Neurocomputing ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Multi-view clustering via pairwise sparse subspace representation

Qiyue Yin, Shu Wu, Ran He, Liang Wang*

Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, PR China

ARTICLE INFO

Article history:
Received 26 September 2014
Received in revised form
4 January 2015
Accepted 4 January 2015
Communicated by Yi Zhe Song

Keywords:
Multi-view clustering
Subspace clustering
Pairwise sparse representation

ABSTRACT

Multi-view clustering, which aims to cluster datasets with multiple sources of information, has a wide range of applications in the communities of data mining and pattern recognition. Generally, it makes use of the complementary information embedded in multiple views to improve clustering performance. Recent methods usually find a low-dimensional embedding of multi-view data, but often ignore some useful prior information that can be utilized to better discover the latent group structure of multi-view data. To alleviate this problem, a novel pairwise sparse subspace representation model for multi-view clustering is proposed in this paper. The objective function of our model mainly includes two parts. The first part aims to harness prior information to achieve a sparse representation of each high-dimensional data point with respect to other data points in the same view. The second part aims to maximize the correlation between the representations of different views. An alternating minimization method is provided as an efficient solution for the proposed multi-view clustering algorithm. A detailed theoretical analysis is also conducted to guarantee the convergence of the proposed method. Moreover, we show that the must-link and cannot-link constraints can be naturally integrated into the proposed model to obtain a link constrained multi-view clustering model. Extensive experiments on five real world datasets demonstrate that the proposed model performs better than several state-of-the-art multi-view clustering methods.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Many kinds of real-world data appear in multiple views. For example, web pages contain both images and corresponding texts, and images can be encoded by different features such as color histogram and Fourier shape descriptors. Although learning tasks such as classification and clustering can be approached based on one single view, multiple views providing complementary information can improve the performance of learning tasks [1]. This leads to a surge of interest in multi-view learning, whose goal is to exploit multiple views to obtain better performance rather than relying on every single view. Till now, multi-view learning has been widely studied in different areas such as data mining, multimedia, computer vision and natural language processing [2–5].

As one of the basic tasks of multi-view learning, multi-view clustering has attracted more and more attention because it can handle large numbers of unlabeled datasets. The objective of multi-view clustering is to cluster multi-view datasets based on their latent groups. Generally, the main challenge lies in how to make use of the complementary characteristics embedded in the multiple sources of information. Plenty of multi-view clustering algorithms have been developed to solve this problem. Some methods aim to find a unified low-dimensional embedding to fuse the multi-view representations, and clustering is then performed when the unified representation is obtained [6,7]. These methods often map the original high-dimensional feature space to a latent low-dimensional space so as to well explore the feature correlation between different views. On the other hand, some methods perform multi-view clustering through merging the clustering results from different individual views [8,9]. These methods, called late fusion, obtain the final clustering results by voting or other fusion strategies. For more details about multi-view clustering, refer to Section 2.

Although various existing methods indeed improve the clustering performance for multi-view data, they often do not take some useful prior knowledge into consideration, such as collaborative [10], sparse [11] and low-rank [12] information, which has been shown to be helpful for clustering in some data mining applications. On the other hand, spectral-based subspace clustering methods [13] are recently developed, which can take advantage of such prior information and achieve promising results. These methods bring in different prior

* Corresponding author.
E-mail addresses: qyyin@nlpr.ia.ac.cn (Q. Yin), shu.wu@nlpr.ia.ac.cn (S. Wu), rhe@nlpr.ia.ac.cn (R. He), wangliang@nlpr.ia.ac.cn (L. Wang).

http://dx.doi.org/10.1016/j.neucom.2015.01.017
0925-2312/© 2015 Elsevier B.V. All rights reserved.

knowledge to constrain the self-representation matrix of the dataset to be the ideal block diagonal matrix, and then use the spectral method to obtain the final clustering results. Moreover, algorithms in this class can well discover the relationship between data points and reflect the latent group structure of the dataset. However, these methods often focus on the single view data and could not been directly applied for multi-view datasets.

Inspired by the recent advances in subspace clustering, this paper proposes a novel multi-view clustering framework based on sparse subspace representation. The proposed model resorts to subspace clustering for efficiently using the prior knowledge compared with conventional multi-view clustering methods. Besides, pairwise co-regularization is developed to explore the complementary information embedded in the multi-view data. More specifically, the sparse representation of the dataset for each view is firstly constructed. At the same time, a pairwise co-regularization constraint is utilized to capture the interaction between the correlated view-specific sparse representations. Then, we develop an iterative algorithm to efficiently solve the proposed framework, and provide rigid theoretical analysis on the convergence of this algorithm. Moreover, we discuss the impacts of the proposed different co-regularization forms in exploring the correlation between views. In addition, we show that the link prior can be easily integrated into our proposed model and a link constrained multi-view clustering method is accordingly developed. Extensive experiments are conducted to demonstrate the effectiveness of the proposed methods.

Main contributions in this paper are summarized as follows:

(1) A novel pairwise co-regularization model is proposed for the multi-view clustering problem. It harnesses the prior information to obtain the view specific sparse representation and meanwhile utilizes the correlation between different views. Besides, different co-regularization forms are discussed as special examples in our framework.
(2) A novel link constrained multi-view clustering algorithm is developed to naturally integrate the partially observed supervisory information (e.g., must-link and cannot-link). To the best of our knowledge, this is rarely studied in the literature of multi-view clustering.
(3) We verify the effectiveness of the proposed multi-view clustering algorithms with extensive experiments on five real world datasets, achieving state-of-the-art results in terms of accuracy and normalized mutual information.

The rest of this paper is organized as follows. In Section 2, we briefly review multi-view clustering and subspace clustering algorithms. Then our multi-view sparse subspace clustering method is introduced in Section 3. Section 4 gives the extensions of our multi-view clustering model. Extensive experimental results and analysis are given in Section 5. Finally, Section 6 concludes the paper.

2. Related work

In this section, we briefly introduce the background of our proposed model, which consists of multi-view clustering and subspace clustering.

2.1. Multi-view clustering

Multi-view clustering, which aims to cluster the dataset with multiple views, can be roughly classified into three categories based on the usage of multiple sources of information in the clustering process [1,14]. Algorithms in the first category find a unified low-dimensional embedding of multi-view data, and then cluster the dataset using this representation like the single view clustering methods [15,6,2,16–19]. These methods, also called subspace learning-based methods, are widely studied. Kumar et al. [6] proposed a co-regularization framework to regularize the difference between view-specific Laplacian embeddings. Liu et al. [7] developed a multi-view non-negative matrix factorization framework to gain a consensus low-dimensional feature matrix from the original high-dimensional data, and He et al. [20] further improved the idea of multi-view non-negative matrix factorization based clustering algorithms. Recently, Wang et al. [3] proposed a regression-like clustering method, which directly obtains the final consensus label matrix.

The second category directly integrates the information of different views in the clustering process. Popular examples are the co-EM clustering algorithm [21] and the co-training framework [22–24]. Kumar and Daume [23] resorted to the co-training framework, which is widely used in the semi-supervised learning, to design the first co-training based multi-view spectral clustering algorithm. Zhao et al. [22] combined LDA, K-means with the co-training framework and developed a subspace co-training framework for the multi-view clustering task. In contrast, the third category is late fusion (or called ensemble clustering). That is, the final clustering result is derived from integrating each individual clustering result [25,9,26]. Long et al. [8] proposed to use mapping functions to make clusters from different views comparable and learn the best clusters from these multiple views. Greene and Cunningham [9] developed a matrix factorization based method to group the clusters obtained from each view.

Overall, these multi-view clustering methods indeed improve clustering performance for multi-view datasets. However, they rarely consider some useful prior knowledge, such as sparse or low rank information of the latent group structure, which has been shown to be helpful for clustering in some data mining applications.

