

Locality-constrained least squares regression for subspace clustering

Yuanyuan Chen, Zhang Yi*

Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, PR China



ARTICLE INFO

Article history:

Received 8 April 2018

Received in revised form 9 August 2018

Accepted 12 August 2018

Available online 16 August 2018

Keywords:

Subspace clustering

Locality

Least squares regression

Low-rank representation

ABSTRACT

Low-rank representation (LRR) is a typical data representation method that has been developed in recent years. Based on the main idea of LRR, least squares regression (LSR) constructs a new optimization problem to group the highly correlated data together. Compared with other LRR algorithms that are sophisticated, LSR is simpler and more efficient. It is a linear representation method that captures the global structure of data with a low-rank constraint. Research has shown that locality constraints typically improve the discrimination of the representation and demonstrate better performance than those “non-local” methods for image recognition tasks. To combine LSR and the locality constraints, we propose a novel data representation method called locality-constrained LSR (LCLSR) for subspace clustering. LCLSR forces the representation to prefer the selection of neighborhood points. The locality constraint is calculated based on the Euclidean distances between data points. Under the locality constraint, the obtained affinity matrix captures the locally linear relationship for data points lie on a nonlinear manifold. We propose three approaches to obtain the locality constraint and compare their performance with other related work on subspace clustering. We conducted extensive experiments on several datasets for subspace clustering. The results demonstrated that the proposed DCLSR, LCLSR_s, and LCLSR_k substantially outperformed state-of-the-art subspace clustering methods.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Subspace clustering is one of the most important tools in the quest to analyze and interpret big data, particularly for clustering of high-dimensional data, such as image, voice, video, and gene data. Given a dataset drawn from a union of linear subspaces that consists of massive high-dimensional data points, subspace clustering aims to group points according to some certain similarity metrics and grouping rules. In most applications, such as image processing, video analysis and DNA microarray analysis, the underlying subspaces of the data are low dimensional. We can consider subspace clustering as an extension of feature selection or data representation that attempts to determine clusters in relatively low-dimensional subspaces of the same dataset. Among different subspace clustering methods, spectral clustering is an important branch for managing high-dimensional data and is attracting an increasing amount of researchers' attention.

One key step in the spectral clustering algorithm is to construct a “good” affinity matrix. The classical spectral clustering algorithm constructs an affinity matrix by calculating the pairwise distance between each pair of points. In 2009, Wright et al. [1] proposed

sparse representation (SR) for robust face recognition. This linear representation-based method reveals the linear relationship between data. The application of the obtained representation coefficients to the classification task resulted in a breakthrough in face recognition, which has created an upsurge in applying representation learning to data analysis and machine learning. In [2], the authors proposed the nearest subspace classifier. An approximation of the testing data was calculated using a linear combination of available training samples per class. The testing data was labeled according to the class that minimized the reconstruction error. In [3], Zhang et al. indicated that it is the collaborative representation rather than ℓ_1 norm sparsity that makes SRC powerful for face classification. In [4], Elhamifar et al. proposed the sparse subspace clustering (SSC) algorithm, which constructs the data affinity matrix using the SR coefficient. Peng et al. extended SSC, and proposed scalable sparse subspace clustering (SSSC) [5].

SR usually reveals the linear relationship between a single point and other points. To capture the global structure of the data, Liu et al. proposed low-rank representation (LRR) [6] for subspace clustering. LRR aims to solve an optimization problem to determine a matrix to represent the data using points in a dictionary. Compared with the traditional algorithm that uses distance to measure data similarity, representation learning based algorithms, such as SSC, or LRR, achieve better clustering results overall. However, because

* Corresponding author.

E-mail address: zhangyi@scu.edu.cn (Z. Yi).

the objective function in most LRR problems is not differentiable, it requires a relatively long time to solve the rank minimization problem in real-world applications. Researchers have attempted to determine new optimization problems to solve LRR more efficiently [7,8]. In [9], the authors proposed least squares regression (LSR) for subspace clustering. LSR groups the highly correlated data together, and also it is robust to noise. Compared with other LRR algorithms, LSR is simpler and more efficient.

