Spectral Clustering Algorithm Based on Local Sparse Representation

Sen Wu, Min Quan, and Xiaodong Feng

Dongling School of Economics and Management,
University of Science and Technology Beijing
Beijing, 100083, P.R. China
wusen@manage.ustb.edu.cn,
quanmin900112@hotmail.com,
fengxd1988@gmail.com

Abstract. Clustering based on sparse representation is an important technique in machine learning and data mining fields. However, it is time-consuming because it constructs l_1-graph by solving l_1-minimization with all other samples as dictionary for each sample. This paper is focused on improving the efficiency of clustering based on sparse representation. Specifically, the Spectral Clustering Algorithm Based on Local Sparse Representation (SCAL) is proposed. For a given sample the algorithm solves l_1-minimization with the local k nearest neighborhood as dictionary, constructs the similarity matrix by calculating sparsity induced similarity (SIS) of the sparse coefficients solution, and then uses spectral clustering with the similarity matrix to cluster the samples. Experiments using face recognition data sets ORL and Extended Yale B demonstrate that the proposed SCAL can get better clustering performance and less time consumption.

Keywords: Spectral Clustering, Weight Matrix, Sparse Representation, $k - nn$.

1 Introduction

Spectral clustering algorithm, as one kind of important clustering algorithms, searches for clusters in the full feature space, and it is equivalent to graph min-cut problem based on a graph structure constructed from the original samples in vector space. It has been increasingly and widely developed and applied in recent years, such as financial time series [1], cluster analysis of spam images[2], cancer diagnosis and treatment[3]. Compared with the traditional clustering methods, spectral clustering has the advantage to cluster the arbitrary shape of samples space and finds the global optimal solution [4].

In recent years, sparse representation, the algorithmic problem of computing sparse linear representation with respect to an over-complete dictionary of base elements, has attracted a great attention in machine learning and pattern recognition [5]. Semi-supervised learning by sparse representation [6] assumes

H. Yin et al. (Eds.): IDEAL 2013, LNCS 8206, pp. 628–635, 2013.

that each sample can be reconstructed by the sparse linear combination of other samples and constructs a l_1 graph by solving l_1 minimization problems.

The basic idea of unsupervised spectral clustering based on sparse representation (SCSR)[7] solves the sparse decomposition of each sample in the form of the linear combination of other samples; constructs the weight matrix between data samples using the sparse coefficients solution; then the weight matrix of this graph is used as the weight matrix for spectral clustering to get the cluster result. In SCSR, finding sparse representation is fast in theory if the sparsest solution is found(i.e. few non-zero coefficients)[5]. In practice, however, when the number of samples is large, the computational time cost of solving the sparse representation from all other instances is too expensive.

To overcome the time consumption deficiency of finding sparse representation, we propose to solve the sparse representation problem in a local domain to obtain an approximate solution and then to construct a local graph for spectral clustering, called spectral clustering algorithm based on local sparse representation (SCAL). Experiments conducted on real datasets demonstrated that the proposed SCAL algorithm can reduce the computational complexity.

The rest of the paper is organized as follows: in Section 2, we review an related work of spectral clustering based on sparse representation, followed by our proposed algorithm in Section 3. Section 4 demonstrates the detailed experiment results, and our work is concluded in Section 5.

2 Spectral Clustering Based on Sparse Representation

Sparse representation means to represent a sample as a linear combination of a few atoms of a given dictionary. Mathematically, given sample set X= $[x_1, x_2, ..., x_N] \in R^{M \times N}$,where N is the total number of samples and M is the number of fatures of each sample; a dictionary $D \in R^{M \times K}$, where K is the numeber of atoms in D, the sparse representation problem can be stated as :

$$\hat{\alpha} = \arg\min_{\alpha} ||\alpha||_0 \quad s.t. x_i = D\alpha \tag{1}$$