2.2. Subspace clustering

Subspace clustering aims to cluster the high-dimensional data into multiple subspaces as well as find the subspaces fitting each group of data points. Generally it can be divided into four categories based on different techniques [13], and our discussion mainly focuses on the recently developed spectral-based subspace clustering methods [27,28,10,29]. The key idea of these approaches is to obtain a self-representation matrix by taking different prior information into consideration. Usually, these prior information is utilized as different constraints to achieve different self-representation matrices. Then the above matrix is applied to construct the affinity matrix, which is used for the final spectral clustering. Examples lie in this category are briefly introduced as follows.

Sparse Subspace Clustering (SSC) [11,28] is based on the fact that each point in a union of subspaces can be written as a linear (affine) combination of points belonging to the same subspace. Thus, the representation coefficients for a data point should be sparse, and this prior information is brought into the model by using an l_1-norm to constrain the representation coefficients. Different from SSC, Low Rank Representation (LRR) based subspace segmentation algorithms [12,27] seek the lowest rank representation for all points. The prior information lies in the low rank characteristic of the optimal representation matrix. Whereas, the Multi-Subspace Representation (MSR) based subspace segmentation methods [30,31] regularize the representation matrix to be both low rank and sparse. By a careful parameter configuration, the subspace structure can be well revealed. Least Squares Regression (LSR) [10,32] applies ridge regression to

Please cite this article as: Q. Yin, et al., Multi-view clustering via pairwise sparse subspace representation, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.01.017

construct the representation of all points and it can be proven that LSR has a good grouping effect that tends to group highly correlated data together.

Although many spectral-based subspace clustering methods are recently developed and make use of prior information to obtain promising clustering results, they often focus on single view data and have rarely been applied to multi-view datasets. Hence, in the following sections, we employ the idea of subspace clustering methods for the problem of multi-view clustering.

3. Multi-view sparse subspace clustering (MultiSSC)

Assume that we have data $X^v=[(X^1)^T,(X^2)^T,\ldots,(X^l)^T]\in R^{d_v\times n}$ with l views sampled from c classes and $X^v=[x_1^v,\ x_2^v,\ldots,\ x_n^v]\in R^{d_v\times n}$ denotes the vth view of X with dimension d_v. Our goal is to group the n data points into their corresponding classes. Here, we make two assumptions: (1) Each view is sufficient for the clustering problem, which indicates that $X^v(v=1,2,\ldots,l)$ has complete feature representations. (2) The underlying clustering would assign a point to the same cluster irrespective of the view, which means that the true class labels for x_i^v and x_i^w $(v\neq w)$ should be the same. This is a natural and reasonable assumption to avoid ambiguity in the clustering results.

In the following, we first formulate the *multi-view sparse subspace representation* in Section 3.1, which learns view specific sparse representation by taking the multi-view characteristic into consideration. Then we give the solution of the *multi-view sparse subspace representation* in Section 3.2. Section 3.3 gives the complete *multi-view sparse subspace clustering algorithm* (MultiSSC). Convergence and computational complexity analysis of the proposed MultiSSC model are then described in Section 3.4. Finally, the variants of MultiSSC using different co-regularization forms are also developed to explore the correlation between different views.

3.1. Multi-view sparse subspace representation

The key step in our model is to construct the sparse representation of each view and maximize the correlation between views based on the unified latent group structure. Then the learned sparse representations are utilized to achieve the final clustering.

To begin with, we give the key formulation of the sparse representation based subspace clustering method. Using the above notation, we have

$$\min_{Z^v}\|X^v-X^vZ^v\|_F^2+\alpha\|Z^v\|_1\quad \text{s.t } \operatorname{diag}(Z^v)=0 \tag{1}$$

where $Z^v\in R^{n\times n}$ is the sparse representation matrix, with each column being the reconstruction coefficients for its corresponding data point. And the regularizer $\|.\|_1$ forces the target points to be reconstructed using the samples belonging to the same subspace. $\alpha$is a positive parameter trading off the reconstruction error and the sparse constraint. diag(C) is the diagonal elements of matrix C, and the zero constraint on the sparse representation matrix is used to avoid trivial solution.

Multi-view data often consist of multiple heterogeneous representations, and cannot be directly adapted to the above model. To handle the multi-view data and resort to the subspace clustering model, we need to maximize the correlation of sparse representations between different views. Generally, the values of the sparse representation can be used to measure the similarity between data points and reflect the group structure. Under our assumption, the latent group structure of different views should be the same. So a general framework that can explore the correlation among different views is written as

$$\min_{Z^v}\sum_v\|X^v-X^vZ^v\|^2+\alpha\sum_v\|Z^v\|_1+\beta\varphi(Z^1,\ldots,Z^l)$$
$$\text{s.t. } \operatorname{diag}(Z^v)=0,\quad \forall v\in\{1,\ldots,l\} \tag{2}$$

where $\varphi(Z^1,\ldots,Z^c)$ is the regularizer to utilize multiple sources of information. α and $\beta$are two positive parameters balancing the three items, and the other variables are the same as in Eq. (1).

Pairwise constraint, which has been designed to restrict the coupled relationship, can be utilized in Eq. (2) to explore the view–view correlation. So some specifically designed pairwise co-regularization constraints can be used to make use of the multiple sources of information. We discussed this in Section 5.7. Here, a feasible objective with pairwise constraint can be formulated as

$$\min_{Z^v}\sum_v\|X^v-X^vZ^v\|_F^2+\alpha\sum_v\|Z^v\|_1+\beta\sum_{1\le v<w}\|Z^v-Z^w\|_1$$
$$\text{s.t. } \operatorname{diag}(Z^v)=0,\quad \forall v\in\{1,\ldots,l\} \tag{3}$$

where $\|Z^v-Z^w\|_1$ is an l_1-norm based pairwise co-regularization constraint to be explained in the next paragraph. It should be noted that we can learn pairwise specific importance parameter for the co-regularization term in Eq. (3) like the method proposed in [2], but here we just use the same β for simplicity.

The real-world data, especially those on the web, are often polluted for various reasons. Thus view-specific noise and outliers lead to the corruptions of the representation matrix. For example, when data point x_i^v is seriously polluted, the ith column of Z^v will be affected, leading to negative influence on the interaction between view-specific reconstruction coefficients. So we propose an l_1-norm based pairwise constraint $\|Z^v-Z^w\|_1$ to alleviate the corruption problem and meanwhile to make use of the multiple sources of information.

3.2. Solution to multi-view sparse subspace representation

In this subsection, we give the optimization method for the proposed formulation of multi-view sparse subspace representation. Since it is difficult to estimate all the variables at the same time, we propose an iterative updating algorithm.

We iteratively solve Z^v with Z^w $(w\neq v)$ fixed. Then Eq. (3) can be rewritten as

$$\min_{Z^v}\|X^v-X^vZ^v\|_F^2+\alpha\|Z^v\|_1+\beta\sum_{w\neq v}\|Z^v-Z^w\|_1$$
$$\text{s.t. } \operatorname{diag}(Z^v)=0 \tag{4}$$

To eliminate the equality constraint and make Eq. (4) easy to be optimized, we solve Z^v once per column. Thus Eq. (4) is rewritten as

$$\min_{Z_i^v}\|x_i^v-X_{-i}^vZ_i^v\|^2+\alpha\|Z_i^v\|_1+\beta\sum_{w\neq v}\|Z_i^v-Z_i^w\|_1 \tag{5}$$

where Z_i^v is the ith column of Z^v[1] and X_{-i}^v represents all data points excluding x_i^v in the vth view.

Since Eq. (5) consists of two non-smooth regularization terms, it is not easy to obtain the analytic solution directly. Similar to [3,33], an iteratively re-weighted method is applied to solve the problem. Taking the derivative of Eq. (5) with respect to Z_i^v, we have

$$(X_{-i}^v)^TX_{-i}^vZ_i^v-(X_{-i}^v)^TX_i^v+\alpha M_vZ_i^v+\beta\sum_{w\neq v}N_wZ_i^v=0 \tag{6}$$

where M_v and $N_w(w\neq v)$ are diagonal matrices with their jth diagonal elements calculated as[2]

$$(M_v)_{jj}=\frac{1}{2|Z_i^v|_j}(N_w)_{jj}=\frac{1}{2|Z_i^v-Z_i^w|_j} \tag{7}$$

where $|.|_j$ is the jth absolute value of the target vector. And the

[1] After obtaining Z_i^v, we insert a zero element to the ith position of Z_i^v and it will then consist of the ith column of Z^v as claimed in Eq. (4).