These representation-based algorithms can be summarized as constructing and solving an optimization problem, and then using the coefficients to represent the similarity between data. From SR to LRR, the given data is typically represented by the points in a dictionary. Considering the subspace clustering task, these algorithms construct an affinity matrix by solving the representation coefficients between points in the dataset. The coefficients represent the similarity between data. Therefore, constructing an affinity matrix using the optimal coefficients followed by spectral clustering results in better performance than a number of traditional methods.

Under the sparse or low-rank constraints on the coefficients, the optimal solution reveals the linear relationship between data very well. Usually, these methods assume that the data can be linearly represented with respect to each other. However, the relation between data in real-world often shows highly nonlinear. Then these aforementioned algorithms might achieve degraded performance. Some researchers realized it and they tried to search novel methods to reveal the nonlinear relation between data points of interests. In [10–12], the authors used deep neural networks to search latent representation of data. Fuzzy rules were also applied to establish optimization problems, which showed better performance [13–15]. In [16], the authors proposed the famous locality linear embedding (LLE), which is an unsupervised learning algorithm that computes low-dimensional, neighborhood-preserving embeddings of high-dimensional nonlinear inputs. LLE assumes that each data point and its neighbors lie on or close to a locally linear patch of the nonlinear manifold. What is more, researchers [17] have empirically observed that, in the results of sparse coding, nonzero coefficients are often assigned to bases that are near to the query data. They have proposed local coordinate coding, which explicitly encourages the coding to be local. Moreover, locality features have been applied in many representation learning algorithms for classification tasks [18,19]. In [20], Peng et al. proposed locality-constrained collaborative representation for face recognition. The study demonstrated that locality constraints typically improve the discrimination of the representation and demonstrate better performance than those “non-local” methods for image recognition.

Inspired by the progress in research on representation learning and image classification, we combine locality constraints with LSR for the subspace clustering problem. Under the locality constraints, the affinity matrix obtained not only capture the global structure of data points, but also reveal the locally linear relationship for data points lie on a nonlinear manifold. In this paper, We propose a novel optimization problem that combines the locality feature of data and representation learning using the LSR method. Under the constraint of a pre-calculated similarity matrix that describes the locality relationships between data, the coefficient matrix prefers to select nearby points to represent the query data. The solution to the optimization problem reveals the structure of the dataset, and thus, performs better while performing subspace clustering.

The remainder of this paper is organized as follows: In Section 2, we briefly discuss recent related work on representation learning, particularly LRR and locality constraints for representation learning. In Section 3, we present the objective functions of distance-constrained LSR (DCLSR) and locality-constrained LSR (LCLSR) and their optimal solution. We report the results of experiments and

discuss the analysis in Section 4. We present our conclusions and further studies in Section 5.

2. Related work

2.1. Preliminaries

Given dataset $X = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{d \times N}$ that contains N data points which drawn from K different subspaces, spectral clustering methods construct affinity matrix W . W_{ij} denotes the similarity between data points x_i and x_j according to certain distance metrics.

For example, $W_{ij} = \sqrt{(x_i - x_j)^2}$ calculates the Euclidean distance between two data points x_i and x_j in the dataset. A “good” affinity matrix produces high quality clusters that have high within-cluster similarity and low between-class similarity. In recent years, representation learning-based algorithms that have constructed the affinity matrix by solving an optimization problem to obtain a coefficient matrix have achieved state-of-the-art performance in subspace clustering. Of these representation learning-based algorithms for subspace clustering, LRR is particularly attractive.

2.2. Low-rank representation

In [6], Liu et al. proposed an LRR to determine the coefficient matrix Z of data X with the lowest rank:

$$\min_Z \text{rank}(Z) \quad \text{s.t.} \quad X = AZ, \quad (1)$$

where A is a dictionary of the data. When applying LRR to doing subspace clustering, the dataset X can be set as the dictionary. LRR attempts to reveal the similarity between data by determining the representation coefficients. If the data in X are arranged to satisfy the true segmentation of the data, then the optimal solution to (1), Z^* , is block diagonal:

$$Z^* = \begin{bmatrix} Z_{11}^* & 0 & \dots & 0 \\ 0 & Z_{22}^* & \dots & 0 \\ 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Z_{kk}^* \end{bmatrix}. \quad (2)$$

Considering this representation matrix as the affinity matrix for spectral clustering, the between-cluster affinities are all zeros, which captures both dense within-class affinities and zero between-class affinities. The original LRR problem is not easy to solve because the rank function in the objective function is discrete. Based on Liu’s main concept, many researchers have attempted to search for another approximate objective function to improve the efficiency of the original algorithm [7–9,21].