Where $||\alpha||_0$ is the l_0 pseudo-norm of the coefficient vector $\alpha \in R^K$,which counts the number of nonzero entries in a vector. However, finding the sparsest solution of (1) is NP-hard. So Wright et al. [8] proposed to solve the following l_1-minimization problem instead:

$$\hat{\alpha} = \arg\min_{\alpha} ||\alpha||_1 \quad s.t. x_i = D\alpha \tag{2}$$

Where $||\alpha||_1$ is the l_1 norm of the coefficient vector α, which is simply the sum of the absolute values of the columns. In the noisy case the equality constraint must be relaxed as well. An alternative then is to solve the unconstrained problem,

$$\hat{\alpha} = \arg\min_{\alpha} ||x_i - D\alpha||_2^2 + \lambda||\alpha||_1 \tag{3}$$

Where λ is a parameter that balances the tradeoff between reconstruction error and sparsity.

Spectral Clustering Based on Sparse Representation is depended on a basic assumption that any data sample in the data set can be well represented by the linear combination of a small number samples from the same cluster. Nevertheless, for each sample, unsupervised clustering does not have any priori information which samples are in the same cluster, To successfully identify the potential small set to reconstruct each sample,John Wright and Yi Ma etc. [9] used equation (2) to represent each sample in form of the linear combination of other samples in the dataset. For each sample x_i, $i\in$ 1,2,...,N, set X_i=X\x_i=[x_1,x_2,...,x_{i-1}, x_{i+1},...,x_N], then the reconstruction weight α_i for x_i can be calculated by solving the following l_1-minimization problem:

$$\hat{\alpha_i} = \arg\min_{\alpha_i} ||x_i - X_i\alpha_i||_2^2 \ + \lambda||\alpha_i||_1 \tag{4}$$

Theoretically, the computational complexity to obtain the sparse representation is approximately to $O(t^2$(N-1)) for each sample, where t is the number of non-zero entries in reconstruction coefficients and N-1 is the number of samples in dictionary X_i. In practice, due to there are many very small non-zero reconstruction coefficients in the obtained solution, the computation complexity to find the sparse reconstruction coefficients tend to be about $O(N^3)$ for each sample as t tends to be N.

According to the coefficient vector α, a sparse weight graph G can be constructed using different methods to characterize the relationship between samples. The commonly used weight matrix construction approaches are as follows.

Direct Construction[9]: $W_{ij} = |\alpha_j^i|$, if $i > j$, and W_{ij}=$|\alpha_{j-1}^i|$ if $i < j$,where α_j^i is the coefficient corresponding to the jth sample basis function in representation of sample x_i.

Symmetric Weight (CSR)[10]: on the basis of Direct Construction, if $\alpha_j^i \neq 0$, set the weight W_{ij}=$|\alpha_j^i|$, $1 \leq i, j \leq N$, and then W=$(W^T + W)/2$.

Sparsity Induced Similarity (SIS)[11]: The similarity between x_i and x_j, $1 \leq i, j \leq N$, $i \neq j$ is defined as

$$W_{ij} = \frac{\max\{\alpha_i^j, 0\}}{\sum\limits_{k=1,k\neq i}^{N} \max\{\alpha_i^k, 0\}}$$

This matrix is not necessarily symmetric, to ensure symmetric, the final similarity between x_i and x_j is defined as $W_{ij} = (W_{ij} + W_{ji})/2$, and set $W_{ii} = 0$.

In the concrete realization of the process, the similarity matrix W is constructed differently in different algorithms. When we get the weight matrix W, we can use the basic framework of spectral clustering[12] to cluster the dataset into K clusters.

3 The Proposed Algorithm: SCAL

In this section, we propose to solve the sparse decomposition problem in a local domain to obtain an approximate solution, this method solves the same problem

in the local neighborhood of each sample to improve the efficiency. Furthermore, in order to better reflect the relationship between samples, we use Sparsity Induced Similarity (SIS) of sparse coefficients to induce the weights of the directed l_1-graph.

Firstly, *k-nn* rules is used to find the k neighbors for each sample x_i from X_i=$X\backslash x_i$=[x_1,x_2,...,x_{i-1}, x_{i+1},...,x_N] , and defining $N\_id \in R^{N\times k}$ as the location index of the nearest k neighbors in the original data X. Notice that the k nearest neighbors are measured by the traditional Euclidean distance.