[2] When $|Z_i^v|_j=0$, Eq. (5) is not differentiable. And a small perturbation is introduced as in [34] to smooth the objective as $1/(2\sqrt{|Z_i^v|_j^2+\zeta})$. Similarly, when $|Z_i^v-Z_i^w|_j=0$, $1/(2|Z_i^v-Z_i^w|_j)$ can be reformulated as $1/(2\sqrt{|Z_i^v-Z_i^w|_j^2+\zeta})$. In our following experiments, we set $\zeta=1e-8$.

Please cite this article as: Q. Yin, et al., Multi-view clustering via pairwise sparse subspace representation, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.01.017

solution of Eq. (6) is

$$Z_i^v = \left((X^v_{-i})^T X^v_{-i} + \alpha M_v + \beta \sum_{w \neq v} N^v\right)^{-1} \left((X^v_{-i})^T X_i^v + \beta \sum_{w \neq v} N_w Z_i^w\right) \tag{8}$$

Note that M_v and N_w are dependent on Z_i^v, so we can iteratively solve M_v, N_w and Z_i^v. After an iteration of M_v, N_w and Z_i^v, we use the same method to solve Z_i^w $(w \neq v)$, and the whole procedure is repeated until convergence. Finally, we obtain $Z^v (v = 1, \ldots, l)$ after all their columns being calculated. The whole algorithm is summarized in Algorithm 1.

Algorithm 1. Solving Eq. (3) for multi-view sparse subspace representation.

Input:

Multi-view dataset $X = \{X^1; X^2; \ldots; X^l\}$, parameter α and β.

1: $t = 1$. Initialize $(Z^v)^t = 0 (v = 1, 2, \ldots, l)$;
2: **for** $i = 1 : n$ **do**
3: **while** not converge
4: Calculate M_v^{t+1} and $N_w^{t+1} (w = 1, \ldots, l; w \neq v)$ using Eq. (7);
5: Solve $(Z_i^v)^{t+1}$ using Eq. (8);
6: Repeat 4 and 5 to solve $(Z_i^w)^{t+1} (w \neq v))$ with the calculated $(Z_i^v)^{t+1}$ to replace $(Z_i^v)^t$;
7: $t = t + 1$;
8: **end while**
9: **end for**

Output:

Sparse representation matrices $Z^v (v = 1, 2, \ldots, l)$.

3.3. Algorithm of MultiSSC

After solving Eq. (3) as in Algorithm 1, we obtain the sparse representation matrices $Z^v \in R^{n \times n} (v = 1, 2, \ldots, l)$ for all the views. Generally, the elements of the representation matrices reflect the pairwise relationship between data points of different views. Similar to the spectral based subspace clustering method, we use them to construct the affinity matrix and then apply the spectral clustering algorithms such as the Normalized Cuts (Ncuts) [35] to group the multi-view dataset. We make the affinity matrix symmetric by setting $A = \frac{1}{2}(|Z|^T + l|Z|)$ like [12,10]. Similar to [6], we can use either one of $Z^v (v = 1, 2, \ldots, l)$ or the average of all $|Z^v|$ to obtain the affinity matrix. This procedure is summarized in Algorithm 2.

Algorithm 2.

Input:

Multi-view data $X = \{X^1; X^2; \ldots; X^l\}$ and the number of clusters c.

1: Obtain the multi-view sparse subspace representation $Z^v (v = 1, 2, \ldots, l)$ by Algorithm 1.
2: Define the affinity matrix: $A = \frac{1}{2}(|Z^v|^T + |Z^v|)$;
3: Apply the Ncuts [35] to the affinity matrix A.

Output:

c groups of the multi-view dataset X.

3.4. Convergence and computational complexity

Convergence analysis: To prove the convergence of Eq. (3), it should be guaranteed that Algorithm 1 decreases the objective value in each iteration. Specifically, once solving $(Z_i^v)^{t+1}$ with $(Z_i^w)^t (w \neq v)$ fixed,[3] Step 5 in Algorithm 1 will decrease the objective value. To simplify the notation, we use z_v, z_w, y and Y to represent Z_i^v, Z_i^w, x_i^v and X^v_{-i} respectively.

According to Step 5 in Algorithm 1, we can derive that

$$z_v^{t+1} = \min_{z_v} \|y - Y z_v\|^2 + \alpha z_v^T M_v^{t+1} z_v + \beta \sum_{w \neq v} (z_v - z_w)^T N_w^{t+1} (z_v - z_w) \tag{9}$$

Since Eq. (8) is the analytic solution of Eq. (9), we can derive that

$$\begin{aligned} &\|y - Y z_v^{t+1}\|^2 + \alpha (z_v^{t+1})^T M_v^{t+1} z_v^{t+1} + \beta \sum_{w \neq v} (z_v^{t+1} - z_w^t)^T N_w^{t+1} (z_v^{t+1} - z_w^t) \\ &\leq \|y - Y z_v^t\|^2 + \alpha (z_v^t)^T M_v^{t+1} z_v^t + \beta \sum_{w \neq v} (z_v^t - z_w^t)^T N_w^{t+1} (z_v^t - z_w^t) \end{aligned} \tag{10}$$

By substituting M_v^{t+1} and $N_w^{t+1} (w = 1, \ldots, l; w \neq v)$ into the above equation, we have

$$\begin{aligned} &L_{t+1} + \alpha \sum_i \frac{|z_v^{t+1}|_i |z_v^{t+1}|_i}{2|z_v^t|_i} + \beta \sum_{w \neq v} \sum_i \frac{|z_v^{t+1} - z_w^t|_i |z_v^{t+1} - z_w^t|_i}{2|z_v^t - z_w^t|_i} \\ &\leq L_t + \alpha \sum_i \frac{|z_v^t|_i |z_v^t|_i}{2|z_v^t|_i} + \beta \sum_{w \neq v} \sum_i \frac{|z_v^t - z_w^t|_i |z_v^t - z_w^t|_i}{2|z_v^t - z_w^t|_i} \end{aligned} \tag{11}$$

where $L_t = \|y - Y z_v^t\|^2$, and $|k|_i$ is the ith absolute value for vector k. Here we introduce a function $f(x) = x - x^2/2a$, which has the property: $\forall x \in R, a > 0, f(x) \leq f(a)$. Then we substitute a with $|z_v^t|_i$ and $|z_v^t - z_w^t|_i$, and let x be $|z_v^{t+1}|_i$ and $|z_v^{t+1} - z_w^t|_i$ respectively. Then the following equations hold

$$\sum_v |z_v^{t+1}|_i - \sum_i \frac{|z_v^{t+1}|_i |z_v^{t+1}|_i}{2|z_v^t|_i} \leq \sum_v |z_v^t|_i - \sum_i \frac{|z_v^t|_i |z_v^t|_i}{2|z_v^t|_i} \tag{12}$$

$$\begin{aligned} &\sum_{w \neq v} \sum_i |z_v^{t+1} - z_w^t|_i - \sum_{w \neq v} \sum_i \frac{|z_v^{t+1} - z_w^t|_i |z_v^{t+1} - z_w^t|_i}{2|z_v^t - z_w^t|_i} \\ &\leq \sum_{w \neq v} \sum_i |z_v^t - z_w^t|_i - \sum_{w \neq v} \sum_i \frac{|z_v^t - z_w^t|_i |z_v^t - z_w^t|_i}{2|z_v^t - z_w^t|_i} \end{aligned} \tag{13}$$

Adding both sides of Eqs. (12) and (13) into Eq. (11), we obtain

$$\begin{aligned} &L_{t+1} + \alpha \sum_i |z_v^{t+1}|_i + \beta \sum_{w \neq v} \sum_i |z_v^{t+1} - z_w^t|_i \\ &\leq L_t + \alpha \sum_i |z_v^t|_i + \beta \sum_{w \neq v} \sum_i |z_v^t - z_w^t|_i \end{aligned} \tag{14}$$

Namely, the following equation holds

$$\begin{aligned} &\|y - Y z_v^{t+1}\|^2 + \alpha \|z_v^{t+1}\|_1 + \beta \sum_{w \neq v} \|z_v^{t+1} - z_w^t\|_1 \\ &\leq \|y - Y z_v^t\|^2 + \alpha \|z_v^t\|_1 + \beta \sum_{w \neq v} \|z_v^t - z_w^t\|_1 \end{aligned} \tag{15}$$

Therefore, Algorithm 1 decreases Eq. (3) in each iteration.