By solving a nuclear norm minimization problem, Favaro [7] proposed a closed-form solution to the low-rank minimization problem:

$$\min_Z \|Z\|_* + \frac{\lambda}{2} \|X - XZ\|_F^2, \quad (3)$$

where $\|Z\|_*$ is the nuclear norm of Z . The optimal solution to Eq. (3) has a closed form by performing singular value decomposition on X . This algorithm is called low-rank subspace clustering (LRSC), which is more efficient than the classical LRR methods.

2.3. Least squares regression

Among massive studies on regression problems [15,22,23], subspace clustering by solving regression problems has attracted more and more attention [9,24,25]. In 2012, Lu et al. proposed a robust and efficient subspace segmentation algorithm [9] using LSR. By

applying the Frobenius norm, LSR solves the following optimization problem:

$$\min_Z \|X - XZ\|_F^2 + \frac{\lambda}{2} \|Z\|_F^2, \quad (4)$$

where $\lambda > 0$ is a parameter used to balance the two terms. Eq. (4) has an analytical solution:

$$Z = (X^T X + \lambda I)^{-1} X^T X \quad (5)$$

Compared with most LRR problems, LSR can be solved more efficiently. According to the experimental results in [9], LSR also achieves higher clustering accuracy for both image clustering and motion segmentation.

3. Locality-constrained least squares regression (LCLSR)

3.1. Optimization problem

Considering dataset $X \in \mathbb{R}^{d \times N}$ that contains N data points, the pairwise distances are calculated and stored in a matrix D . Here D_{ij} is defined as:

$$D_{ij} = e^{\text{dist}(x_i, x_j)^2}, \quad (6)$$

where the function $\text{dist}(x_i, x_j)$ is a certain distance metric, such as the Euclidean distance that we typically use. To make the distance have a value in the range $[0, 1]$, we typically normalize it by dividing it by the maximum distance of each point x_i . Therefore, D_{ij} represents the distance-based similarity between point x_i and x_j and its value is in the range $[e, 1]$. Smaller values of D_{ij} indicate that the two related points are closer to each other.

Considering distance matrix D between data as the constraint, DCLSR attempts to solve the following optimization problem:

$$\arg \min_Z \frac{1}{2} \|X - XZ\|_F^2 + \frac{\lambda_1}{2} \|Z\|_F^2 + \frac{\lambda_2}{2} \|Z \odot D\|_F^2, \quad (7)$$

where λ_1 and λ_2 are the two balance parameters, and the symbol “ $\odot$ ” denotes the Hadamard product. The first two items in Eq. (7) are the same as those in the optimization problem in LSR [9]. The third term represents the constraint that the coefficients matrix prefers to select the closer points to represent a certain point.

3.2. Solution

It is not an easy work to solve problem (7) directly in matrix form.

We rewrite the distance matrix D in another form as

$$D = [d_1, d_2, \dots, d_N], \quad (8)$$

where the vector d_i represents the distance between x_i and other points.

Similarly, the matrix Z is rewritten as:

$$Z = [z_1, z_2, \dots, z_N]. \quad (9)$$

We define two intermediate variables $A = X^T X$ and $B = A + \lambda_1 \cdot I$, where I is the identity matrix. We rewrite A as $A = [a_1, a_2, \dots, a_N]$.

Then, the solution to (7) is written in vector form as:

$$z_i = (B + \lambda_2 \cdot G_i \cdot G_i)^{-1} \cdot a_i, \quad (10)$$

where

$$G_i = \text{diag}(d_i) \\ = \begin{bmatrix} d_{i1} & 0 & \dots & 0 \\ 0 & d_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{iN} \end{bmatrix}. \quad (11)$$

It can be easily proved that the matrix $B + \lambda_2 \cdot G_i \cdot G_i$ is a positive definite matrix because all the elements in d_i is greater than zero. Therefore, $B + \lambda_2 \cdot G_i \cdot G_i$ is reversible.

