator matrix is utilized to capture the interaction between the correlated clustering indicator matrices. We also develop an iterative optimization algorithm to solve the proposed framework. At last, extensive experiments on several real-world multi-view data sets demonstrate the effectiveness of our method.

In summary, the main contributions of our work include:

- We propose a novel multi-view subspace clustering method based on pairwise co-regularization to guarantees the consistency across views, considering the same data point in different views should have the same membership.

- Instead of learning subspace representations of each view first, then applying spectral clustering to subspace representations, we integrate learning representation and the preceding-step of spectral clustering into the objective equation. In other words, we use an indicator matrix to performing clustering on the subspace representation of each view simultaneously.
- We verify the effectiveness of our proposed method. Experimental results on six real-world datasets show our method outperforms other baseline algorithms.

The rest of this paper is organized as follows. In Section 2, we introduce subspace clustering and multi-view subspace clustering briefly. In Section 3, we describe the proposed co-regularized multi-view subspace clustering method in detail. Section 4 shows the optimization of our algorithm. Extensive experimental results and analysis are given in Section 5. Finally, we conclude this paper in section 6.

## 2. Related Work

In this section, we will introduce subspace clustering and multi-view clustering briefly.

### 2.1. Subspace Clustering

Subspace clustering naturally arises with the appearance of high-dimensional data. It is now widely known that many high dimensional data can be modelled as samples drawn from the union of multiple low-dimensional subspaces (Li and Vidal (2015); Gao et al. (2015)). Thus, the data points can be represented by a low-dimension subspace. Subspace clustering aims to allocate data in the same subspace into the same cluster according to the subspace structure of data (Liu et al. (2010b)).

Consider $n$ data points $X = \{x_1, x_2, ..., x_n\} \in \mathbb{R}^{d\times n}$, where each column represents a data vector, $d$ denotes the feature dimension. The subspace clustering is based on the self-representation property (Elhamifar and Vidal (2013)). That is, each data point in the space can be represented by other points in the same subspace. The matrix form is written as:

$$X = XZ + E \tag{1}$$

where $Z \in \mathbb{R}^{n\times n}$ is the subspace representation matrix, each column of $Z$ is a reconstruction coefficients of the original data point. $E \in \mathbb{R}^{d\times n}$ is the error matrix resulting from the possible corruptions in the representations. In order to make the presentation error as small as possible, the subspace clustering aims to find the self-representation by solving the following optimization problem:

$$\min_{E} ||E||_1 \quad s.t.\ X = XZ + E, Z(i,i) = 0, 1 \le i \le n \tag{2}$$

where $\|\cdot\|_1$ is $\ell_1$-norm, and the constraint $Z(i,i) = 0$ is to prevent the data point from expressing itself. After solving the problem (2), we can get the subspace representation matrix $Z$. The similarity matrix $S$ can be constructed as:

$$S = (|Z| + |Z|^T)/2 \tag{3}$$

where $|\cdot|$ is the absolute operator. Afterwards, we can perform spectral clustering on such subspace similarity matrix to obtain the final clustering results:

$$\min_{F} Tr(F^T(D-S)F) \quad s.t\ F^TF = I \tag{4}$$

Where $F \in \mathbb{R}^{n\times k}$ is the cluster indicator matrix, which can be given to the k-means algorithm to obtain cluster memberships. And $k$ is the number of clusters. $D \in \mathbb{R}^{n\times n}$ is a diagonal matrix whose diagonal elements are defined as $d_{ii} = \sum_j s_{ij}$, where $s_{ij}$ is the element of the $i$-th row and the $j$-th column of $S$.

### 2.2. Multi-view Subspace Clustering

The main challenge of multi-view subspace clustering is to integrate the features of individual views through subspace representations. Note that $Z_v$ is the subspace representation of $v$-th view, where $v = 1, 2, ..., l$, and $l$ is the number of views. In order to combine the multi-view subspace learning results, existing methods prefer to learn the subspace representation of each view first, then obtain an affinity matrix by using either one of $Z_v$ or the average of all $|Z_v|$. Finally, performing spectral clustering on such affinity matrix. The process of learning subspace representation and clustering are carried out separately (Yin et al. (2015); Wang et al. (2016)).

Instead of considering integrate the features of each view by adding various regular constraints to $Z_v$. We propose to perform clustering on the subspace representation of each view simultaneously. Specifically, a cluster indicator matrix $F_v \in \mathbb{R}^{n\times k}$ is propose to learn the subspace features of each view( with $i$-th row of $F_v$ mapping the original $i$'th sample to the $k$ dimensional embedding space). Thus we can define $K_v = F_vF_v^T$ as the similarity matrix of $F_v$. According to the the fact that the membership of a data point should be the same across all the views, the indicator matrices of pairwise views should be as similar as possible. Therefore we use the inner product of indicator matrices to measure the disagreement between clustering of two views:

$$D(F_v, F_w) = \|\frac{K_v}{\|K_v\|_F^2} - \frac{K_w}{\|K_w\|_F^2}\|_F^2 \tag{5}$$

where $\|\cdot\|_F$ is the Frobenius norm. Since $\|K_v\|_F^2 = k$ and $\|K_w\|_F^2 = k$. Ignoring the constant $k$, we can get:

$$D(F_v, F_w) = -Tr(F_vF_v^TF_wF_w^T) \tag{6}$$

where $Tr(\cdot)$ is the trace of a matrix.

