



HAL
open science

Sparse Subspace Clustering Based on Adaptive Parameter Training

Kexuan Zhu, Min Li

► **To cite this version:**

Kexuan Zhu, Min Li. Sparse Subspace Clustering Based on Adaptive Parameter Training. 12th International Conference on Intelligent Information Processing (IIP), May 2022, Qingdao, China. pp.56-64, 10.1007/978-3-031-03948-5_5. hal-04178750

HAL Id: hal-04178750

<https://inria.hal.science/hal-04178750v1>

Submitted on 8 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Sparse Subspace Clustering Based on Adaptive Parameter Training

Kexuan Zhu¹, Min Li*

School of Information Engineering, Nanchang Institute of Technology

No. 289 Tianxiang Road, Nanchang Jiangxi, P. R. China

Abstract There are many researches on sparse subspace clustering, but there are few related studies on its parameter optimization. In this paper, we propose an adaptive training parameter method to improve the manual selection process of convex optimization regularization parameters and improve the accuracy of subspace clustering. Experiments were carried out on multiple datasets, and the clustering accuracy is improved. The results prove that the improved parameter training process can improve the clustering accuracy of subspace clustering.

Keywords Adaptive parameter training · Spectral clustering · Sparse subspace clustering · Curve fitting

¹ ORCID: 0000 - 0002 - 9434 - 4802

* Email address: liminghuadi@hotmail.com (Min Li)

1 Introduction

Cluster analysis belongs to the category of data mining and refers to the process of dividing a given physical or abstract object into classes composed of similar individuals. Traditional clustering methods can be roughly divided into division method, hierarchical method, grid-based method, density-based method, model-based method. The advantage of these clustering methods is that they can successfully cluster low-dimensional data [1]. However, with the development of computer technology and the improvement of people's ability to obtain data, the scale of the acquired data is larger and the data structure is more complicated. Such as images, texts, web documents, and gene expression data, their dimensionality often reaches hundreds or thousands of dimensions, or even higher. To solve the problem of clustering data in high-dimensional spaces, R. Agrawal proposed the Subspaces Clustering algorithm [2].

The subspace clustering algorithm has attracted much attention since it was proposed. Subspace clustering solves the problem of poor performance of traditional clustering algorithms on high-dimensional datasets. However, it is not practical to represent complex high-dimensional data only utilizing a single subspace, so scholars consider using multiple subspaces to solve this problem. How to better divide the subspace is the focus of subspace clustering research. After continuous improvement and experiments, the subspace clustering algorithm has been widely used in clustering problems of high-dimensional data in various fields [3].

Elhamifar and Vidal et al. proposed the Sparse Subspace Clustering algorithm (SSC) in 2009 [4]. The algorithm constructs an affinity matrix through self-representation between data and uses spectral clustering to obtain clustering results. To improve the performance of sparse subspace clustering, the following aspects are worth studying.

1. The design of regularization[5-6]makes the generated affinity matrix more conducive to the clustering process.
2. Robustness research[7-8], to obtain a good clustering effect under various noise interference.
3. Exploration in application[9-10].

In-depth optimization studies have been carried out in these aspects, and significant results have been achieved.

This paper aims to improve the clustering accuracy of the original sparse subspace. In this study, we improve the training process of regression parameters in the algorithm to obtain a better coefficient matrix, to have better clustering results. The results of experiments indicate that sparse subspace clustering based on adaptive parameter training proposed by us can improve clustering accuracy.

2 Sparse subspace clustering

The idea of the Sparse Subspace Clustering (SSC) algorithm is that the point x_i is represented by a linear combination of other data points in the subspace. A sparse coefficient matrix is obtained by imposing constraints. The obtained coefficient matrix is

input into the spectral clustering to obtain the clustering results. If we superimpose all data points in the same subspace to the column of the data matrix \mathbf{X} , SSC can be expressed as a linear equation $\mathbf{X} = \mathbf{X}\mathbf{C}$, where \mathbf{C} is a self-expression coefficient matrix.