3.3. Locality constraint

In Eq. (7), other methods can be used in addition to Eq. (6) to compute constraint matrix D . For example, the distance matrix D calculated by Eq. (6) can be converted to a locality constraint by setting some small distances to value of zero. In this paper, we attempt to apply two locality constraints that are frequently used to constrain representation matrix Z : k nearest neighbor (k -NN) and ε -ball. k -NN constraint set k largest values in each row of D be nonzero, while ε -ball constraint set the distance values that smaller than ε be zeros. Then, the related matrix is denoted by $\hat{D}$, which is used as the constraint in Eq. (7). The zeros in $\hat{D}$ force Z to have more non-zero coefficients in positions that correspond to nearby points.

In [9], the authors proposed two optimization problems called LSR1 and LSR2. LSR2 has another constraint on the objective function. That is,

$$\arg \min_Z \frac{1}{2} \|X - XZ\|_F^2 + \frac{\lambda}{2} \|Z\|_F^2, \quad \text{s.t. } Z_{ii} = 0. \quad (12)$$

Z_{ii} is the self-representation coefficient. The constraint $Z_{ii} = 0$ emphasize the representation more on the relation between points, i.e., samples in the dataset.

Adding the constraint $Z_{ii} = 0$ used in LSR2, problem (7) becomes:

$$\arg \min_Z \frac{1}{2} \|X - XZ\|_F^2 + \frac{\lambda_1}{2} \|Z\|_F^2 + \frac{\lambda_2}{2} \|Z \odot D\|_F^2, \quad \text{s.t. } Z_{ii} = 0. \quad (13)$$

It is not easy to solve problem (13) directly. To guarantee that the elements on the diagonal of Z equal to zero, the elements on the diagonal of D are set to large values. Therefore, the solution to Eq. (13) has the same form as that of Eq. (7). The only difference is the value of the pre-calculated matrix D .

3.4. Algorithm

In conclusion, the LCLSR algorithm for subspace clustering is described as follows:

Algorithm 1 LCLSR

Input:

Input dataset, X . Appropriate values of parameters, such as λ_1 , λ_2 .

Constraint calculation:

Calculate the constraint matrix D (using Eq. (6)) or $\hat{D}$ (using ε -ball or k -NN) on X .

Optimization:

Solve the optimization Eq. (7) or (13) using Eq. (10), to obtain the coefficient matrix Z .

Clustering:

Perform spectral clustering on the representation Z using some methods.

Output:

Clustering results.

4. Experiments

In this section, we report the performance of the proposed DCLSR and two different LCLSR methods for subspace clustering on several image datasets. The two LCLSR methods are denoted by LCLSR $_{\varepsilon}$ for ε -ball and LCLSR $_k$ for the k -NN constraint. We compare



Fig. 1. Example images in the datasets that used in the experiments.

Table 1

Performance of different subspace clustering algorithms on the ORL dataset(%).

Alg.	PCA				HOG feature			
	para.	ACC	NMI	ARI	para.	ACC	NMI	ARI
DCLSR	(0.8,3)	84.63 ± 1.87	92.09 ± 0.65	77.91 ± 1.68	(0.8,3)	75.72 ± 0.75	85.93 ± 0.23	62.15 ± 1.20
LCLSR _{ε}	(1,2)	85.01 ± 1.08	91.56 ± 0.26	77.51 ± 0.63	(1,2)	74.13 ± 0.89	85.66 ± 0.32	61.21 ± 0.81
LCLSR _{k}	(0.8,0.9)	86.25 ± 0.75	92.52 ± 4.84	79.08 ± 0.48	(0.5,0.4)	73.75 ± 1.51	83.93 ± 1.22	60.26 ± 2.13
LSR	0.8	80.17 ± 1.36	89.14 ± 1.04	71.25 ± 2.27	1	74.01 ± 1.51	85.44 ± 0.76	61.53 ± 1.23
LRR	0.6	78.83 ± 2.18	88.53 ± 0.53	68.97 ± 1.57	0.7	76.53 ± 2.03	85.59 ± 0.08	62.67 ± 0.06
LRSC	2	80.17 ± 0.94	89.10 ± 0.31	70.97 ± 0.49	1	74.75 ± 2.13	85.60 ± 0.45	62.72 ± 0.91
SSC	8	76.58 ± 0.83	87.90 ± 1.91	66.01 ± 1.01	7	71.94 ± 0.45	84.28 ± 0.56	59.38 ± 1.26

the proposed methods with other representation-based subspace clustering algorithms that proposed in recent years. All the experiments were conducted using MATLAB 2015b on a 2.20 GHz PC with 3.25 GB RAM. Considering the randomness of k -means that is applied in spectral clustering, we report the mean and standard variance results of 10 independent experiments.