Computational complexity analysis: We briefly discuss the computational complexity of the proposed multi-view sparse subspace clustering algorithm. In Algorithm 1, the main calculation lies in the inverse problem as in Eq. (8), which has a cubic complexity. However, by solving a system of linear equations instead, we can obtain a quadratic complexity of the number of points [3]. Moreover, since each column of $Z^v (v = 1, \ldots, l)$ is solved individually, a parallel computing strategy can be easily adopted for efficiency.

3.5. Discussion of different co-regularization forms

As explained in Section 3.1, different co-regularization constrains can be utilized to explore the correlation between different views. In this subsection, we discuss two extra co-regularization forms that can be used to make use of multiple sources of information, and their objectives are respectively formulated as

$$\min_{Z^v} \sum_v \|X^v - X^v Z^v\|_F^2 + \alpha \sum_v \|Z^v\|_1 + \beta \sum_{1 \leq v < w} \|Z^v - Z^w\|_F^2$$

[3] Or $(Z_i^w)^{t+1}$ that has been calculated already for some w. Since it does not influence the proof of the convergence analysis, we use $(Z_i^w)^t$ for all w for convenience.

Please cite this article as: Q. Yin, et al., Multi-view clustering via pairwise sparse subspace representation, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.01.017

$$\text{s.t. } \mathrm{diag}(Z^v)=0, \quad \forall v\in\{1,\ldots,l\} \tag{16}$$

$$\min_{Z^v}\sum_v\|X^v-X^vZ^v\|_F^2+\alpha\sum_v\|Z^v\|_1+\beta\sum_{i=1}^n\sum_{1\le v<w}\|[Z_i^v,Z_i^w]\|_{21}$$

$$\text{s.t. } \mathrm{diag}(Z^v)=0, \quad \forall v\in\{1,\ldots,l\} \tag{17}$$

where in Eq. (16), $\|Z^v-Z^w\|_F^2$ is an l_2-norm based pairwise co-regularization to force the view specific representations to be similar. In Eq. (17), $\|[Z_i^v,Z_i^w]\|_{21}$ is the l_{21}-norm on one point's two reconstruction coefficient vectors, which enforces the sparse representations of the two views have the same nonzero group structure.

Since the $\|Z^v-Z^w\|_F^2$ term is differentiable, the optimization and the convergence of Eq. (16) can be obtained by following Sections 3.2 and 3.4. As for Eq. (17), since the added term is non-smooth at the origin, an iteratively re-weighted method is applied like in Eq. (5). And the derivative of Eq. (17) with respect to Z_i^v is

$$(X^v_{-i})^TX^v_{-i}Z_i^v-(X^v_{-i})^TX_i^v+\alpha M_vZ_i^v+\beta\sum_{w\ne v}N_wZ_i^v=0 \tag{18}$$

where N_w is a diagonal matrix with its jth diagonal element calculated as $(N_w)_{jj}=\frac{1}{2}[(Z_i^v)_j,(Z_i^w)_j]_2$ and the other variables are the same as in Eq. (6). Then the optimization and convergence can be obtained like MultiSSC in Sections 3.2 and 3.4.

To distinguish the above two co-regularization based clustering methods from MultiSSC as in Eq. (3), we name them as MultiSSC_1 and MultiSSC_2 respectively.

4. Extensions of MultiSSC

In the previous section, we give our algorithm for multi-view clustering (MultiSSC). Now we give another two extensions of our proposed MultiSSC method, which consist of a weakly semi-supervised link constrained multi-view clustering (L-MultiSSC) and a special co-regularization based multi-view clustering method (S-MultiSSC).

4.1. Link constrained MultiSSC (L-MultiSSC)

In many applications, we can easily obtain some pairwise link information that tells us whether two points are in the same class or not. This prior information is also called must-link and cannot-link constraints [36]. A must-link constraint enforces that two points must be in the same category while a cannot-link constraint enforces that two points are placed in different classes.

By using a partially observed link matrix that has binary entries indicating which candidate data points are in the same category, we can guide the learning of multi-view sparse representation under properly designed constraints. Since the partially observed link matrix reflects the high-level semantics of data and provides groundtruth of the pairwise relationship, we enforce the corresponding entries of each view-specific sparse representation matrix to be consistent with the observed link entities as in [37], which uses a similar constraint in image classification application. The objective is then formulated as

$$\begin{aligned}\min_{Z^v}&\sum_v\|X^v-X^vZ^v\|_F^2+\alpha\sum_v\|Z^v\|_1\\&+\beta\sum_{1\le v<w}\|Z^v-Z^w\|_1+\gamma\sum_v\sum_{ij\in O}(Z_{ij}^v-L_{ij})^2\\\text{s.t. }&\mathrm{diag}(Z^v)=0, \quad \forall v\in\{1,\ldots,l\}\end{aligned} \tag{19}$$

where L is the partially observed link matrix with its observed (ij) th element being 1 or 0 depending on whether point i and point j are in the same category. O is the observed must-link and cannot-link set. The other variables are the same as in Eq, (3).

We define an indicator matrix R with its (ij) th value assigned as follows:

$$R_{ij}=\begin{cases}1 & \text{if } L_{ij} \text{ is observed}\\ 0 & \text{if } L_{ij} \text{ is unobserved}\end{cases} \tag{20}$$

Then the introduced link constrained term $\sum_v\sum_{ij\in O}(Z_{ij}^v-L_{ij})^2$ can be formulated as $\|Z\odot R-L\odot R\|_F^2$, where the operator $\odot$represents the element-wise product of two matrices. And Eq. (19) can be further rewritten as

$$\begin{aligned}\min_{Z^v}&\sum_v\|X^v-X^vZ^v\|_F^2+\alpha\sum_v\|Z^v\|_1\\&+\beta\sum_{1\le v<w}\|Z^v-Z^w\|_1+\gamma\|Z\odot R-L\odot R\|_F^2\\\text{s.t. }&\mathrm{diag}(Z^v)=0, \quad \forall v\in\{1,\ldots,l\}\end{aligned} \tag{21}$$

Since the added link constraint term is differentiable, the optimization and convergence analysis of Eq. (21) can be derived by following Sections 3.2 and 3.4.

4.2. Shared co-regularization for MultiSSC (S-MultiSSC)

Different from the l_1-norm based pairwise co-regularization that each view has its individual sparse representation, we assume that all the views share a unified representation Z^*. Namely, we consider the intrinsic spare representation can associate all the views. Even though this assumption is a little arbitrary, especially for the noise data, it can lead to an interesting observation. Firstly, the model can be formulated as

$$\min_{Z^*}\sum_v\|X^v-X^vZ^*\|_F^2+\alpha\|Z^*\|_1 \quad \text{s.t. } \mathrm{diag}(Z^*)=0 \tag{22}$$

Since all the views share the same sparse representation, there will be no co-regularization between different views.

We will derive a simple equivalent form of Eq. (22). Firstly, the sum of all views' reconstruction errors can be written as

$$\sum_v\|X^v-X^vZ^*\|_F^2=\mathrm{tr}\left[(I-Z^*)^T\left(\sum_v(X^v)^T(X^v)\right)(I-Z^*)\right]=\|X-XZ^*\|_F^2 \tag{23}$$

where $X=[X^1;X^2;\ldots;X^l]$, namely a concatenation of different views of the dataset. Thus this model can be finally formulated as

$$\min_{Z^*}\|X-XZ^*\|_F^2+\alpha\|Z^*\|_1 \quad \text{s.t. } \mathrm{diag}(Z^*)=0 \tag{24}$$

This multi-view clustering model equals to the feature concatenation of all views and then uses this new feature to obtain the sparse representation matrix. The above observation in our model gives a justification for using concatenation of the multi-view features. As for its optimization, we can resort to Section 3.2.