## 3. Co-regularized Multi-view Subspace Clustering

Based on the above, we leverage the self-representation of data to get the subspace representation, graph regularization to extracting local manifold structure and pariwise co-regularization of similarity between $F_v$ and $F_w$ to keep the consistency across views. There-by, our co-regularization multi-view subspace clustering method can be expressed as follows:

$$\begin{aligned}&\min_{Z_v,E_v,F_v}\sum_v\|X_v-X_vZ_v\|_F^2+\lambda_1\sum_v Tr(F_v^T(D_v-S_v)F_v)-\lambda_2\sum_{v\neq w}Tr(F_vF_v^TF_wF_w^T)\\&s.t.\ Z_v(i,i)=0,F_v^TF_v=I,1\leq i\leq n,1\leq v\leq l\end{aligned}\tag{7}$$

In real applications, the data tend to be corrupted by noise. In order to improve the robustness of our method, we take outlying entries matrix $E_v$ into consideration. With the combination above, our final object function can be rewritten as follows:

$$\begin{aligned}&\min_{Z_v,E_v,F_v}\underbrace{\sum_v\|X_v-X_vZ_v-E_v\|_F^2}_{\text{Subspace representation}}+\underbrace{\lambda_1\sum_v Tr(F_v^T(D_v-S_v)F_v)}_{\text{Graph regularization}}\\&\underbrace{-\lambda_2\sum_{v\neq w}Tr(F_vF_v^TF_wF_w^T)}_{\text{Pairwise co-regularization}}+\underbrace{\lambda_3\sum_v\|E_v\|_1}_{\text{Noise robustness}}\\&s.t.\ Z_v(i,i)=0,F_v^TF_v=I,1\leq i\leq n,1\leq v\leq l\end{aligned}\tag{8}$$

Where $S_v=\frac{|Z_v|+|Z_v|^T}{2}$, $Z_v$ is the subspace representation matrix of the $v$-th view, $D_v$ is a diagonal matrix whose diagonal element are defined as $d_{ii}=\sum_j s_{ij}$, and the rows of matrix $F_v$ are the embeddings of the data points that can be given to the k-means algorithm to obtain cluster memberships.

## 4. Optimization Algorithm

The goal of optimization is to minimize Eq. (8). It is difficult to optimize Eq. (8) directly because Eq. (8) is non-convex. In this paper, we employ the alternative optimization to solve Eq. (8) by iteratively optimizing one variable with others fixed.

### 4.1. Update $F_v$

Fixing $Z_v$, $E_v$, update $F_v$.

$$\begin{aligned}&\min_{F_v}\lambda_1\sum_v Tr(F_v^T(D_v-S_v)F_v)-\lambda_2\sum_{v\neq w}Tr(F_vF_v^TF_wF_w^T)\\&s.t.\ Z_v(i,i)=0,F_v^TF_v=I,1\leq i\leq n,1\leq v\leq l\end{aligned}\tag{9}$$

We have

$$\begin{aligned}&\min_{F_v}Tr\left\{F_v^T\left((D_v-S_v)-(\frac{\lambda_2}{\lambda_1})\sum_{v\neq w}F_wF_w^T\right)F_v\right\}\\&s.t.\ Z_v(i,i)=0,F_v^TF_v=I,1\leq i\leq n,1\leq v\leq l\end{aligned}\tag{10}$$

Such a problem (10) is a standard spectral clustering object on view $v$ with graph Laplacian matrix $(D_v-S_v)-(\frac{\lambda_2}{\lambda_1})\sum_{v\neq w}F_wF_w^T$, whose solution are the eigenvectors corresponding to the smallest $k$ eigenvalue of the Laplacian matrix.