4.1. Datasets

In our experiments, we used four well-known datasets: ORL [26] and three AR [27] datasets. The details of these datasets are described as follows and some images contained in the datasets are shown in Fig. 1.

The AR dataset contains over 4000 color images that corresponds to 126 people's faces. These faces have different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). Accordingly, the dataset can be divided into three parts: AR (without any occlusion), AR-glass(with sun glasses) and AR-scarf (with scarf). The images in the dataset are resized to 55×40 .

The ORL dataset consists of 10 frontal face images each for 40 human subjects. For some subjects, the images were taken at different times and have varied lighting, facial expressions, and facial details. All the images in the ORL database are resized to 56×46 .

For a more comprehensive study, we performed three different feature extraction on the raw data rather than using the raw data directly; that is, principle component analysis (PCA) and the histogram of oriented gradients (HOG). For computational efficiency, we performed PCA to reduce the features calculated by HOG.

4.2. Experimental settings

We compared the proposed DCLSR, LCLSR _{ε} and LCLSR _{k} algorithms with well-known representation learning-based subspace clustering algorithms, LSR [9], LRR [6], LRSC [7] and SSC [4], using different settings. The experiments of LCLSR in this section were carried out based on problem (13).

For different datasets, the best parameter of each algorithm was also different. It is difficult to select appropriate parameters for

all datasets. Some hyper-parameters used in the algorithms were tuned using a grid search, which allowed us to test different combinations of hyper-parameters and determine the combination with the best performance. In LCLSR _{ε} , because the matrix D has been normalized by dividing the maximum distance of each point, we set ε to 0.3. In LCLSR _{k} , k is set to half of the number of data points, i.e., $k = 0.5 \times N$. The other related parameters of each algorithm are also reported accompany with results.

4.3. Clustering results

The clustering performance of these algorithms on the four datasets are reported in Tables 1–4. The bold values indicate the best performance according to the highest accuracy. The parameters reported are: λ_1 and λ_2 for the three proposed algorithms; λ for LRR, LRSC, and LSR; and α for SSC.

According to the results shown in Tables 1–4, it can be seen that the three proposed algorithms outperformed the other subspace clustering methods in a whole. Particularly when performing PCA, all the proposed algorithms demonstrated better performance than the other algorithms. Among the three proposed algorithms, LCLSR _{k} achieved best performance overall.

4.4. Discussion

The LCLSR added a locality constraint on LSR. Therefore, considering the optimization problem (13), two parameters λ_1 and λ_2 affected the performance. In this section, we compared the performance of LCLSR with different λ_1 and λ_2 on ORL dataset after PCA feature extraction. LCLSR _{k} was used because it had the best performance on the ORL dataset, as shown in Table 1. In Fig. 2, it can be observed that the best performance was achieved when $\lambda_1 = 1$. Furthermore, when λ_1 took different values, the performance (ACC, NMI and ARI) of LCLSR increased with the increase of λ_2 as a whole.

5. Conclusion and further study

We proposed three locality-constrained LSR algorithms: DCLSR, LCLSR _{ε} , and LCLSR _{k} . We constructed a novel optimization problem,