5. Experiments

In this section, extensive experiments are conducted to demonstrate the effectiveness of our proposed multi-view sparse subspace clustering model.

5.1. Databases

We report experimental results on five public datasets and the description of the databases are summarized in Table 1.

- *UCI handwritten digit dataset*:[4] It consists of features of handwritten numerals (0–9) extracted from a collection of Dutch utility maps. The dataset consists of 2000 samples with 200 in each category, and it is represented in terms of six features. Similar with [6], we select the 76 Fourier coefficients of the character

[4] http://archive.ics.uci.edu/ml/datasets/Multiple+Features

Please cite this article as: Q. Yin, et al., Multi-view clustering via pairwise sparse subspace representation, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.01.017

Table 1
Information of the multi-view datasets.

Info.	Digits	3-Source	Movies617	Animal	WebKB
# of Views	2	3	2	3	2
# of Clusters	10	6	17	10	5

shapes and the 216 profile correlations as two views of the original dataset.

- *3-Sources text dataset*:[5] It was collected from three well-known online news sources: BBC, Reuters and the Guardian. In total it consists of 416 distinct news manually categorized into six classes. Among them, 169 are reported in all three sources and are used in our experiments as in [7] with each source serving as one independent view of a story. And the feature describing stories for all the three views is the word frequency.
- *WebKB dataset*: It consists of webpages collected from four universities: Texas, Cornell, Washington and Wisconsin, and each webpage can be described by the content view and the link view. Since the four subsets of the dataset are similar in content and organization, we just test our method on the Texas dataset. Texas is divided into five categories (course, project, student, faculty and staff) and the data can be downloaded from the Internet.[6]
- *Movies617 dataset*:[6] It was extracted from IMDb (http://www.imdb.org), and consists of 617 movies over 17 labels. The two views are the 1878 keywords and the 1398 actors with a keyword used for at least 2 movies and an actor appeared in at least 3 movies.
- *Animal dataset*:[7] It consists of 30,475 images of 50 animals with six pre-extracted features for each image. In this dataset, we select three kinds of features, namely PyramidHOG (PHOG), color-SIFT and SURF as three views. Besides, we select the first ten categories with each class consisting of randomly sampled 50 points as a subset to evaluate the proposed method.

5.2. Experimental settings

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

S_Spectral: Using the spectral clustering method in [35] to cluster each view's data and selecting the best clustering result.

S_Subspace: Using sparse subspace representation of each view's data as shown in Eq. (1) to construct the affinity matrix, and then using the spectral clustering method in [35] to cluster the dataset. Similar with S_Spectral, the best clustering results are reported.

PairwiseSC, *CentroidSC*: Kumar [6] proposed two objective functions to co-regularize the eigenvectors of all views' Laplacian matrices.

Co_Training: Alternately modifying one view's graph structure using the other view's information [23].

Multi_NMF: Liu et al. [7] developed a multi-view non-negative matrix factorization method to group the multi-view data.

Multi_CF: Wang et al. [3] proposed a structure sparsity based multi-view clustering and feature learning framework, which is used for multi-view data clustering.

MultiSSC: Our proposed multi-view sparse subspace clustering method as in Eq. (3).

It should be noted that the essence of one of our extensions, i.e., S-MultiSSC, is the concatenation of different views, and then uses subspace representation to construct the affinity matrix. This method also serves as one important baseline, and its results are displayed in Section 5.6.

In this work, we are not concentrating on how to choose the clustering numbers, which can be solved through existing algorithms and we just set it to the true number of categories. For those methods using the Gaussian kernel to construct the affinity matrix, we use the mean value of the Euclidean distance between all data points as the standard deviation [6]. For PairwiseSC, CentroidSC, Co_Training, Multi_NMF methods, we use the codes released by their authors and follow the suggestions the authors have given to achieve their best clustering results. For Multi_CF, we implement the code and follow the clustering rules the authors suggested to achieve the final clustering results. For our method, $\alpha$and $\beta$are empirically selected to reach the best clustering performance. Besides, we choose the view, which achieves the best clustering results in S_Spectral as the one to construct the affinity matrix when all view's subspace representations are obtained. As K-means is used in all the experiments, it is run 20 times with random initialization. Two measures, the accuracy (ACC) and the normalized mutual information (NMI) are used to measure the clustering results. Readers can refer to [38] for more details about their definitions.

5.3. Experimental results

Tables 2 and 3 respectively show the clustering accuracy and normalized mutual information of different algorithms on the five public datasets. Overall, it can be seen that our method almost outperforms all the compared algorithms. Compared with the single view spectral clustering method (S_Spectral), sparse subspace based single view clustering algorithm (S_Subspace) gains better results. This demonstrates that the sparse representation can better capture the latent group structure of the dataset, namely the (dis)similarity between data points, than the Gaussian kernel based Euclidean distance.

Compared with the S_Subspace method, our proposed MultiSSC gains promising results, and has at least 3 percent improvement in terms of both accuracy and normalized mutual information in four databases. Furthermore, in Digits and 3-Source databases, the proposed method achieves more than 9 percent improvement in terms of both evaluation measures. This validates the effectiveness of the proposed l_1-norm based pairwise co-regularization, which can take advantage of the multiple sources of information.

CentroidSC and PairwiseSC algorithms aim to find low dimensional embedding of the Laplacian matrices from multi-view datasets. However, these methods use the Gaussian kernel to construct the affinity matrix, which may not truly reflect the (dis)similarity between data points. This is because even the Euclidean distance of two points is large, the points can still be sampled from the same class because of the large within class variance. And this may be one reason that our method performs better than these two algorithms.

Multi_NMF is based on the non-negative matrix factorization and it aims to find an intrinsic low dimensional representation of the dataset. From the experimental results, Multi_NMF has poor performance compared with our method. Besides, Multi NMF is limited for it needs the feature matrix to be non-negative, which fails to deal with the feature matrix with negative values.

As for the Multi_CF algorithm, it is a regression-like clustering method, which can directly obtain the final cluster indicator matrix. By properly designing regularizers, this method can make use of the type-wise relevance and feature-wise information among different views. Compared with our method, it explicitly makes use of the relationship of the multi-view features and can

[5] http://mlg.ucd.ie/datasets/3sources.html
[6] http://membres-liglab.imag.fr/grimal/data.html
[7] http://attributes.kyb.tuebingen.mpg.de/

Please cite this article as: Q. Yin, et al., Multi-view clustering via pairwise sparse subspace representation, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.01.017

obtain the final group information without relying on extra clustering methods. However, the regression-based method, whose clustering results are not as good as ours, cannot explicitly discover the latent group structure.

Table 2
Clustering results in terms of accuracy on the Digits, 3-Sources, WebKB, Movies617 and Animal databases. Both mean value and standard deviation are reported.

ACC(%)	Digits	3-Source	Movies617	Animal	WebKB
S_Spectral	66.37(4.44)	52.93(3.59)	25.70(1.13)	27.21(1.50)	56.79(1.25)
S_Subspace	76.60(4.78)	64.85(0.73)	28.56(1.15)	28.11(1.36)	64.79(6.61)
PairwiseSC	80.82(6.30)	58.37(3.28)	27.89(1.64)	31.65(1.59)	51.63(1.76)
CentroidSC	82.77(7.14)	58.93(3.07)	28.57(1.17)	31.06(2.02)	53.90(2.31)
Co_Training	80.22(6.84)	58.37(3.47)	30.74(1.28)	30.35(1.48)	54.25(2.70)
Multi_NMF	69.24(6.28)	68.40(0.06)	26.99(1.19)	30.56(1.02)	–
Multi_CF	72.45(4.10)	69.23(3.54)	29.60(1.10)	32.11(1.86)	59.79(0.33)
MultiSSC	86.16(5.31)	73.99(1.76)	33.13(1.44)	32.51(1.36)	65.53(2.34)

Table 3
Clustering results in terms of normalized mutual information on the Digits, 3-Sources, WebKB, Movies617 and Animal databases.

