SUBSPACE CLUSTERING WITH A LEARNED DIMENSIONALITY REDUCTION PROJECTION

Qiang Zhang, Zhenjiang Miao

Beijing Jiaotong University

Abstract

Subspace clustering aims to separate data from a union of low dimensional linear subspaces. Many recent subspace clustering methods based on self-representation are popular and achieve state-of-art performance. Dimensionality reduction is a common preprocessing procedure before applying these clustering methods. In this paper, we present an algorithm to segment subspaces with a learned dimensionality reduction projection instead of simply using PCA (Principal Component Analysis). We propose an objective function which simultaneously learns the dimensionality reduction projection and self-representation coefficients. We integrate SMR (Smooth Representation subspace clustering) into this framework and propose SMR_LP (Smooth Representation clustering with Learned Projection). We also propose an efficient method to optimize the cost function. A well learned projection helps preserving the data structure and improves the clustering performance. Experimental results demonstrate the effectiveness of our proposed method.

Index Terms—Motion segmentation, Face clustering, Dimensionality reduction, Subspace clustering, Spectral clustering

1. INTRODUCTION

High-dimensional data are ubiquitous in many practical computer vision and image processing applications, e.g. motion segmentation, face clustering. Often these high-dimensional vison data lie in or near to a union of low-dimensional subspaces. Subspace clustering is a technology to find this multiple low-dimensional structure and partition data into their underlying subspaces. Numerous subspace clustering algorithms have been suggested in the literature. They can be divided into four categories: iterative methods (K-subspaces [4], K-flats [5]), algebraic methods [8-11], statistical methods (MPPCA [6], MSL [7]), and spectral clustering based methods [3, 12-18], as summarized in [13] [17].

Recent works based on spectral clustering have attracted more eyes and achieve excellent performance. They construct the affinity matrix based on local or global information around each data and spectral clustering is applied on the affinity matrix to segment the data into appropriate number of groups. In particular, self-representation based methods (SSC [12, 13], LRR [3, 14, 15, 16], LSR [17], SMR [18]) try to write each data point as a linear of affine combination of other data points (In this paper, we refer to it as self-representation coding) and construct the affinity matrix based on the self-representation matrix. A self-representation matrix C is obtained by solving

$$\min_{C} \lambda \|X\text{-}XC\|_{\ell} + f(C) \quad (1),$$

where X is the collection of data with each column being a signal, $\|\bullet\|_{\ell}$ is a proper norm measuring the self-reconstruction error. $f(C)$ is the regularization term which is the essential difference between these self-representation methods. λ is a positive constant that control the tradeoff between reconstruction error and regularization. They utilize C to construct an affinity matrix $(|C|+|C|)/2$. The final segmenting result is produced by Spectral Clustering. Ideally, the affinity matrix should be block diagonal. To enforce that the problem receives a block diagonal solution, they adopt different regularization terms $f(C)$. SSC(Sparse Subspace Clustering)[12,13] use $\|C\|_1$, l-1 norm of C, (a surrogate of l-0 norm $\|C\|_0$) to pursue a sparse self-representation. [3,14,15,16] use $\|C\|_*$, nuclear norm of C (a surrogate of $Rank(C)$) to pursue a low rank representation. LSR(Least Square Regression) [17] theoretically shows when Enforced Block Diagonal Conditions are satisfied, we are able to get a block diagonal solution. They propose LSR using $\|C\|_F^2$, Frobenius norm of C. Further [18] indicates that grouping effect leads to a well affinity graph which is helpful for spectral clustering. Guided by Enforce Grouping Effect Conditions, they propose smooth representation subspace clustering[18] (SMR) method which results in competitive state-of-art performance in many subspace clustering tasks.

As the dimension of vision data is often very

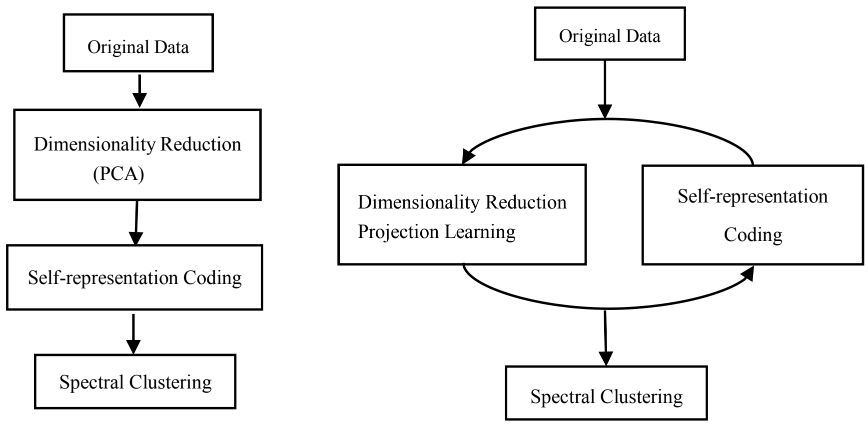

Figure 1. Flowchart of traditional subspace clustering method (Left) and our proposed method (Right). The main difference is that in our proposed method the dimensionality reduction projection is learned along with self-representation coding instead of using a fixed projection from PCA.