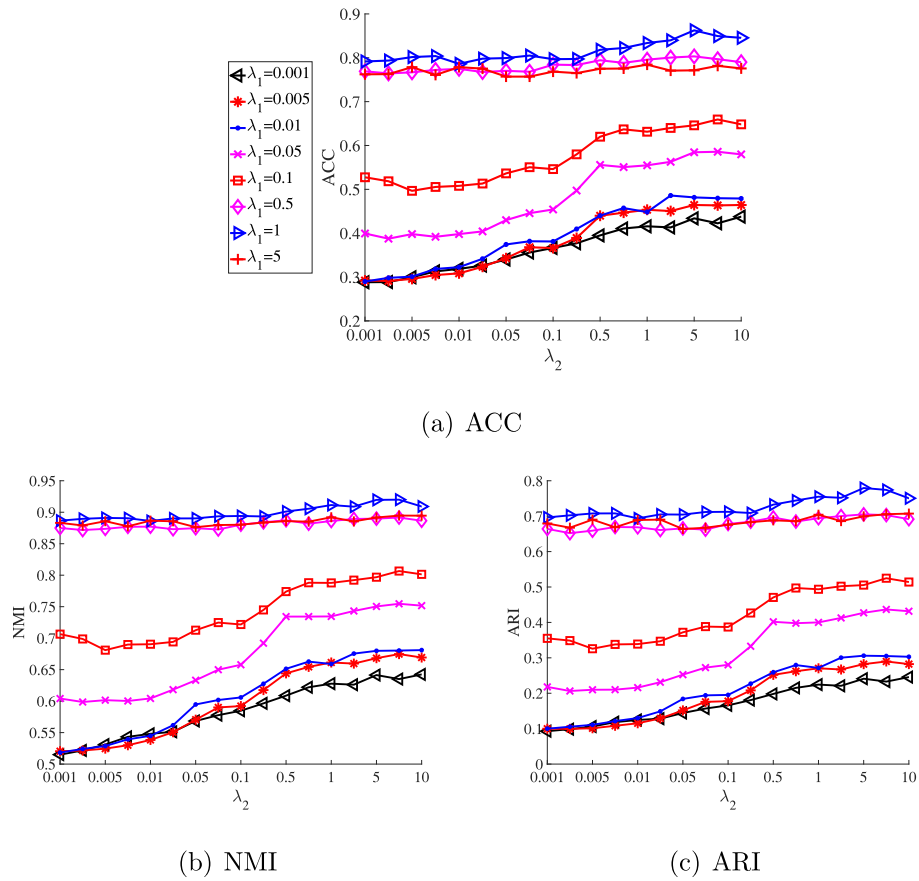


Fig. 2. Performance of LCLR algorithms on ORL dataset with different parameters (λ_1 and λ_2).

Table 2

Performance of different subspace clustering algorithms on the AR dataset (%).

Alg.	PCA				HOG feature			
	para.	ACC	NMI	ARI	para.	ACC	NMI	ARI
DCLR	(0.2,0.2)	85.81 $\pm$ 1.57	93.00 $\pm$ 0.51	76.78 $\pm$ 1.94	(0.7,0.7)	83.86 $\pm$ 0.66	91.36 $\pm$ 0.21	72.09 $\pm$ 1.36
LCLR _r	(0.1,0.1)	86.62 $\pm$ 0.44	93.32 $\pm$ 0.39	77.53 $\pm$ 0.93	(3,0.5)	83.02 $\pm$ 1.53	91.83 $\pm$ 0.78	72.39 $\pm$ 2.43
LCLR _k	(0.3,0.7)	85.67 $\pm$ 1.27	93.31 $\pm$ 0.53	78.33 $\pm$ 1.68	(3,3)	85.76 $\pm$ 1.12	92.88 $\pm$ 0.36	76.47 $\pm$ 0.34
LSR	0.5	82.91 $\pm$ 0.34	92.51 $\pm$ 0.13	75.62 $\pm$ 0.37	3	83.93 $\pm$ 0.46	91.94 $\pm$ 0.11	72.99 $\pm$ 0.26
LRR	1	84.24 $\pm$ 1.13	92.47 $\pm$ 0.15	75.25 $\pm$ 1.19	0.3	81.79 $\pm$ 0.99	91.21 $\pm$ 0.42	71.37 $\pm$ 0.52
LRSC	0.3	84.41 $\pm$ 0.84	92.33 $\pm$ 0.62	75.07 $\pm$ 1.13	0.5	83.50 $\pm$ 0.36	91.86 $\pm$ 0.10	72.46 $\pm$ 0.54
SSC	10	75.11 $\pm$ 2.62	88.62 $\pm$ 1.47	66.08 $\pm$ 1.34	5	85.29 $\pm$ 1.57	93.78 $\pm$ 0.22	79.35 $\pm$ 1.32

Table 3

Performance of different subspace clustering algorithms on the AR-glass dataset(%).