NMI(%)	Digits	3-Source	Movies617	Animal	WebKB
S_Spectral	62.30(1.85)	53.38(2.12)	25.47(0.85)	15.70(0.65)	40.53(1.91)
S_Subspace	71.69(1.96)	54.37(0.00)	29.37(1.01)	17.36(1.06)	42.58(3.45)
PairwiseSC	75.84(2.37)	62.25(2.76)	28.04(0.73)	19.90(1.51)	38.54(0.69)
CentroidSC	76.76(2.58)	62.65(2.51)	28.02(0.72)	18.50(1.48)	38.17(1.05)
Co_Training	75.90(2.27)	63.15(1.79)	30.74(1.28)	18.98(0.73)	36.24(1.74)
Multi_NMF	65.05(2.30)	60.20(0.06)	27.45(0.55)	18.77(0.71)	–
Multi_CF	74.55(2.49)	67.91(4.41)	30.09(1.32)	21.25(1.76)	32.94(3.39)
MultiSSC	83.30(3.75)	67.42(2.23)	33.37(0.98)	21.28(0.57)	45.01(3.51)

5.4. Parameter selection

For the proposed multi-view subspace clustering model, there are two parameters balancing the reconstruction error term, sparse constraint term and the l_1-norm based co-regularization term. α is a parameter that controls the regularization on the sparse representation, and the parameter β controls the pairwise co-regularization between view-specific sparse representations. In this section, we investigate how the performance varies with the change of these two parameters. Here we just report the accuracy and NMI on the 3-Source, WebKB and Movies617 databases due to space limitation, and the Digits and Animal datasets show similar results.

From Figs. 1 and 2, we have the following conclusions. When α is too small, the sparse reconstruction will lose its effect. In the case when α is too big, the sparse characteristic will lead to reconstructing the target data with only a few points, and will harm the subspace clustering effect. As for the parameter β, when β is small, the pairwise co-regularization effect is weak, so the view-specific sparse representations cannot promote each other very well. In the case when β is too big, the punishment of the function will be mainly imposed on the co-regularization item and the regularizer $\|z_i^v\|_1$ will lose its effect. So α and β should be carefully selected. In all our experiments, when α and β are chosen from the interval [0.005,0.1], acceptable results can be obtained.

5.5. Convergence

As discussed in Section 3.4, the updating rule decreases the objective function in each iteration. In Fig. 3, we give the convergence curves together with the final clustering accuracy on the 3-Source, WebKB and Movies617 databases (In order to

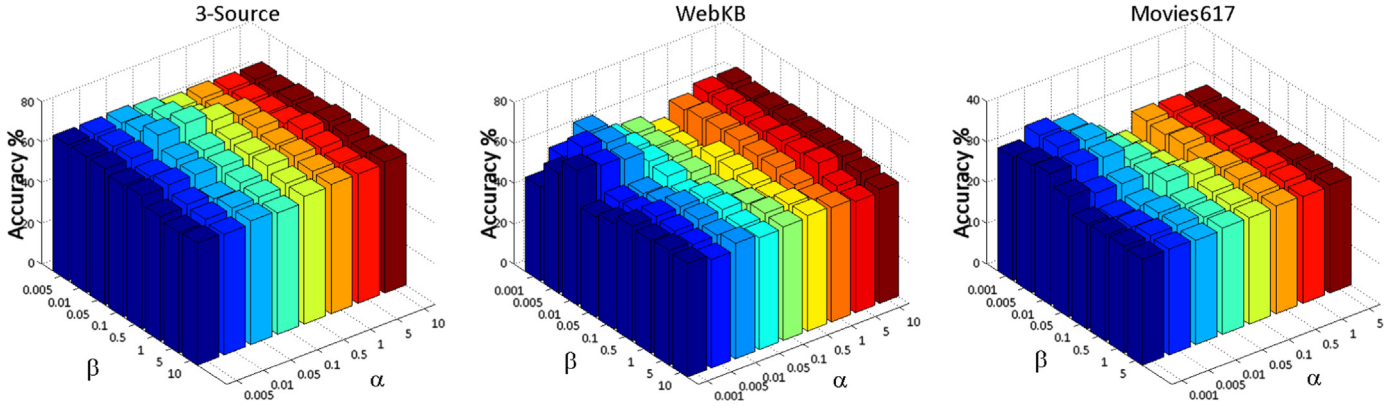


Fig. 1. Accuracy of the proposed MultiSSC vs. parameters α and β.

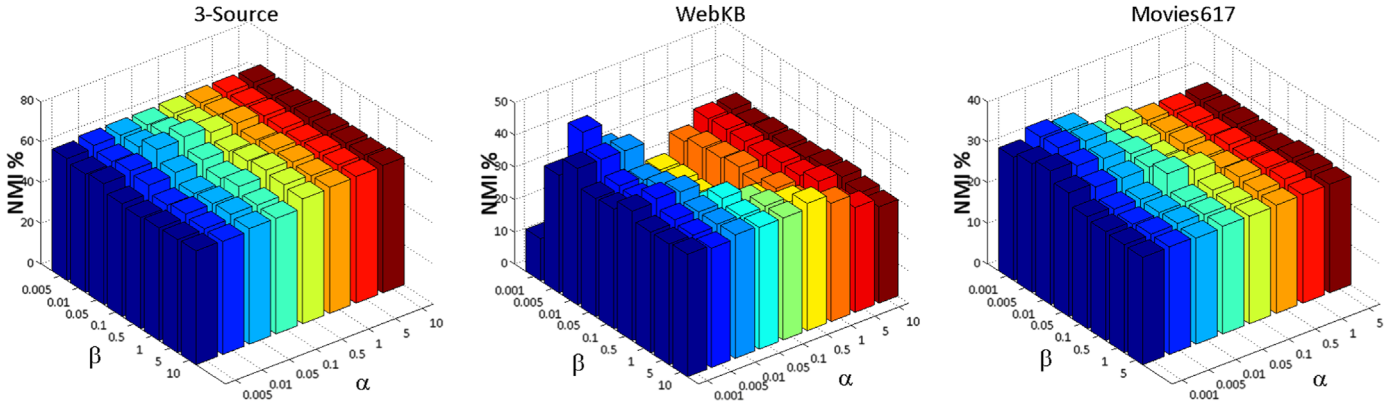


Fig. 2. NMI of the proposed MultiSSC vs. parameters α and β.

Please cite this article as: Q. Yin, et al., Multi-view clustering via pairwise sparse subspace representation, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.01.017

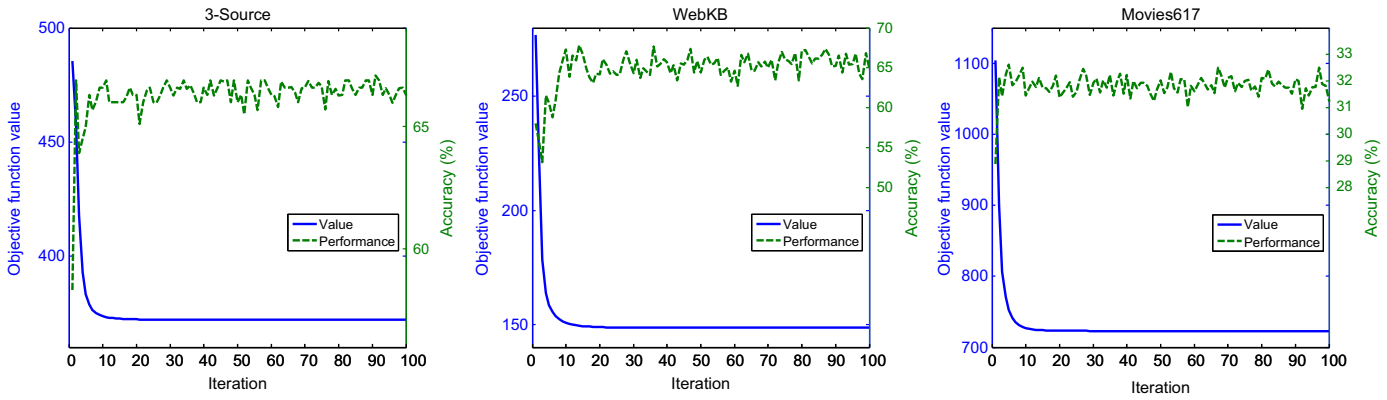


Fig. 3. Convergence and the corresponding accuracy curves of MultiSSC.