Then, l_1-norm minimization is bulit to slove the sparse representation problem in a local domain to obtain sparse coefficients for each sample x_i

$$\hat{\alpha}_i = \arg \min_{\alpha_i} ||X_{k(x_i)}\alpha_i - x_i||_2^2 + \lambda||\alpha_i||_1 \tag{5}$$

$X_{k(x_i)}$ denotes the data matrix of the k nearest neighbors of x_i. So, we can get sparse coefficients $\alpha_i \in R^k$ as the best reconstruction coefficients to represent the sample x_i. Hence the coefficient matrix is noted as A=[$\alpha_1, \ldots, \alpha_N$] $\in R^{k\times N}$.

Due to the size of coefficient matrix A is $k\times N$, it should be transformed to a $N\times N$ matrix W before calculating the similarity of two samples. The transformation function is described as follows:

$$W_{ij} = \begin{cases} A_{ri} \;\; if \; j = N\_id_{ir} \;\; i, \; r \in \{1, \ldots, k\} \\ 0 \;\; otherwise \end{cases} \tag{6}$$

Then we use Sparsity Induced Similarity (SIS) to define the similarity between x_i and x_j:

$$S_{ij} = \begin{cases} \frac{\max\{W_{ij},0\}}{\sum_{j=1}^{N} \max\{W_{ij},0\}} & if \; i \neq j \\ 0 \quad otherwise \end{cases} \tag{7}$$

Note that this matrix is not necessarily symmetric, it can be converted to be a symmetric matrix , that is: $S_{ij} = (S_{ij} + S_{ji})/2$.

After getting the weight matrix S, it can be seen as the input weight matrix of the basic framework of spectral clustering to group the sample sets X into K clusters. When the number of local neighborhood k is set N-1, our proposed method is the equal to SCSR. In manifold learning the local linearity is used to capture the local geometric structure[13] and each data point on the manifold can be locally approximated by a linear combination of its nearby points[14]. Especially the local nonnegative linear reconstructing coefficients are used to discover the natural class structure[15]. So the advocated algorithm in this paper is at least theoretically possible.

In our algorithm, the system of liner equations in (5) is over-determined. The computational complexity to obtain the sparse representation is about $O(t^2k)$ for each sample, where k is the number of the neighbors of the sample, t is the number of non-zero entries in reconstruction coefficients, and $t \leq k$. Compared with the original spectral clustering based on sparse representation, the computational cost will be saved remarkably when $k \ll N$.

4 Experiments

We conduct experiments on face recognition data sets ORL and Extended Yale B to evaluate the effectiveness of the proposed algorithm SCAL. 400 and 600 images are randomly selected from ORL data set and Extended Yale B data set respectively where each class contains the same number of images, and each image is manually cropped and normalized to the size of 32-by-32 pixels in our experiments. The cluster number (K) of two data sets are 40 and 10 respectively. Two spectral clustering algorithms based on sparse representation with all other samples as dictionary in [10](denoted as CSR) and [11] (denoted as SIS) are compared with the proposed SCAL. The difference of CSR and SIS is the different similarity construction described in the above section.

In our experiment, we use the approach and open source tool in [16] to solve the l_1-norm constraint least square minimization problem[1] in Equation(5), since it is a specialized interior-point method for solving large scale problem. The regularization parameter λ and the relative target duality gap tolerance ε for the two data sets is set with the highest clustering accuracy[17], described in Table 1.