Alg.	PCA				HOG feature			
	para.	ACC	NMI	ARI	para.	ACC	NMI	ARI
DCLR	(0.1,0.1)	73.93 $\pm$ 0.39	86.54 $\pm$ 0.46	61.96 $\pm$ 1.69	(2,3)	61.39 $\pm$ 1.29	79.06 $\pm$ 0.68	46.13 $\pm$ 1.76
LCLR _r	(0.2,0.1)	74.08 $\pm$ 0.71	86.52 $\pm$ 0.17	62.19 $\pm$ 0.30	(2,1)	62.21 $\pm$ 0.13	79.81 $\pm$ 0.10	47.08 $\pm$ 0.22
LCLR _k	(0.1,0.3)	75.67 $\pm$ 0.20	87.45 $\pm$ 0.21	63.52 $\pm$ 0.78	(0.5,0.2)	48.92 $\pm$ 1.29	71.06 $\pm$ 0.64	32.44 $\pm$ 1.18
LSR	0.1	72.75 $\pm$ 0.54	86.88 $\pm$ 0.59	60.71 $\pm$ 1.99	0.8	62.28 $\pm$ 1.16	78.81 $\pm$ 0.96	45.66 $\pm$ 1.57
LRR	5	67.58 $\pm$ 2.29	82.83 $\pm$ 0.92	52.04 $\pm$ 2.04	5	49.92 $\pm$ 1.16	71.04 $\pm$ 0.88	30.84 $\pm$ 1.50
LRSC	13	73.42 $\pm$ 0.01	86.76 $\pm$ 0.68	61.36 $\pm$ 0.62	1	63.83 $\pm$ 1.33	80.83 $\pm$ 0.87	48.90 $\pm$ 1.69
SSC	33	70.25 $\pm$ 0.35	85.58 $\pm$ 0.45	57.82 $\pm$ 0.98	13	65.69 $\pm$ 1.06	82.51 $\pm$ 0.69	51.26 $\pm$ 1.71

in which a pre-calculated similarity matrix was used to allow the data to prefer the selection of nearby points nearby to represent themselves. Therefore, the representation coefficient matrix revealed the nonlinear relationship between data points, and

grouped the points in the same class together. Extensive experimental results demonstrated that the proposed LCLR outperformed state-of-the-art subspace clustering methods on different datasets overall.

Table 4

Performance of different subspace clustering algorithms on the AR-scarf dataset (%).

Alg.	PCA				HOG feature			
	para.	ACC	NMI	ARI	para.	ACC	NMI	ARI
DCLSR	(0.1,1)	73.56 ± 1.33	86.65 ± 0.15	61.59 ± 0.99	(2,1)	47.78 ± 1.46	72.17 ± 0.68	31.55 ± 1.15
LCLSR _e	(0.1,0.5)	74.53 ± 1.02	87.47 ± 0.29	63.55 ± 0.32	(2,0.2)	48.90 ± 1.30	72.54 ± 0.69	32.02 ± 1.16
LCLSR _k	(0.2,0.2)	75.28 ± 0.62	87.63 ± 0.19	64.16 ± 0.61	(3,0.3)	50.75 ± 1.29	73.92 ± 0.29	35.72 ± 0.94
LSR	0.1	72.44 ± 0.97	86.52 ± 0.39	60.21 ± 0.99	2	50.08 ± 2.36	73.33 ± 1.21	33.71 ± 2.31
LRR	2	65.69 ± 1.75	82.77 ± 0.45	54.43 ± 4.18	5	35.48 ± 1.55	63.34 ± 0.66	17.52 ± 1.18
LRSC	13	73.92 ± 1.29	87.23 ± 8.43	62.39 ± 1.77	1	49.22 ± 0.28	73.53 ± 0.18	32.86 ± 0.54
SSC	25	65.97 ± 1.39	83.87 ± 0.49	53.39 ± 1.51	11	47.81 ± 0.39	74.38 ± 0.35	32.30 ± 4.56

Recently, deep neural networks have been successfully applied to solve supervised and unsupervised learning tasks. One of the main concepts in convolutional neural networks is receptive field. It reduces connection to local regions. In future work, we plan to combine the concept of locality constraint with neural networks and investigate the performance of our method on supervised tasks.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 61702349, 61432012, and U1435213) and Science and Technology Department of Sichuan Province (Grant No. 2017JY0265).