Table 4
Clustering results of S-MultiSSC and L-MultiSSC methods.

Methods	3Source		WebKB		Movies617	
	ACC (%)	NMI (%)	ACC (%)	NMI (%)	ACC (%)	NMI (%)
MultiSSC	73.99(1.76)	67.42(2.23)	65.53(2.34)	45.01(3.51)	33.13(1.44)	33.37(0.98)
S-MultiSSC	72.37(3.52)	65.67(2.47)	60.72(1.82)	30.32(1.52)	27.35(1.06)	27.79(0.58)
L-MultiSSC	75.61(3.57)	68.58(0.71)	67.00(0.74)	44.63(5.61)	34.41(1.28)	35.65(1.43)

Table 5
Clustering results in terms of accuracy for MultiSSC, MultiSSC_1 and MultiSSC_1.

ACC(%)	Digits	3-Source	Movies617	Animal	WebKB
MultiSSC	86.16(5.31)	73.99(1.76)	33.13(1.44)	32.51(1.36)	65.53(2.34)
MultiSSC_1	80.12(4.61)	74.79(1.70)	33.72(1.15)	32.10(1.11)	66.36(1.45)
MultiSSC_2	83.95(6.78)	71.72(4.00)	33.19(0.94)	30.41(0.71)	65.27(2.24)

Table 6
Clustering results in terms of NMI for MultiSSC, MultiSSC_1 and MultiSSC_1.

NMI(%)	Digits	3-Source	Movies617	Animal	WebKB
MultiSSC	83.30(3.75)	67.42(2.23)	33.37(0.98)	21.28(0.57)	45.01(3.51)
MultiSSC_1	73.61(0.70)	69.38(2.02)	33.22(0.76)	20.83(0.63)	43.27(0.66)
MultiSSC_2	80.11(2.60)	64.42(3.93)	32.86(0.95)	19.41(1.21)	42.51(2.66)

save space, only these three databases with their corresponding curves are displayed and the other two datasets show similar results).

As can be seen from Fig. 3, the algorithm converges after about 15 iterations. As for the clustering performance, the accuracy stops changing a lot after the algorithm converges. The slightly small changes of the performance may be because of the use of the K-means algorithm to obtain the accuracy.

5.6. Results of S-MultiSSC and L-MultiSSC

We test the performance of S-MultiSSC and L-MultiSSC methods on the aforementioned datasets. To obtain the partially observed link information, we randomly label 10% of the dataset and construct the pairwise must-link and cannot-link prior based on the label information. And this setting is repeated 5 times for fair comparison. It should be noted that we are concentrating on testing the effect of this link prior information, so we will not compare our algorithm with semi-supervised clustering methods. The clustering results are shown in Table 4.

S-MultiSSC equals to concatenating features of all the views and then using the new feature to obtain the sparse representation matrix. Since this strategy may bring in information redundancy and has a high probability of adding more noise to the data representation, it cannot obtain the clustering performance as good as the MultiSSC.

Compared with the MultiSSC method, L-MultiSSC constrains the entities of all views' sparse representation matrices to be consistent with the observed link matrix, thus guiding the learning of the representation matrices. Since the link information reflects the high-level semantics of the data, it can promote the learning process. Thus, L-MultiSSC gains better clustering results compared with MultiSSC.

5.7. Results of MultiSSC_1 and MultiSSC_2

Further, we compare the performance of MultiSSC with the other two co-regularization based multi-view clustering methods as introduced in Section 3.5. The clustering results on the five datasets are shown in Tables 5 and 6. Overall, MultiSSC obtains slightly better clustering results in terms of both accuracy and NMI, and MultiSSC_1 and MultiSSC_2 also achieve promising results compared with previous multi-view clustering methods as introduced in Section 5.2. This further validates our proposed multi-view clustering framework.

Compared with MultiSSC_1, MultiSSC uses an l_1-norm based co-regularization instead of the l_2-norm co-regularization. When a point in one view is polluted, its sparse representation will be affected, and this will then affect its correlation effect with other

Please cite this article as: Q. Yin, et al., Multi-view clustering via pairwise sparse subspace representation, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.01.017

Table 7
Clustering results vs. the number of views on animal database.

# of views	1	2	3	4	5	6
ACC (%)	28.11(1.36)	30.71(2.17)	33.34(1.58)	33.17(1.72)	32.86(1.84)	32.93(1.57)
NMI (%)	17.36(1.06)	19.79(0.76)	20.44(0.71)	21.51(0.76)	21.81(1.33)	20.12(0.90)

view's representations. The l_1-norm based pairwise constraint will relieve this situation by putting relatively small punishment on the co-regularization than the l_2-norm based co-regularization. Thus, MultiSSC may be more robust to this situation.

Compared with the MultiSSC, MultiSSC_2 only requires the nonzero positions of the view-specific representation to be the same. Maybe it is a bit weak to maximize the correlation between views, since it does not constrain the values of the sparse representations. Maybe it is one reason that the performance of MultiSSC is better than MultiSSC_2.

5.8. Performance vs. the number of views

We give the results of how the clustering performance varies as the number of views increases. As for the order of the added views, we follow the idea of boosting for feature selection [39], namely the added view can achieve the best performance together with the previous views. The steps are detailed as follows:

1. Choose one view which can achieve the best clustering results.
2. Choose one of the rest views which together with the former views can achieve the best performance.
3. Repeat step 2 until all views are selected.

In the experiments, our model is tested on the six-view Animal dataset, which is introduced in Section 5.1. And the other experimental settings are the same as in Section 5.2.

From Table 7, we can see that the clustering results stop improving when the total number of views is four. It is because in the real world, as the number of views increases, the complementary information will be saturated, thus making the clustering performance cease to increase. Besides, increasing the number of views may lead to the risk of bringing in outliers and noise, which may harm the final clustering performance.

6. Conclusion

In this paper, we have proposed a novel multi-view sparse subspace clustering framework to cluster the multi-view data. Being formulated as a joint subspace segmentation problem with a pairwise co-regularization constraint, the proposed MultiSSC model can take advantage of the prior sparse information and the complementary information embedded in the multi-view data. We have also developed an iterative optimization algorithm with rigorous theoretical proof on its convergence to solve the regularization problem. Besides, different pairwise co-regularization forms are also discussed to maximize the correlation between views. Moreover, a must-link and cannot-link constrained multi-view clustering algorithm is devised to integrate the partially available supervisory information. Extensive experiments have demonstrated the effectiveness of our method compared with several state-of-the-art multi-view clustering algorithms.

Acknowledgments

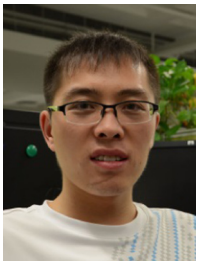
This work is jointly supported by National Basic Research Program of China (No. 2012CB316300) and National Science Foundation of China (No. 61175003).