### 4.2. Update $Z_v$

Fixing $F_v$, $E_v$, update $Z_v$. We have

$$\begin{aligned} &\min_{Z_v} \sum_v \|X_v - X_v Z_v - E_v\|_F^2 + \lambda_1 \sum_v Tr\left(F_v^T (D_v - \frac{|Z_v| + |Z_v|^T}{2}) F_v\right) \\ &s.t.\ Z_v(i,i) = 0, F_v^T F_v = I, 1 \le i \le n, 1 \le v \le l \end{aligned} \tag{11}$$

there is a basic but significantly important equation in spectral analysis (Belkin and Niyogi (2002))

$$Tr(F^T L F) = \frac{1}{2} Tr(SY) \tag{12}$$

where $Y_{ij} = \|f^i - f^j\|$, $f^i$ is the $i$-th row of matrix $F$. So we have

$$\begin{aligned} &\min_Z \|X - XZ - E\|_F^2 + \frac{\lambda_1}{2} Tr(|Z|^T Y) \\ &s.t.\ Z_v(i,i) = 0, 1 \le i \le n, 1 \le v \le l \end{aligned} \tag{13}$$

When all rows except the $i$-th row fixed, updating the $i$-the row of $Z$:

$$\begin{aligned} &\min_z \|X_1 - xz^T\|_F^2 + \frac{\lambda_1}{2} |z|^T y \\ &s.t.\ z_i = 0, 1 \le i \le n \end{aligned} \tag{14}$$

where $z^T$ is the $i$-th row of $Z$, and $y$ is the $i$-th column of $Y$, and $X_1 = X - (XZ - xz^T) - E$. Eq. (14) differs Eq. (15) only by a constant.

$$\begin{aligned} &\min_z x^T x z^T z - 2z^T X_1^T + x + + \frac{\lambda_1}{2} |z|^T y \\ &s.t.\ z_i = 0, 1 \le i \le n \end{aligned} \tag{15}$$

Also, Eq. (15) differs the follow problem only by a constant:

$$\begin{aligned} &\min_z \|z - v\|_2^2 + \frac{\lambda_1}{2} |z|^T y \\ &s.t.\ z_i = 0, 1 \le i \le n \end{aligned} \tag{16}$$

where $\nu = \frac{X_1^T x}{x^T x}$. Thus Eq. (14) has the same solution with Eq. (16). In detail, if $k = i$, then $z_k = 0$. If $k \ne i$,we have :

$$\min_{z_K} \frac{1}{2}(z_k - \nu_k)^2 + \frac{\lambda_1 y_k}{4} |z_k| \tag{17}$$

Finally we have:

$$z_k = \begin{cases} \nu_k - \frac{\lambda_1 y_k}{4}, & if\ \nu_k > \frac{\lambda_1 y_k}{4} \\ \nu_k + \frac{\lambda_1 y_k}{4}, & if\ \nu_k < \frac{\lambda_1 y_k}{4} \\ 0, & otherwise \end{cases} \tag{18}$$

### 4.3. Update $E_v$

Fixing $F_v$, $Z_v$, update $E_v$. We have

$$\min_{E_v} \|X_v - X_v Z_v - E_v\|_F^2 + \lambda_3 \sum_v \|E_v\|_1 \tag{19}$$

For convenience, we ignore the subscript as follows:

$$\min_{E} \frac{1}{2}\|E - (X - XZ)\|_F^2 + \frac{\lambda_3}{2}\|E\|_1 \tag{20}$$

We solve the $i$-th column of $E$:

$$\min_{e_i} \frac{1}{2}\|e_i - (X - XZ)_i\|_2^2 + \frac{\lambda_3}{2}\|e_i\|_1 \tag{21}$$

Finally we have:

$$E_{ij} = \begin{cases} (X - XZ)_{ij} - \frac{\lambda_3}{2}, & if\ (X - XZ)_{ij} > \frac{\lambda_3}{2} \\ (X - XZ)_{ij} + \frac{\lambda_3}{2}, & if\ (X - XZ)_{ij} < \frac{\lambda_3}{2} \\ 0, & otherwise \end{cases} \tag{22}$$

According to the functions above, we can iteratively update the $Z_v$, $E_v$, $F_v$ until convergence. Finally applying k-means to the indicator matrix $F_v$ to get the final clustering result. The whole algorithm is summarized in **Algorithm 1**.

**Algorithm 1** Co-regularized Multi-view Subspace Clustering.

**Input**: Unlabeled multi-view data $D = \{X_1, X_2, ..., X_v\}$, parameters $\lambda_1$, $\lambda_2$, $\lambda_3$.
**Output**: Clustering result $C$.

```
Initialize Z_v = 0, E_v = 0, F is initialized to the result of spectral clustering.
while not converge do
    Update the i-th row of Z_v :
    if k = i then
        z_k = 0.
    end
    if k ≠ i then
        Update the i-th row of Z_v by Eq. (18).
    end
    Update the E_v by Eq. (22).
    Update the F_v by Eq. (10).
end
Apply k-means to F_v. if the j-th row of F_v is assigned to cluster c, example j is assigned
to cluster c.
```

## 5. Experiment

In this section, experiments on several real-world multi-view datasets are conducted to demonstrate the effectiveness and superiority of our proposed method.

### 5.1. Datasets

We evaluate our method on six benchmark datasets. The information of these datasets is summarized in Tab. 1.