Under the assumption that the subspaces are independent of each other, we can find the optimal structure of the coefficient matrix \mathbf{C} by minimizing some norms of \mathbf{C} . That is, solving:

$$\min_c \frac{1}{2} \|\mathbf{X} - \mathbf{X}\mathbf{C}\|_2^2 + \lambda \|\mathbf{C}\|_1, s. t. \text{diag}(\mathbf{C}) = 0 \quad (1)$$

The first term of the formula ensures the minimum error between the data represented by the coefficient matrix and the original data. The second term of the formula aims to solve the sparse form of the coefficient matrix \mathbf{C} . SSC needs to constrain the number of zero elements in \mathbf{C} to be much greater than the number of non-zero elements, that is, constrain $\min\|\mathbf{C}\|_0$. But the problem here is non-convex. the l_0 -norm constraint is replaced by the convex approximation problem $\min\|\mathbf{C}\|_1$. The third term of the formula restricts the diagonal elements of \mathbf{C} to 0 to avoid self-representation. The obtained coefficient matrix is processed to obtain the similarity matrix. Process the obtained coefficient matrix to obtain the similarity matrix:

$$\mathbf{W} = |\mathbf{C}| + |\mathbf{C}|^T \quad (2)$$

Then the graph is constructed with the similarity matrix \mathbf{W} , and the spectral clustering method is used for clustering. Let the Laplacian matrix of data \mathbf{X} is:

$$\mathbf{L}_c = \text{Diag}(\mathbf{W} * \mathbf{1}) - \mathbf{W} \quad (3)$$

According to graph theory, the Laplacian matrix of the graph is obtained by subtracting the weight matrix from the corresponding degree matrix. Similarly, the similarity matrix \mathbf{W} is used as the weight matrix when the original data is constructed as a graph. $\mathbf{1}$ is the column vector of all 1, then $\text{Diag}(\mathbf{W} * \mathbf{1})$ is the degree matrix of the graph.

The constructed Laplacian matrix is standardized by $\mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$ and the eigenvectors corresponding to the smallest k eigenvalues of the standardized \mathbf{L}_c are calculated. The k eigenvectors are formed into an eigen matrix \mathbf{f} , and then k -means clustering is performed on each row of the eigen matrix. The sparse subspace clustering process is as follows:

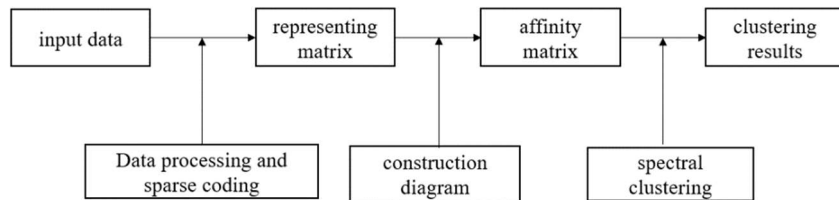


Figure 1 Flow chart of SSC model

3 Sparse subspace clustering based on adaptive parameter training

3.1 Algorithmic ideas

The sparse subspace clustering algorithm uses Lasso regression to train the coefficient matrix. The algorithm requires repeated experiments to train the regression parameters λ which is quite cumbersome. A large number of results of experiments show that the value of λ will affect the clustering results. To solve this problem, we propose an adaptive parameter training method, which avoids the process of selecting parameters manually. Concretely as the following optimization model:

$$\min_c \frac{1}{2} \|\mathbf{X} - \mathbf{XC}\|_2^2 + \lambda \|\mathbf{C}\|_1, s. t. \text{diag}(\mathbf{C}) = 0, \lambda \propto f(\|\mathbf{C}\|_1), \lambda > 0 \quad (4)$$

λ is used to constrain the two properties of the coefficient matrix. On the one hand, it constrains the sparseness of the coefficient matrix, on the other hand, it ensures that the matrix has a good representation ability[11].

If the number of subspaces k and the dimension d are known, Soltanolkotabi et al. in 2014 revealed the phenomenon found in the parameter training process: the l_1 -norm of \mathbf{C}_j will swing up and down according to a certain trend, changing the value of λ will cause the fluctuation of the l_1 -norm of \mathbf{C}_j to change. In the problem of motion segmentation, this trend is described by $1/\sqrt{d}$ [12]. Inspired by this when facing the data of unknown subspace dimensions, the method of curve fitting is used to describe the l_1 -norm of the coefficient matrix by us. Then the regular parameter is written $\lambda \propto f(\|\mathbf{C}\|_1)$.