References

- [1] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (2) (2009) 210–227.
- [2] K.C. Lee, J. Ho, D.J. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (5) (2005) 684–698, <http://dx.doi.org/10.1109/TPAMI.2005.92>.
- [3] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: Which helps face recognition? in: *IEEE International Conference on Computer Vision*, 2011, pp. 471–478.
- [4] E. Elhamifar, R. Vidal, Sparse subspace clustering, in: *IEEE International Conference on Computer Vision and Pattern Recognition*, 2009, pp. 2790–2797.
- [5] X. Peng, L. Zhang, Z. Yi, Scalable sparse subspace clustering, in: *IEEE International Conference on Computer Vision and Pattern Recognition*, 2013, pp. 430–437.
- [6] G. Liu, Z. Lin, Y. Yu, Robust subspace segmentation by low-rank representation, in: *International Conference on Machine Learning*, 2010, pp. 663–670.
- [7] P. Favaro, R. Vidal, A. Ravichandran, A closed form solution to robust subspace estimation and clustering, in: *IEEE International Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1801–1807.
- [8] Y. Chen, L. Zhang, Z. Yi, A novel low rank representation algorithm for subspace clustering, *Int. J. Pattern Recognit. Artif. Intell.* 30 (04) (2016) 1650007.
- [9] 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*, Springer, 2012, pp. 347–360.
- [10] X. Peng, J. Feng, S. Xiao, W.-Y. Yau, J. Zhou, S. Yang, Structured AutoEncoders for subspace clustering, *IEEE Transaction on Image Processing* 27 (10) (2018) 5076–5086.
- [11] Y. Chen, L. Zhang, Z. Yi, Subspace clustering using a low-rank constrained autoencoder, *Inform. Sci.* 424 (2017) 28–38.
- [12] K.G. Dizaji, A. Herandi, C. Deng, T.W. Cai, H. Huang, Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization, in: *IEEE International Conference on Computer Vision*, 2017, pp. 5747–5756.
- [13] P.J.F. Groenen, K. Jajugab, Fuzzy clustering with squared Minkowski distances, *Fuzzy Sets and Systems* 120 (2) (2006) 227–237.
- [14] L. Zhang, W. Lu, X. Liu, W. Pedrycz, C. Zhong, Fuzzy c-means clustering of incomplete data based on probabilistic information granules of missing values, *Knowl.-Based Syst.* 99 (2016) 51–70.
- [15] H. Zuo, G. Zhang, W. Pedrycz, V. Behbood, J. Lu, Fuzzy regression transfer learning in Takagi-Sugeno fuzzy models, *IEEE Trans. Fuzzy Syst.* 25 (6) (2017) 1795–1807.
- [16] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science* 290 (5500) (2000) 2323–2326.
- [17] K. Yu, T. Zhang, Y. Gong, Nonlinear learning using local coordinate coding, in: *Advance in Neural Information Processing Systems*, 2009, pp. 2223–2231.
- [18] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in: *IEEE International Conference on Computer Vision and Pattern Recognition*, IEEE, 2010, pp. 3360–3367.
- [19] D. Arpit, G. Srivastava, Y. Fu, Locality-constrained low rank coding for face recognition, in: *International Conference on Pattern Recognition*, IEEE, 2012, pp. 1687–1690.
- [20] X. Peng, L. Zhang, Z. Yi, K.K. Tan, Learning locality-constrained collaborative representation for robust face recognition, *Pattern Recognit.* 47 (9) (2014) 2794–2806.
- [21] R. Vidal, P. Favaro, Low rank subspace clustering (LRSC), *Pattern Recognit. Lett.* 43 (2014) 47–61.
- [22] A. Suarez, J.F. Lutsko, Globally optimal fuzzy decision trees for classification and regression, *IEEE Trans. Pattern Anal. Mach. Intell.* 21 (12) (1999) 1297–1311.
- [23] Q. Feng, Y. Zhou, R. Lan, Pairwise linear regression classification for image set retrieval, in: *IEEE International Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4865–4872.
- [24] B. Li, Y. Zhang, Z. Lin, H. Lu, Subspace clustering by Mixture of Gaussian Regression, in: *IEEE International Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2094–2102.
- [25] C. Peng, Z. Kang, Q. Cheng, Subspace clustering via variance regularized ridge regression, in: *IEEE International Conference on Computer Vision and Pattern Recognition*, 2017, pp. 682–691.
- [26] F.S. Samaria, A.C. Hartert, Parameterisation of a stochastic model for human face identification, in: *Workshop on Applications of Computer Vision*, 1994, pp. 138–142.
- [27] M. Aleix, B. Robert, The AR Face Database, *Tech. Rep.*, 1998.