Table 1: Information of the multi-view datasets.

| Datasets | Samples | Views | Clusters |
|---|---|---|---|
| Cora | 2708 | 2 | 7 |
| WebKB | 307 | 3 | 4 |
| 3-source | 169 | 3 | 6 |
| ALOI | 1080 | 4 | 10 |
| Reuters | 1200 | 5 | 6 |
| NUS | 900 | 5 | 6 |

- *Cora*:[1] It contains 2708 documents over 7 labels. It is made of 4 views (content, inbound, outbound, cites) on the same documents. We consider the following two views in our experiments: number of citations between documents and the term-document matrix.
- *WebKB*:[2] It consists of webpages collected from four universities: Texas, Cornell, Washington and Wisconsin. The web pages are classified into 7 categories. Here, we choose four most populous categories (course, faculty, project,student) for clustering. And a web page is made of three views: the text on it, the anchor text on the hyperlinks pointing to it and the text in its title.
- *3-sources*:[3] It is a text dataset collected form three online news sources: BBC, Reuters, and the Guardian. It contains a total of 948 news articles covering 416 distinct news stories. Among them, there are 169 articles reported in all three sources. Each story is correspond to one or more of six topical labels. We use all 169 news in our experiments, while each source is taken as one independent view of the story.
- *ALOI*:[4] It is a collection of 110250 images of 1000 small objects. We select 1080 samples on 10 classes with four views: RGB color histograms, HSB color histograms, color similarity and haralick features.
- *Reuters*:[5] It contains 1200 documents and each document is translated into five languages (English, French, German, Spanish and Italian). Each language version can be as a view. And it has six classes.
- *NUS*:[6] It is a web image dataset of National University of Singapore. The dataset contains 30000 images in 31 categories. Each picture can be represented by five

1. http://lig-membres.imag.fr/grimal/data.html
2. http://membres-liglab.imag.fr/grimal/data.html
3. http://mlg.ucd.ie/datasets/3sources.html
4. http://elki.dbs.ifi.lmu.de/wiki/DataSets/MultiView
5. http://membres-liglab.imag.fr/grimal/data.html
6. http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm

views: color histogram(CH), color correlation(CoRR), edge estimation(EDH), wavelet texture(WT) and block-wise color moment (CM). In our experiments, we use 900 samples with five views.

### 5.2. Experimental Settings

To evaluate the performance of our method, we compare our method with the following algorithms.

- PairwiseSC : Kumar et al. proposed the multi-view spectral clustering method to co-regularize pairwise eigenvectors of all views' Laplacian matrices and achieve consensus clusters across views (Kumar et al. (2011)). In our experiments, we set the parameter of this methond to 0.01 as the author suggested.

- CentroidSC : Kumar proposed the multi-view spectral clustering method to regularize each view-specific set of eigenvectors towards a common centroid eigenvector (Kumar et al. (2011)). The value of the parameter is the same as PairwiseSC.

- Co_training : According to the idea of co-training, learning the clustering in one view and use it to "label" the data in other views so as to modify the graph structure. And iterating this process until convergence (Kumar and III (2011)). In our experiments, we set the parameter of this methond to 1.5 as the author suggested.

- Multi_NMF : Gao et al. proposed a NMF-based multi-view clustering algorithm by searching for a factorization that gives compatible clustering solutions across multiple views (Gao et al. (2013)). The parameter of this method is set to 0.01 as authors suggested.

- RMSC : A robust multi-view spectral clustering method, which learns the shared transition probability matrices and their respective error matrices from each view. And then use the shared transition probability matrix as the input matrix of Markov chain (Xia et al. (2014)). As the author claims that when setting the parameter of this methond to 0.005, it work well in all of the datasets. So we also use the same value in our experiments..

- MVSC : A multi-view subspace clustering method, which performs clustering on the subspace representation of each view simultaneously (Gao et al. (2015)). Because the author did not give a suggestion of parameter settings in the paper, we adjusted the parameters and set the two parameters to 1.5 and 0.5 finally, which works better.

In this paper, for all comparison experiments, we use the source codes provided by the corresponding paper authors and follow the suggestions the authors have given. Each experiment is repeated 10 times and the average value are reported. For our method, we set parameters $\lambda_1 = 0.5$, $\lambda_2$=0.5 and $\lambda_3 = 0.05$, because the experimental results prove that they work well on all datasets. To measure the performance of our proposed algorithm, we use three widely used clustering metrics: clustering accuracy (ACC), normalized mutual information(NMI) and Purity. ACC measures the proportion of correct clustering. The definition of ACC is as follows:

$$ACC(\bar{y_i}, y_i) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{\bar{y_i} = y_i\} \tag{23}$$

where $N$ is the number of samples. $\bar{y_i}$ is the predicted result of the $i$-th sample and $y_i$ is the ground truth classification. $\mathbf{1}\{x\}$ is the indicator equation, when the predicted result is exactly the same as the true result, the value of the equation is 1, otherwise it is 0.