The training model is updated to:

$$\min_c \frac{1}{2} \|\mathbf{X} - \mathbf{XC}\|_2^2 + \lambda_0 \|\mathbf{C}\|_1 + \lambda_i \|\mathbf{C}\|_1$$

$$s. t. \text{diag}(\mathbf{C}) = 0, \lambda_i \propto f(\|\mathbf{C}\|_1), \lambda_i > 0, i = 1, 2, \dots, n \quad (5)$$

According to the sparsity constraint of the original algorithm, the initial training sparse matrix regular parameter is 0.001[4]. Since repeated constraints and training will cause the performance of the coefficient matrix to deteriorate, the optimization model is modified as:

$$\min_c \frac{1}{2} \|\mathbf{X} - \mathbf{XC}\|_2^2 + \lambda_0 \|\mathbf{C}\|_1 + \lambda_i \|\mathbf{B}\|_1$$

$$s. t. \text{diag}(\mathbf{C}) = 0, \lambda_i \propto f(\|\mathbf{B}\|_1), \lambda > 0, \mathbf{C} = \mathbf{B}, \mathbf{XB} = \mathbf{X}, i = 1, 2, \dots, n \quad (6)$$

To retain the coefficient matrix with global sparsity and local performance, we take the coefficient matrix:

$$\mathbf{D} = (\mathbf{C} + \mathbf{B})/2 \quad (7)$$

Then the alternating minimization method is used to solve the model.

3.2 Model training process

Model updating

To facilitate the optimization process:

$$\min_c \frac{1}{2} \|X - XC\|_2^2 + \lambda_0 \|C\|_1 + f(\|B\|_1) \|B\|_1 \quad (8)$$

1. Update C

Fixed B, solve C^{t+1} corresponding subproblems:

$$C^{t+1} = \arg \min \frac{1}{2} \|X - XC\|_2^2 + \lambda_0 \|C\|_1 \quad (9)$$

Derivation is equal to 0, then C^{t+1} is:

$$C^{t+1} = E - \lambda_0 (X)^{-1} (X^{-1})^T \quad (10)$$

2. Update B

Fixed C to solve the corresponding problem B^{t+1} :

$$B^{t+1} = \arg \min f(\|B\|_1) \|B\|_1 \quad (11)$$

Let the above derivation:

$$f'(B)B + f(B) = 0 \quad (12)$$

Since the two-step operation is used in this paper, C and B are updated successively. And $C = B$ before updating B, so replace the fitting variable with the updated C^{t+1} to get:

$$f'(C)B + f(C) = 0 \quad (13)$$

Then B^{t+1} is:

$$B^{t+1} = -\frac{f(C)}{f'(C)+\varepsilon} \quad (14)$$

Where ε is an arbitrary integer, avoiding the zero-dividing operation.

Model processes and algorithms

The original SSC clustering process is improved by the above model as shown in Figure 2:

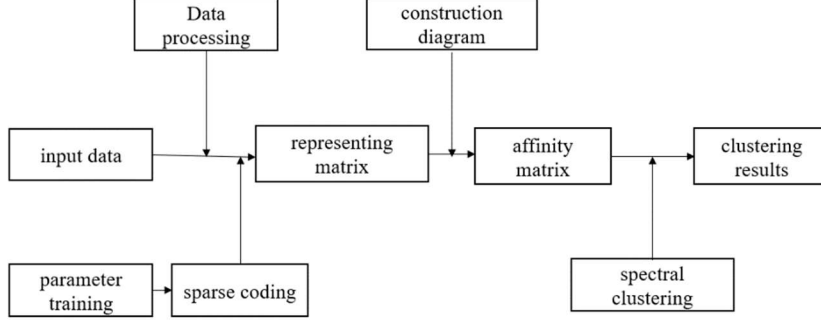


Figure 2 Flow chart of parameter training SSC model

The model training algorithm is as follows:

Algorithm 1: APSSC

Input: $D \times N \in R^n$ data matrix \mathbf{X}

Step 1 Data preprocessing and outlier deletion;

Step 2.1 Solves the optimal \mathbf{C} matrix and train the fitting coefficient;

Step 2.2 Obtains $f(\|\mathbf{C}\|_1)$, as the constraint parameter of solving \mathbf{B} matrix, solving coefficient matrix \mathbf{B} ;

Step 3 Uses Equation (2) to construct the similarity matrix \mathbf{W} ;

Step 4 Solves the corresponding Laplacian matrix for spectral clustering;

Output: Clustering results.