Table 1. The regularization parameter λ and the relative target duality gap tolerance ε in 2 data sets respectively

parameters	ORL	Extended Yale B
λ	0.05	0.001
ε	1	0.005

4.1 Clustering Accuracy

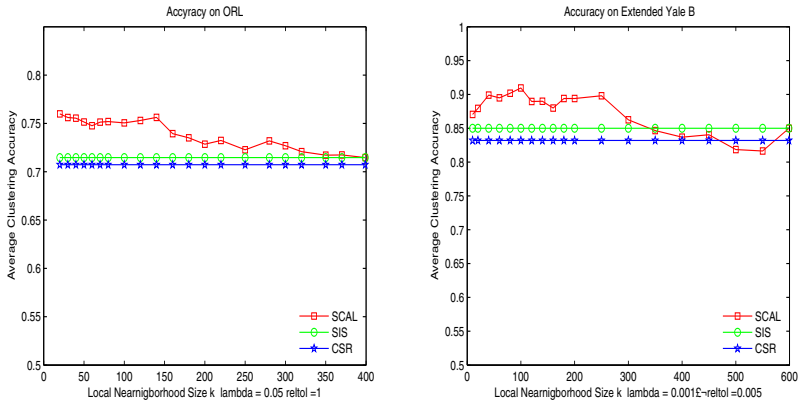
In this subsection, we adopt an external criterion which measures the degree of correspondence between the clusters obtained from our clustering algorithms and the true classes. The clustering accuracy r [17] is defined as

$$r = \frac{\sum_{i=1}^{K} a_i}{N}$$

Where a_i is the number of instances occurring in both cluster i and its corresponding class, K is the cluster number and N is the number of instances in the data set.

In ORL and Extended Yale B, all the samples are resized into 32×32 pixels. For two datasets we computed the clustering accuracy with different local neighbor number as following figures. Each step in the K-means of spectral clustering is repeated for fifty times to reduce the random influence, and the average clustering accuracy results are given in Fig.1.

[1] We use the solver for l_1 constraint minimization l_1ls.m which is downloaded from (`http://www.stanford.edu/~boyd/l1_ls/`).



(a) `ORL data set` (b) `Extended Yale B data set`

Fig. 1. Clustering accuracy on ORL data set and Extended Yale B data set

As can be seen from these figures, when the number of local neighbors is much small, the clustering accuracy of SCAL is evidently higher than other two algorithms. It's because local neighbors with appropriate number can well reflect the local geometric structure of data points. When the number of local neighbors become bigger, the clustering accuracy of SCAL approximates to that of SIS[11] using sparsity induced similarity to character the similarity of two data points. Moreover, it can get the best clustering accuracy in SCAL under small number of local neighbors. The clustering accuracy and the corresponding number of local neighbors are represented in Table 2.

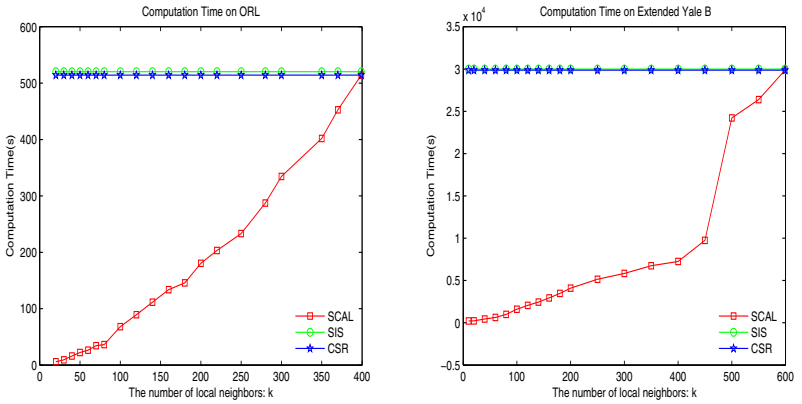
Table 2. The clustering accuracy and the corresponding number k of local neighbors

Data Sets	Symmetric Weight	SIS	Local Sparse Representation	k
ORL	0.7073	0.7147	**0.7599**	20
Extended Yale B	0.8320	0.8500	**0.9096**	100

4.2 Computation Time on Two Datasets

Considering the same condition described in Subsection 4.1, we record the computational time of SCAL and compared CSR and SIS to solve the sparse representation problem on two data sets. The computational times are presented in Fig.2.