The definition of NMI is as follows:

$$NMI(\bar{y_i}, y_i) = \frac{MI(\bar{y_i}, y_i)}{\sqrt{H(\bar{y_i})H(y_i)}} \tag{24}$$

where $MI(\cdot)$ denotes the mutual information between $\bar{y_i}$ and $y_i$. $H(\cdot)$ denotes their entropy.

Purity is a simple and transparent evaluation measure. It only need to calculate the proportion of the number of correct clustering to the total number of samples. Given some set of clusters $M = \{m_1, m_2, ..., m_K\}$, a set of classes $D = \{d_1, d_2, ..., d_J\}$, and $N$ data points, purity can be defined as:

$$Purity = \frac{1}{N}\sum_{k}\max_{j}|m_k \cap d_j| \tag{25}$$

where $m_k$ indicates the set of samples in the $k$-th cluster and $d_j$ indicates the set of samples in the $j$-th class. The advantage of the purity method is that it is easy to calculate and the value is between 0 and 1.

### 5.3. Experimental Results

The experimental results in terms of ACC, NMI and Purity on six real-world datasets are reported in Tab. 2, Tab. 3 and Tab. 4. The larger value of the three metrics, the better clustering performance we get. From Tab. 2, Tab. 3 and Tab. 4, we can observe that our method outperforms other algorithms.

Table 2: Clustering results in terms of ACC(%) on five datasets.

| **Methods** | **Cora** | **WebKB** | **3-source** | **ALOI** | **Reuters** | **NUS** |
|---|---|---|---|---|---|---|
| PairwiseSC | 37.03 | 70.39 | 54.67 | 80.43 | 50.01 | 26.93 |
| CentroidSC | 32.17 | 72.08 | 54.79 | 80.95 | 50.35 | 34.57 |
| Co_training | 52.60 | 57.65 | 54.69 | 80.94 | 43.96 | 32.66 |
| Multi_NMF | 41.36 | 68.70 | 47.93 | 83.72 | 49.14 | 34.53 |
| MVSC | 48.24 | 80.90 | 57.52 | 52.27 | 37.36 | 29.78 |
| RMSC | 38.82 | 64.46 | 48.58 | 71.65 | 53.47 | 36.52 |
| Co_MVSC | **58.12** | **85.49** | **62.37** | **85.91** | **56.16** | **58.12** |

PairwiseSC, CentroidSC , Co_training are three classic multi-view clustering algorithms. RMSC is a multi-view spectral clustering method. They all focus on how to integrate the feature information of multiple views without considering the priori information of the

Table 3: Clustering results in terms of NMI(%) on five datasets.

| Methods | Cora | WebKB | 3-source | ALOI | Reuters | NUS |
|---|---|---|---|---|---|---|
| PairwiseSC | 17.14 | 29.29 | 44.81 | 80.95 | 31.24 | 6.23 |
| CentroidSC | 16.24 | 57.81 | 47.64 | 80.65 | 30.81 | 13.68 |
| Co_training | 32.15 | 47.99 | 55.41 | 79.81 | 27.12 | 11.67 |
| Multi_NMF | 28.75 | 51.81 | 47.92 | 81.11 | 29.91 | 12.17 |
| MVSC | 22.53 | 55.45 | 43.66 | 50.54 | 17.26 | 12.53 |
| RMSC | 15.75 | 41.74 | 45.08 | 66.42 | 35.08 | **14.55** |
| Co_MVSC | **35.27** | **62.92** | **57.93** | **81.12** | **35.23** | 13.74 |

dataset. The clustering results are good, but they do not consider the priori information of the datasets. Our proposed method Co_MVSC not only keeps the the consistency across views, but also considers the priori subspace information. MVSC is a multi-view subspace algorithm. From the results in the tables, we find this method is not very stable. It could get good results on some datasets, such as Cora, WebKB, 3-source, while does not work well on others like ALOI, Reuters and NUS, which have more views. Although MVSC uses the information of the data distributed in a certain low-dimensional subspace, it dose not fully exploit the complementary information across the views. This leads to more views not necessarily guaranteeing better clustering performance. Our Co_MVSC method uses a pair co-regularization to explore the complementary information embedded in the multi-view data. And the results show that the performance of our method generally outperforms other algorithms.