4 Experimental results and analysis

To evaluate the performance of our algorithm, we have conducted extensive experiments on two kinds of standard datasets: low-dimensional UCI datasets and face image datasets (Extended Yale B). We chose both the classical algorithms and the latest proposed algorithms for comparison, such as Sparse Subspace Clustering (SSC) [4], k-means algorithm, spectral clustering algorithm (SC), random sparse subspace clustering(S³COMP-C) [13], and sparse subspace clustering based on orthogonal matching pursuit (SSCOMP) [14]. For most of the algorithms mentioned above, the code is released by the authors based on the Matlab platform. For a fair comparison, all algorithms are run with the suitable parameter setting.

We use the clustering error to measure the performance of algorithms. The subspace clustering error is defined as the ratio of the misclassified data to the total number of data, with a value range of [0, 1]. The larger the error rate is, the worse the clustering effect is.

$$error\ rate = \frac{error}{n} \quad (15)$$

This paper evaluates clustering performance by accuracy, which is defined as:

$$accuracy = 1 - error\ rate \quad (16)$$

4.1 Low-dimensional data clustering

This part of the experiment was carried out on the commonly used UCI data set. The purpose is to prove the effectiveness of the proposed algorithm for low-dimensional data clustering. The proposed spectral clustering algorithm obtains the value of clustering parameter k through training optimization. Table 1 shows the clustering accuracy of the corresponding optimal k value.

As can be seen from Table 1, APSSC achieves better clustering results in all data sets. Especially for data sets of Iris, Wine, Air, APSSC is significantly better than SSC. This proves that the improved method is feasible.

Table 1 Accuracy of the improved method and original algorithm in UCI datasets

Data	SSC	APSSC
Iris(k=3)	0.51	0.83
Vote(k=2)	0.5	0.61
Wine(k=3)	0.56	0.80
WBC(k=2)	0.66	0.70
Ionosphere(k=2)	0.51	0.63
Diabetes(k=2)	0.54	0.66
Air(k=3)	0.41	0.78

The SSC algorithm has obvious advantages in high-dimensional data, it performs poorly on low-dimensional datasets. Compared with the SSC algorithm, APSSC has significantly improved clustering accuracy on low-dimensional datasets. This proves that our method improves the limitation of the SSC in the low-dimensional data clustering tasks to a certain extent.

4.2 High-dimensional data clustering

The extended Yale B datasets used in this experiment have a total of 38 people, each with 29 photos. The integrated datasets have a dimension of 32,256 dimensions, with a total of 1102 data. To reduce the complexity of the experiment, the number of categories $l \in \{2,3,4,5,6,7\}$ is used for the significance experiment by us. The results are shown in Table 2:

Table 2 Accuracy of improved algorithm and other algorithms in extended Yale-B

Method	2	3	4	5	6	7
SSC	0.9593	0.9521	0.7155	0.8345	0.8103	0.8291
K-means	0.54	0.481	0.301	0.2933	0.2318	0.2245
SSCOMP	0.929	0.8878	0.8385	0.7991	0.7659	0.7432
S ³ COMP-C	0.9593	0.937	0.918	0.899	0.8744	0.8496
SC	0.55	0.362	0.297	0.2534	0.223	0.201
APSSC	0.99	0.965	0.94	0.9124	0.881	0.8567

Compared with other algorithms, APSSC significantly improves the clustering accuracy. The experimental results also show that the adaptive parameter training method reduces the complicated parameter selection process on the one hand, and improves the clustering accuracy of the algorithm on the other hand.