high, dimensionality reduction method is generally applied on the data for reducing data storage costs and computation complexity. Most of earlier works focus on the self-representation coding and neglect the importance of a proper projection for dimensionality reduction. They just simply use PCA which is not designed for subspace clustering. Few works study dimensionality reduction for subspace clustering: [19] present Latent Space Sparse Subspace Clustering (LS3C), which learns a projection for SSC and improves the performance of SSC. Inspired by their idea, we study the dimensionality reduction for general self-representation based methods.

Considering the task of segmenting subspaces, the multiple subspaces structure should be best preserved while projecting origin data into low-dimensional subspace. In this paper, we present a unified framework for simultaneous dimensionality reduction and self-representation coding that the projection matrix and the self-representation matrix are iteratively learned. Further we integrate SMR[18] into this framework and propose a new subspace clustering algorithm called SMR_LP (short for Smooth Representation clustering with Learned Projection), which simultaneously learns the projection and finds smooth self-representation coefficients in the projected low-dimensional space.

Paper contributions are as follows:

A unified framework for simultaneous dimensionality reduction and self-representation coding is proposed. Jointly learning the projection and self-representation matrix ensure that the projection can preserve the multiple subspaces structure, meantime a well learned projection helps finding a more accurate self-representation matrix for subspace clustering ;

We introduce an iterative procedure for optimizing the proposed objective function. It should be noted here, although we have similar idea with LS3C [19], the proposed method is more simple and efficient than LS3C in the step of optimization of transformation matrix. Moreover, as the sparse coding step of LS3C is much computational demanding than smooth representation coding, our proposed method is much faster than LS3C [19].

2. BACKGROUND

In this section, we first overview some basic definition of self-representation based subspace clustering.

Le $X=\left[X_1,X_2,\cdots,X_k\right]=\left[x_1,x_2,\cdots,x_n\right]\in\mathbb{R}^{D\times n}$ be a set of data vectors from a union of k subspaces $\left\{S_i\right\}_{i=1}^{k}$, Let X_i be a collection of data lying in subspace S_i, and $n=\sum_{i=1}^{k}n_i$. The task of subspace clustering is to segment the data into appropriate number of groups according to their underlying subspaces [13] [17].

For spectral clustering based methods, the main challenge is find a good affinity matrix C , with each entry C_{ij} measuring the affinity between data x_i and x_j . An idle affinity matrix should be block-diagonal, with affinities of data points from different subspaces being zeros. Self-representation based methods solve problem (1) to pursue the idle affinity matrix.

SMR (Smooth Representation subspace clustering method) is proposed in [18], which meets Enforce Grouping Effect Conditions and results in competitive state-of-art performance. SMR solves the minimum problem blow to find a self-representation matrix C .

$$\min_{C}\lambda\left\|X-XC\right\|_F^2+tr\left(CLC^T\right) \tag{2}$$

, where L is the Laplacian matrix [18].

Next, we study dimensionality reduction, the common preprocessing procedure on data points before applying self-representation coding. Let $P \in \mathbb{R}^{D\times d}$ be a linear projection matrix that maps the data from $\mathbb{R}^D$ to $\mathbb{R}^d$, with $d < D$. Let $Y = P^T X,\ Y \in \mathbb{R}^{d\times n}$ be the projected data. Then the self-representation matrix is found by solving

$$\min_C \ \lambda \|Y\text{-}YC\|_\ell + f(C) \qquad (3).$$

As the self-representation coding is done on the mapped low dimensional data Y, the computation is more efficient than on the original high dimensional data X. Previous works simply use PCA to build P which is not designed for subspace clustering. Next we will introduce our method which clusters subspaces with a learned dimensionality reduction projection.

3. SELF-REPRESENTATION BASED SUBSPACE CLUSTERING WITH A LEARNED PROJECTION

Here we propose an algorithm that the dimensionality reduction projection P is learned along with self-representation coding rather than using a fixed one as in traditional works. The self-representation matrix C and the projection are simultaneously learned by solving minimization problem as follows:

$$[P^*,\ C^*] = \arg\min_{P,C} J(P,C) \qquad (4)$$

$$J(P,C) = \lambda_1 \|P^T X \text{-} P^T XC\|_F^2 + f(C) + \lambda_2 \|X \text{-} PP^T X\|_F^2$$
$$s.t.\ P^T P = I$$

, where the first two terms promote a block-diagonal self-representation. The third term is a PCA-like regularization, which ensures that the projection retains original information. λ_1 and λ_2 are positive parameters controlling smooth representation and regularization. To avoid degenerate solution, we impose a constraint: $P^T P = I$, where I is a $d \times d$ identity matrix.