Table 4: Clustering results in terms of Purity(%) on five datasets.

| Methods | Cora | WebKB | 3-source | ALOI | Reuters | NUS |
|---|---|---|---|---|---|---|
| PairwiseSC | 42.79 | 71.30 | 65.80 | 83.44 | 50.82 | 27.48 |
| CentroidSC | 40.30 | 83.26 | 67.92 | 82.93 | 50.97 | 36.59 |
| Co_training | 57.34 | 78.86 | 72.38 | 83.58 | 44.77 | 33.98 |
| Multi_NMF | 50.43 | 79.69 | 66.98 | 85.15 | 50.03 | 36.12 |
| MVSC | 48.86 | 82.57 | 63.91 | 53.69 | 38.75 | 31.22 |
| RMSC | 40.22 | 76.06 | 71.85 | 72.43 | 56.12 | 38.01 |
| Co_MVSC | **60.71** | **85.49** | **74.91** | **87.68** | **57.07** | **38.19** |

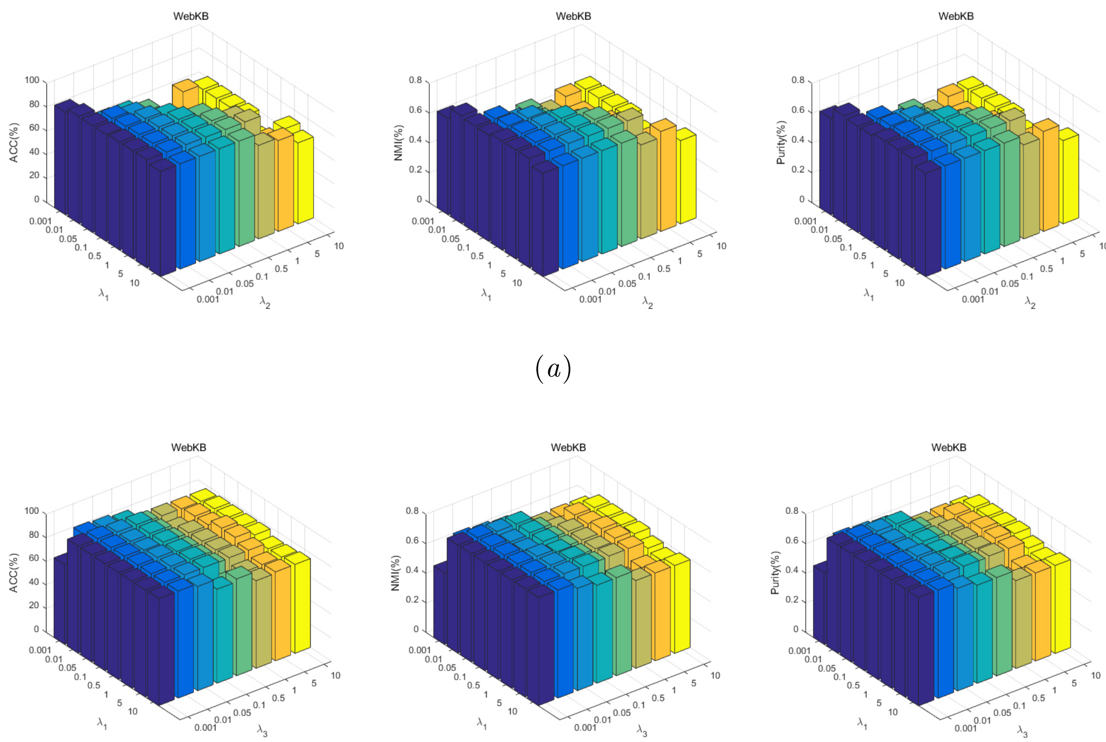
In order to show how the performance of our algorithm changes with the increase of the number of views, we apply our method to increasing-view features on the ALOI data set. The results are shown in Tab. 5. From the results, we can observe that with the increase of views, the value of ACC, NMI and Purity increase, which indicates that the clustering performance of combining features of multiple views is better than only relying on a single view. And the more views means the more complementary information is provided, thus the better clustering performance will be obtained.

Table 5: Increasing views on ALOI dataset.

| No.of views | ACC | NMI | Purity |
|---|---|---|---|
| One | 57.69 | 52.88 | 59.64 |
| Two | 66.46 | 63.40 | 69.96 |
| Three | 76.95 | 73.77 | 78.62 |
| Four | 85.91 | 81.12 | 87.68 |

### 5.4. Parameter Selection

In our proposed co-regularized multi-view subspace clustering method, there are three parameters $\lambda_1$, $\lambda_2$, and $\lambda_3$. Among them, $\lambda_1$ controls the graph regularization term to characterize the local manifold structure. $\lambda_2$ controls the pairwise co-regularization of similarities between two views. $\lambda_3$ controls the possible noise contained by the datasets.