References

[1] S. Sun, A survey of multi-view machine learning, Neural Comput. Appl. 23 (2013) 2031–2038.
[2] G. Tzortzis, A. Likas, Kernel-based weighted multi-view clustering, in: International Conference on Data Mining, 2012, pp. 675–684.
[3] H. Wang, F. Nie, H. Huang, Multi-view clustering and feature learning via structured sparsity, in: International Conference on Machine Learning, 2013, pp. 352–360.
[4] E. Muller, S. Gunnemann, I. Farber, T. Seidl, Discovering multiple clustering solutions: grouping objects in different views of the data, in: International Conference on Data Mining, 2010, pp. 1207–1210.
[5] M.R. Amini, N. Usunier, C. Goutte, Learning from multiple partially observed views—an application to multilingual text categorization, Neural Inf. Process. Syst. 1 (2010) 28–36.
[6] A. Kumar, P. Rai, H. Daume III, Co-regularized multiview spectral clustering, Neural Inf. Process. Syst. (2011) 1413–1421.
[7] J. Liu, C. Wang, J. Gao, J. Han, Multi-view clustering via joint nonnegative matrix factorization, in: SIAM International Conference Data Mining, 2013, pp. 252–260.
[8] B. Long, S. Y. Philip, Z. M. Zhang, A general model for multiple view unsupervised learning, in: SIAM International Conference on Data Mining, 2008, pp. 822–833.
[9] D. Greene, P. Cunningham, A matrix factorization approach for integrating multiple data views, in: European Conference on Machine Learning and Knowledge Discovery in Databases, 2009, pp. 423–438.
[10] C.-Y. Lu, H. Min, Z.-Q. Zhao, L. Zhu, D.-S. Huang, S. Yan, Robust and efficient subspace segmentation via least squares regression, in: European Conference on Computer Vision, 2012, pp. 347–360.
[11] E. Elhamifar, R. Vidal, Sparse subspace clustering, Comput. Vis. Pattern Recognit. (2009) 2790–2797.
[12] G. Liu, Z. Lin, Y. Yu, Robust subspace segmentation by low-rank representation, in: International Conference on Machine Learning, 2010, pp. 663–670.
[13] R. Vidal, Subspace clustering, IEEE Signal Process. Mag. 28 (2011) 52–68.
[14] X. Dong, P. Frossard, P. Vandergheynst, N. Nefedov, Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds, IEEE Trans. Image Process. 62 (2014) 905–918.
[15] X. Cai, F. Nei, H. Huang, F. Kamangar, Heterogeneous image feature integration via multi-modal spectral clustering, Comput. Vis. Pattern Recognit. (2011) 1977–1984.
[16] Y. Guo, Convex subspace representation learning from multi-view data, in: National Conference on Artificial Intelligence, 2013.
[17] T. Xia, D. Tao, T. Mei, Y. Zhang, Multiview spectral embedding, IEEE Trans. Syst. Man Cybern. Part B: Cybern. 40 (2010) 1438–1446.
[18] W. Tang, Z. Lu, I.S. Dhillon, Clustering with multiple graphs, in: International Conference on Data Mining, 2009, pp. 1016–1021.
[19] K. Chaudhuri, S.M. Kakade, K. Livescu, K. Sridharan, Multi-view clustering via canonical correlation analysis, in: International Conference on Machine Learning, 2009, pp. 129–136.
[20] X. He, M.-Y. Kan, P. Xie, X. Chen, Comment-based multi-view clustering of web 2.0 items, in: International Conference on World Wide Web, 2014, pp. 771–782.
[21] S. Bickel, T. Scheffer, Multi-view clustering, in: International Conference on Data Mining, vol. 4, 2004, pp. 19–26.
[22] X. Zhao, N. Evans, J.-L. Dugelay, A subspace co-training framework for multi-view clustering, Pattern Recognit. Lett. 41 (2014) 73–82.
[23] A. Kumar, H. Daume III, A co-training approach for multi-view spectral clustering, in: International Conference on Machine Learning, 2011, pp. 393–400.
[24] B. Gilles, C. Grimal, Co-clustering of multi-view datasets: a parallelizable approach, in: International Conference on Data Mining, 2012, pp. 828–833.
[25] E. Bruno, S. Marchand-Maillet, Multiview clustering: a late fusion approach using latent models, in: ACM Conference on Research and Development in Information Retrieval, 2009, pp. 736–737.
[26] Hussain S, Mushtaq M, Halim Z, Multi-view document clustering via ensemble method, J. Intell. Inf. Syst. 43 (2014) 81–99.

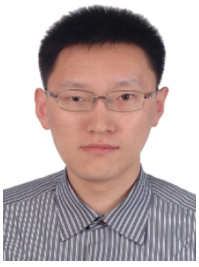
Please cite this article as: Q. Yin, et al., Multi-view clustering via pairwise sparse subspace representation, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.01.017

[27] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 171–184.
[28] E. Elhamifar, R. Vidal, Sparse subspace clustering: algorithm, theory, and applications, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 2765–2781.
[29] Y. Zhang, Z. Sun, R. He, T. Tan, Robust subspace clustering via half-quadratic minimization, in: International Conference on Computer Vision, 2013, pp. 3096–3103.
[30] D. Luo, F. Nie, C. Ding, H. Huang, Multi-subspace representation and discovery, in: European Conference on Machine Learning and Knowledge Discovery in Databases, 2011, pp. 405–420.
[31] W. Tang, R. Liu, Z. Su, J. Zhang, Structure-constrained low-rank representation, IEEE Trans. Neural Netw. Learn. Syst. 25 (2014) 2167–2179.
[32] C. Lu, H. Min, Z. Zhao, L. Zhu, D. Huang, S. Yan, Robust and efficient subspace segmentation via least squares regression, 2014, arXiv:1404.6736.
[33] F. Nie, H. Huang, X. Cai, C. Ding, Effcient and robust feature selection via joint l2,1-norms minimization, In Neural Information Processing Systems (2010) 1813-1821.
[34] I.F. Gorodnitsky, B.D. Rao, A re-weighted minimum norm algorithm, IEEE Trans. Image Process. 45 (1997) 600–616.
[35] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 888–905.
[36] K. Wagstaff, C. Cardie, S. Rogers, S. Schrödl, Community detection via heterogeneous interaction analysis, in International Conference on Machine Learning, vol. 1, 2001, pp. 577–584.
[37] L.S.L. Rank, Representations for image classification. Ensemble clustering with voting active clusters, Comput. Vis. Pattern Recognit. (2013) 676–683.
[38] W.-Y. Chen, Y. Song, H. Bai, C.-J. Lin, E.Y. Chang, Parallel spectral clustering in distributed systems, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 568–586.
[39] P. Viola, M.J. Jones, Robust real-time face detection, Int. J. Comput. Vis. 57 (2004) 137–154.



Qiyue Yin received the B.S. degree in automation control from Harbin Engineering University, Harbin, China in 2012. He is currently pursuing the Ph.D. degree at the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing, China. His research interests include clustering, recommender system, and computer vision.

Shu Wu received the B.S. degree from Hunan University, China, in 2004, the M.S. degree from Xiamen University, China, in 2007, and the Ph.D. degree from the University of Sherbrooke, Canada, in 2012, all in computer science. He is an assistant professor in the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences. His research interests include data mining, recommendation systems, and pervasive computing.



Ran He received the B.S. degree in Computer Science from Dalian University of Technology, the M.S. degree in Computer Science from Dalian University of Technology, and Ph.D. degree in Pattern Recognition and Intelligent Systems from Institute of Automation, Chinese Academy of Sciences in 2001, 2004 and 2009, respectively. Since September 2010, Dr. He has joined NLPR where he is currently a full associate Professor. He currently serves as an associate editor of Neurocomputing (Elsevier) and serves on the program committee of several conferences. His research interests focus on information theoretic learning, pattern recognition, and computer vision.

Liang Wang received both the B.S. and M.S. degrees from Anhui University in 1997 and 2000 respectively, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences (CASIA), in 2004. From 2004 to 2010, he has been working as a Research Assistant at Imperial College London, United Kingdom and Monash University, Australia, a Research Fellow at the University of Melbourne, Australia, and a lecturer at the University of Bath, United Kingdom, respectively. Currently, he is a full-time Professor of Hundred Talents Program at the National Lab of Pattern Recognition, CASIA. His major research interests include machine learning, pattern recognition and computer vision. He has widely published in highly-ranked international journals such as IEEE TPAMI and IEEE TIP, and leading international conferences such as CVPR, ICCV and ICDM. He is an associate editor of IEEE Transactions on SMC-B, International Journal of Image and Graphics, Signal Processing, Neurocomputing and International Journal of Cognitive Biometrics. He is currently an IAPR Fellow and Senior Member of IEEE.

Please cite this article as: Q. Yin, et al., Multi-view clustering via pairwise sparse subspace representation, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.01.017