5 Summary and outlook

This study focuses on improving the accuracy of the sparse subspace clustering algorithm. To obtain the coefficient matrix with good representation ability by considering both the global structure and local sparsity of the data, the sparse coding process of the training matrix is optimized. Experimental results prove that the proposed curve fitting norm matrix parameter method proposed in this paper can improve the clustering accuracy.

For the problem of model optimization, the swarm intelligence algorithm has been developed quite maturely. Combining the good optimization ability of swarm intelligence with sparse subspace clustering algorithm parameter training may be able to obtain better results. This is the direction of the author's next effort.

Acknowledgments

Research on this work was partially supported by grants from and National Nature Science Foundation of China (No.62166028).

Reference

1. WenHua Dong.: Research of sparse subspace clustering. Jiangnan University, (2019). doi: 10.27169/d.cnki.gwqgu.2019.000010
2. Agrawal R, Gehrke J, Gunopulos D, Raghavan P.: Automatic subspace clustering of high dimensional data for data mining applications. *ACM SIGMOD Record*, (1998), 27(2): 94–105. doi: 10.1145/276305.276314
3. Hong W, Wright J, Huang K, Ma Y.: Multi-scale hybrid linear models for lossy image representation. *IEEE Transactions on Image Processing*, (2006), 15(12): 3655–3671. doi: 10.1109/TIP.2006.882016
4. Elhamifar E, Vidal R, "Sparse subspace clustering," 2009 IEEE Conference on Computer Vision and Pattern Recognition, (2009), pp. 2790-2797. doi: 10.1109/CVPR.2009.5206547
5. Elhamifar E, Vidal R.: Clustering disjoint subspaces via sparse representation. In: *Proceedings of the 2010 International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. Dallas, Texas, USA: IEEE, (2010). 1926–1929. doi: 10.1109/ICASSP.2010.5495317
6. Liu G C, Yan S C.: Latent low-rank representation for subspace segmentation and feature extraction. In: *Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV)*. Barcelona, Spain: IEEE, (2011). 1615–1622. doi: 10.1109/ICCV.2011.6126422
7. Fangfang Wu, Peng Yuan, Guangming Shi, Xin Li, Weisheng Dong, Jinjian Wu.: Robust subspace clustering network with dual-domain regularization, *Pattern Recognition Letters*, Volume 149,(2021), Pages 44-50, ISSN 0167-8655. doi: 10.1016/j.patrec.2021.06.009. doi: 10.1016/j.patrec.2021.06.009
8. Li Guo, Xiaoqian Zhang, Zhigui Liu, Xuqian Xue, Qian Wang, Shijian Zheng.: Robust subspace clustering based on automatic weighted multiple kernel learning, *Information Sciences*, Volume 573,(2021), Pages 453-474, ISSN 0020-0255. doi: 10.1016/J.INS.2021.05.070
9. L. Ma, C. Wang, B. Xiao, W. Zhou, "Sparse representation for face recognition based on discriminative low-rank dictionary learning," 2012 IEEE Conference on Computer Vision and Pattern Recognition, (2012), pp. 2586-2593, doi: 10.1109/CVPR.2012.6247977
10. Tron R, Vidal R.: A benchmark for the comparison of 3-d motion segmentation algorithms. In: *Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Minneapolis, MN: IEEE, (2007). 1–8. doi: 10.1109/CVPR.2007.382974
11. Weiwei Wang, Xiaoping Li, Xiangchu Feng, Siqi Wang.: A Survey on Sparse Subspace clustering. *IEEE Journal of Automatica Sinica*, (2015), 41(8): 1373-1384. doi: 10.16383/j.aas.2015.c140891
12. Mahdi Soltanolkotabi, Ehsan Elhamifar, Emmanuel J. Candès. "Robust subspace clustering." *The Annals of Statistics*, 42(2) 669-699 April (2014). doi: 10.1214/13-AOS1199
13. Li CG, You C, Vidal R (2017) Structured sparse subspace clustering: a joint affinity learning and subspace clustering framework. *IEEE Trans Image Process* 26(6):2988–3001. doi: 10.1109/TIP.2017.2691557
14. You C, Robinson D, Vidal R (2016) Scalable sparse subspace clustering by orthogonal matching pursuit. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.3918–3927. doi: 10.1109/CVPR.2016.425