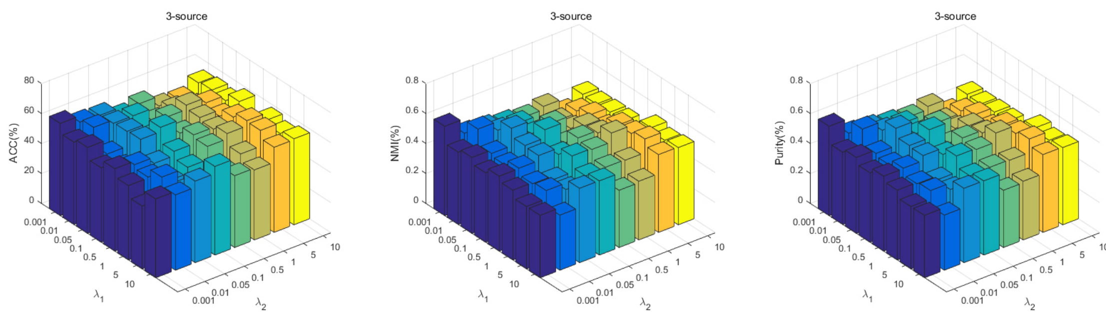

$(a)$

$(b)$

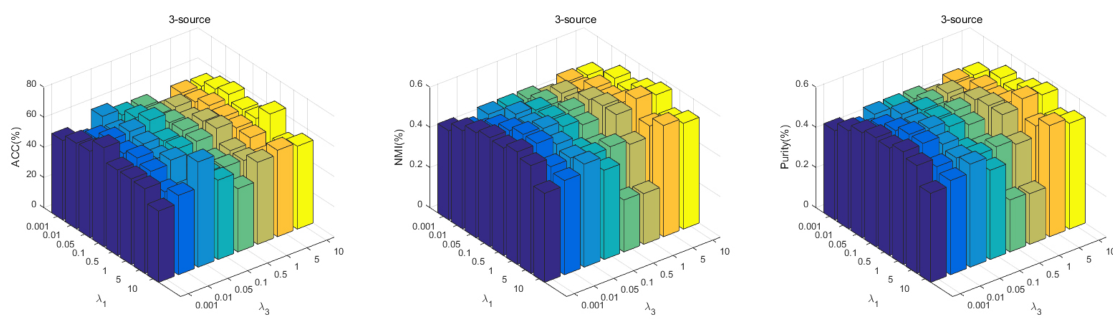
Figure 1: Parameters study of our proposed method on WebKB dataset. (a) ACC, NMI and Purity against parameters $\lambda_1$ and $\lambda_2$. (b) ACC, NMI and Purity against parameters $\lambda_1$ and $\lambda_3$

In order to determine the range of parameters, we first test a relatively large interval [0.001,10]. Firstly, we fix the value of $\lambda_3$ with any value in the interval $[0.001, 10]$, then test the change of each metric with the varying of $\lambda_1$ and $\lambda_2$ on WebKB. From Fig. 1$(a)$, we have the following observations. When $\lambda_1$ it too small, the graph regularization term will lose its effect. When $\lambda_1$ it too big, the effect of other regular terms will be weakened, which will affect the clustering result. And so does $\lambda_2$ . Therefore according to the experimental results, we choose $\lambda_1$ and $\lambda_2$ form the interval [0.05,1]. Secondly, we fix the value $\lambda_2$ to a suitable and increase the value of $\lambda_1$ and $\lambda_3$ in order to determine the range of $\lambda_3$. From Fig. 1$(b)$, we can see the value of $\lambda_3$ should not be too large. This shows that noise has an impact on clustering, but it is not a major factor. So in our experiments, $\lambda_3$ is chosen from the interval [0.001,0.1].

In order to further prove the correctness of our analysis, we do the same experiments on the dataset 3-source. From Fig. 2, it can be seen that the value ranges of $\lambda_1$, $\lambda_2$, $\lambda_3$ are feasible. Finally, we set $\lambda_1 = 0.5$, $\lambda_2 = 0.5$, $\lambda_3 = 0.05$ in our experiments , because they work well on all datasets.




$(a)$




$(b)$

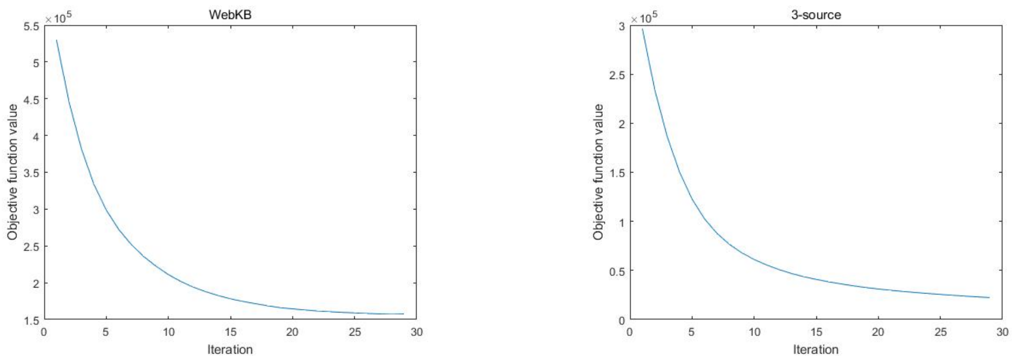
Figure 2: Parameters study of our proposed method on 3-source dataset. (a) ACC, NMI and Purity against parameters $\lambda_1$ and $\lambda_2$. (b) ACC, NMI and Purity against parameters $\lambda_1$ and $\lambda_3$