3.1. Optimization

Problem (4) involves two variables P and C, alternating optimization techniques [20] are adopted to solve this problem by iterative optimization of P and C. We refer the two steps as Projection learning and Self-representation coding respectively.

3.2. Projection learning

In this step, we optimize over P with a fixed C. The second term can be removed and we can obtain P by solving the following problem:

$$\min_P \ \lambda_1 \|P^T X \text{-} P^T XC\|_F^2 + \lambda_2 \|X \text{-} PP^T X\|_F^2 \qquad (5)$$
$$s.t. \quad P^T P = I$$

Then we expand the cost function to

$$tr(\lambda_1 (X - XC)(X - XC)^T PP^T)$$
$$+\lambda_2 tr(X^T PP^T PP^T X - 2X^T PP^T X + X^T X)$$

Using the equality constraint $P^T P = I$ and the fact that $tr(X^T X)$ is constant, we achieve

$$tr(\lambda_1 (X - XC)(X - XC)^T PP^T)$$
$$-\lambda_2 tr(X^T PP^T X)$$

Using the property of matrix trace, the problem further becomes

$$tr(P^T (\lambda_1 (X - XC)(X - XC)^T - \lambda_2 XX^T)P) \quad .$$

Then Problem (4) is equivalent to the optimization problem below,

$$\min \ tr(P^T MP)$$
$$s.t.\ P^T P = I \qquad (6)$$

, where $M = \lambda_1 (X \text{-} XC)(X \text{-} XC)^T \text{-} \lambda_2 XX^T$, Problem (6) is a minimum eigenvalue problem, the optimal solution is the d eigenvectors corresponding to the smallest d eigenvalues of M.

3.3. Self-representation coding

Based on existing P, we can obtain C by solving the following problem:

$$\min_C \ \lambda_1 \|Y \text{-} YC\|_F^2 + f(C) \qquad (7)$$

Where $Y = P^T X$. It is the same as problem (1), expect that the self-representation coding is done on matrix Y rather than on the data matrix X. It is solved by the same way as traditional self-representation method. Algorithm 1 summarized the procedure of self-representation coding along with the projection matrix learning.

Here we integrate SMR [18] method into this framework to jointly learn the smooth representation matrix and the projection. We call this method SMR_LP, in which we substitute (8) for (7).

$$\min_C \lambda_1 \|Y \text{-} YC\|_F^2 + tr\left(CLC^T\right) \qquad (8)$$

Same to most self-representation Clustering methods, once the smooth representation coefficient matrix C is found, the pairwise affinity matrix is constructed as $W = (|C| + |C^T|)/2$, then the segmentation of the projected data is obtained by applying spectral clustering.

Algorithm 1: Self-representation coding with a learned dimensionality reduction projection
Input: Data $X \in R^{D\times n}$, λ_1, λ_2 **Initialization:** Set P to be the d eigenvectors corresponding to the top d eigenvalues of XX^T . **Repeat** **Step 1: Self-representation coding** -Compute $Y = P^T X$. -Solve the self-representation problem (7) to obtain C . **Step 2: Projection learning** -Set $M = \lambda_1 (X - XC)(X - XC)^T - \lambda_2 XX^T$ -Set P to be the d eigenvectors corresponding to the bottom d eigenvalues of M . **Until** stopping conditions reached. **Output:** P **and** C **.**

4. EXPERIMENTS

In this section, we apply our proposed method on two typical subspace clustering tasks: motion segmentation and face clustering. Several self-representation based subspace clustering algorithms such as Least square Regression (LSR) [17], Low-Rank Representation (LRR) [3], SSC [13], Smooth Representation Clustering (SMR)[18] and Latent Space Sparse Subspace Clustering (LS3C) [19] are compared. For all the algorithms, we use Frobenius norm for the self-reconstruction error term for fair comparison. The parameter settings of our method on the two tasks are given in corresponding sections below.

4.1 Motion Segmentation

We evaluate our method for motion segmentation task on the Hopkins155 motion segmentation database [1], which contains 155 video sequences. For each sequence, a number of point trajectories are extracted using standard tracking methods. The task of motion segmentation is to cluster these point trajectories in accordance with different motions in the video sequences. For methods other than LS3C and our proposed SMR_LP, the data is project into a subspace of dimension 12 using PCA as described in their articles. We choose maximum iteration number t = 10 and λ_1=λ_2=20000 .Table 1 reports clustering errors of different methods. As can be seen, our method SMR_LP significantly improves the performance of origin SMR and achieves the best results.