As shown in fig.2., the smaller the number of local neighbors is, the faster it is to resolve the sparse representation. The results validate that the computation time of SCAL can approximate to the CSR and SIS when the number of local neighbors increases.



(a) `ORL data set` (b) `Extended Yale B data set`

Fig. 2. The computation time on ORL data set and Extended Yale B data set

5 Concluding Remarks

In this paper we propose spectral clustering algorithm based on local sparse representation (SCAL) to solve the problem of high computational complexity to find sparse representation in SCSR. In SCAL, the sparse representation is resolved in the local neighborhood of each sample, and the similarity matrix is constructed by calculating sparsity induced similarity of the sparse coefficients solution, using spectral clustering with this matrix as input weight matrix. Experiments on face recognition data sets ORL and Extended Yale B demonstrate that the proposed algorithm can reduce the computation complexity and the clustering accuracy is higher when the number of local neighbors is small.

Acknowledgments. This work is supported by the National Natural Science Foundation of China under Grant No.71271027, by the Fundamental Research Funds for the Central Universities of China under Grant No. FRF-TP-10-006B and by the Research Fund for the Doctoral Program of Higher Education under Grant No. 20120006110037.

References

1. Warren Liao, T.: Clustering of time series data-a survey. Pattern Recognition 38, 1857–1874 (2005)
2. Gao, Y., Choudhary, A., Hua, G.: A nonnegative sparsity induced similarity measure with application to cluster analysis of spam images. In: 2010 IEEE International Conference on Acoustics Speech and Signal Pro-cessing (ICASSP), pp. 5594–5597. IEEE (2010)
3. Yu, Z., Li, L., You, J., Wong, H., Han, G.: SC3: Triple Spectral Clustering-Based Consensus Clustering Framework for Class Discovery from Cancer Gene Expression Profiles. IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 9, 1751–1765 (2012)

4. Wang, H., Chen, J., Huang, S., Guo, K.: A Heuristic Initialization-Independent Spectral Clustering. In: 2010 Fifth International Conference on Internet Computing for Science and Engineering (ICICSE), pp. 81–84. IEEE (2010)
5. Li, C., Guo, J., Zhang, H.: Local sparse representation based classification. In: 2010 20th International Conference on Pattern Recognition (ICPR), pp. 649–652. IEEE (2010)
6. Yan, S., Wang, H.: Semi-supervised Learning by Sparse Representation. In: 2009 SIAM Conference on Data Mining, pp. 792–801 (2009)
7. Elhamifar, E., Vidal, R.: Sparse subspace clustering. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009), pp. 2790–2797. IEEE (2009)
8. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 210–227 (2009)
9. Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T.S., Yan, S.: Sparse representation for computer vision and pattern recognition. P IEEE 98, 1031–1044 (2010)
10. Jiao, J., Mo, X., Shen, C.: Image clustering via sparse representation. In: Boll, S., Tian, Q., Zhang, L., Zhang, Z., Chen, Y.-P.P. (eds.) MMM 2010. LNCS, vol. 5916, pp. 761–766. Springer, Heidelberg (2010)
11. Cheng, H., Liu, Z., Yang, J.: Sparsity induced similarity measure for label propagation. In: 2009 IEEE 12th International Conference on Image clustering via sparse representation, pp. 317–324. IEEE (2009)
12. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: Analysis and an algorithm. In: Advances in Neural Information Processing Systems, vol. 2, pp. 849–856 (2002)
13. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326 (2000)
14. Yu, K., Zhang, T., Gong, Y.: Nonlinear learning using local coordinate coding. In: Advances in Neural Information Processing Systems, pp. 2223–2231 (2009)
15. Wang, F., Zhang, C.: Label propagation through linear neighborhoods. IEEE Transactions on Knowledge and Data Engineering 20, 55–67 (2008)
16. Kim, S.J., Koh, K., Lustig, M., Boyd, S., Gorinevsky, D.: A method for large-scale l1-regularized least squares. IEEE Journal on Selected Topics in Signal Processing 4, 606–617 (2007)
17. Huang, Z.: Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Min. Knowl. Disc. 2, 283–304 (1998)