### 5.5. Convergence analysis

In our experiments, the stop criteria is defined as follows:

$$\frac{|f^{(t+1)} - f^{(t)}|}{f^{(t)}} < 10^{-1} \tag{26}$$

where $f^{(t)}$ denotes the objective value of the $t$-th iteration. Fig. 3 shows the change in objective value with each iteration. Fig. 3, we can see that our proposed method converges quickly.




$(a)$ $(b)$

Figure 3: The convergence curve of our proposed method

## 6. Conclusion

In this paper, we propose a novel multi-view subspace clustering method. Instead of learning subspace representations of each view first, then applying spectral clustering to subspace representations, we integrate learning representation and the preceding-step of spectral clustering into the objective equation. In other words, we use an indicator matrix to performing clustering on the subspace representation of each view simultaneously. And at the same time a co-regularized term is utilized to guarantee the consistency of the indicator matrices. Comparative experiments with six state-of-art algorithms on six multi-view datasests demonstrate the effectiveness of our algorithm.

## Acknowledgments

Research reported in this publication was supported by the National Natural Science Foundation of China(61602081) and Natural Science Foundation of Liaoning Province(201602180).

## References

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in neural information processing systems*, pages 585–591, 2002.

Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang. Diversity-induced multi-view subspace clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–594, 2015.

Ehsan Elhamifar and Rene Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE transactions on pattern analysis and machine intelligence*, 35(11): 2765–2781, 2013.

Hongchang Gao, Feiping Nie, Xuelong Li, and Heng Huang. Multi-view subspace clustering. In *Proceedings of the IEEE international conference on computer vision*, pages 4238–4246, 2015.

Jing Gao, Jiawei Han, Jialu Liu, and Chi Wang. Multi-view clustering via joint nonnegative matrix factorization. In *Proceedings of the 13th SIAM International Conference on Data Mining, May 2-4, 2013. Austin, Texas, USA.*, pages 252–260, 2013.

Yuhong Guo. Convex subspace representation learning from multi-view data. In *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, July 14-18, 2013, Bellevue, Washington, USA.*, 2013.

Abhishek Kumar and Hal Daume III. A co-training approach for multi-view spectral clustering. In *International Conference on International Conference on Machine Learning*, pages 393–400, 2011.

Abhishek Kumar, Piyush Rai, and Hal Daume. Co-regularized multi-view spectral clustering. In *Advances in neural information processing systems*, pages 1413–1421, 2011.

Chun Guang Li and Ren Vidal. Structured sparse subspace clustering: A unified optimization framework. In *Computer Vision and Pattern Recognition*, pages 277–286, 2015.

Jun Li, Yu Kong, Handong Zhao, Jian Yang, and Yun Fu. Learning fast low-rank projection for image classification. *IEEE Transactions on Image Processing*, 25(10):4803–4814, 2016.

Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *arXiv preprint arXiv:1010.2955*, 2010a.

Guangcan Liu, Zhouchen Lin, and Yong Yu. Robust subspace segmentation by low-rank representation. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 663–670, 2010b.

Jialu Liu, Chi Wang, Jing Gao, and Jiawei Han. Multi-view clustering via joint nonnegative matrix factorization. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, pages 252–260. SIAM, 2013.

Can-Yi Lu, Hai Min, Zhong-Qiu Zhao, Lin Zhu, De-Shuang Huang, and Shuicheng Yan. Robust and efficient subspace segmentation via least squares regression. In *European conference on computer vision*, pages 347–360. Springer, 2012.

Feiping Nie, Guohao Cai, and Xuelong Li. Multi-view clustering and semi-supervised classification with adaptive neighbours. In *AAAI*, pages 2408–2414, 2017.

Ming Shao, Sheng Li, Zhengming Ding, and Yun Fu. Deep linear coding for fast graph clustering. In *International Conference on Artificial Intelligence*, pages 3798–3804, 2015.

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. *arXiv preprint arXiv:1608.05560*, 2016.

Rongkai Xia, Yan Pan, Lei Du, and Jian Yin. Robust multi-view spectral clustering via low-rank and sparse decomposition. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 2149–2155, 2014.

Qiyue Yin, Shu Wu, Ran He, Liang Wang, et al. Multi-view clustering via pairwise sparse subspace representation. neurocomputing. 2015.

Handong Zhao, Zhengming Ding, and Yun Fu. Pose-dependent low-rank embedding for head pose estimation. In *Thirtieth AAAI Conference on Artificial Intelligence*, pages 1422–1428, 2016.