Table 1. Clustering errors on the Hopkins155

Algorithm	Median (%)	Mean (%)
SSC	0	2.41
LRR	0	3.74
LSR	0.28	2.30
SMR	0.21	2.34
LS3C	0	2.31
SMR_LP	0	**1.87**

4.2 Face Clustering

We use Extended Yale Face B [2] datasets to test our algorithm for face clustering task. It contains 38 classes of 64 face images under varying illuminations. We use data of the first 10 classes, and resize them to 32×32. Similarly, the data is project into a subspace of dimension 10×6 using PCA for methods other than LS3C and our proposed SMR_LP. We choose maximum iteration number t = 10 and $\lambda_1 = \lambda_2$ = 20. Table 2 presents the clustering errors of different methods on Extended Yale Face B datasets. As can be seen, our SMR_LP also outperforms the others.

Table 2. Clustering errors on Extended Yale B

Algorithm	Clustering error (%)
SSC	48.81
LRR	35.00
LSR	27.50
SMR	26.56
LS3C	41.35
SMR-LP	**24.73**

6. Conclusions

In this paper, we propose a framework for simultaneous dimensionality reduction and self-representation coding and introduce a new algorithm SMR_LP. Experiments on benchmark datasets demonstrate the effectiveness of our method. Our method can also be seen a new dimensionality reduction method for data lying in multiple subspaces. In the future, we plan to study it in a wider scope, e.g., feature extraction and face recognition.

7. Acknowledgement

This work is supported by the NSFC 61273274, 61370127 and 61201158, NSFB4123104, FRFCU 2014JBZ004, Z131110001913143

7. References

[1] R. Tron, R. Vidal, and A. Ravichandran. A benchmark for the comparison of 3-d motion segmentation algorithms. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8, 2007.

[2] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE TPAMI, 23(6):643–660, 2001.

[3] G. Liu, Z. Lin, and Y. Yu. Robust subspace segmentation by low-rank representation. In International Conference on Machine Learning, 2010.

[4] P. Tseng, "Nearest q-flat to m points," Journal of Optimization Theory and Applications, vol. 105, no. 1, pp. 249–252, 2000.F

[5] T. Zhang, A. Szlam, and G. Lerman, "Median k-flats for hybrid linear modeling with many outliers," in Workshop on Subspace Methods, 2009.

[6] M. Tipping and C. Bishop, "Mixtures of probabilistic principal component analyzers," Neural Computation, vol. 11, no. 2, pp. 443–482, 1999.

[7] Y. Sugaya and K. Kanatani, "Geometric structure of degeneracy for multibody motion segmentation," in Workshop on Statistical Methods in Video Processing, 2004.

[8] J. Costeira and T. Kanade, "A multibody factorization method for independently moving objects." Int. Journal of Computer Vision, vol. 29, no. 3, 1998.

[9] K. Kanatani, "Motion segmentation by subspace separation and model selection," in IEEE Int. Conf. on Computer Vision, vol. 2, 2001, pp. 586–591.

[10] C. W. Gear, "Multibody grouping from motion images," Int. Journal of Computer Vision, vol. 29, no. 2, pp. 133–150, 1998.

[11] R. Vidal, Y. Ma, and S. Sastry, "Generalized Principal Component Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 12, pp. 1–15, 2005.

[12] E. Elhamifar and R. Vidal. Sparse subspace clustering. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2790–2797, 2009.

[13] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. IEEE Trans. Pattern Anal. Mach. Intell., 2013.

[14] G. Liu, Z. Lin, and Y. Yu. Robust subspace segmentation by low-rank representation. In International Conference on Machine Learning, 2010.

[15] G. Liu and S. Yan. Latent low-rank representation for subspace segmentation and feature extraction. In IEEE International Conference on Computer Vision, pages 1615–1622, 2011.

[16] R. Vidal and P. Favaro. Low rank subspace clustering. Pattern Recognition Letters, 2013.

[17] C. Y. Lu, H. Min, Z.-Q. Zhao, L. Zhu, D.-S. Huang, and S. Yan. Robust and efficient subspace segmentation via least squares regression. In ECCV, pages 347–360, 2012.

[18] Han Hu, Z. Lin, J. Feng and Jie Zhou, "Smooth Representation Clustering," IEEE Conference on Computer Vision and Pattern Recognition, 2014

[19] Vishal M. Patel, H. Van Nguyen, R. Vidal, "Latent Space Sparse Subspace Clustering," In IEEE International Conference on Computer Vision,2013

[20] J. C. Bezdek1 and R. J. Hathaway, "Some notes on alternating optimization, "in Lecture Notes in Artificial Intelligence, N. R. Pal and M.Sugeno, Eds. Berlin, Germany: Springer-Verlag, 2002, vol. 2275, pp.288–300.