Image Clustering via Sparse Representation

Jun Jiao[1], Xuan Mo[2], and Chen Shen[3]

[1] Computer Science & Technology Department, Nanjing University, Nanjing, China
failedjj@gmail.com

[2] Department of Automation, Tsinghua University, Beijing, China
blademamama@gmail.com

[3] School of Software and Electronics, Peking University, Beijing, China
scv119@gmail.com

Abstract. In recent years, clustering techniques have become a useful tool in exploring data structures and have been employed in a broad range of applications. In this paper we derive a novel image clustering approach based on a sparse representation model, which assumes that each instance can be reconstructed by the sparse linear combination of other instances. Our method characterizes the graph adjacency structure and graph weights by sparse linear coefficients computed by solving ℓ^1-minimization. Spectral clustering algorithm using these coefficients as graph weight matrix is then used to discover the cluster structure. Experiments confirmed the effectiveness of our approach.

Keywords: Image Clustering, Spectral Clustering, Sparse Representation.

1 Introduction

Clustering algorithms are widely used in data mining and pattern recognition problems. The goal of clustering is to determine the intrinsic grouping in a set of data. Spectral clustering[1][7][13] algorithms have been successfully used in computer vision. Compared to the traditional algorithms , spectral clustering has many fundamental advantages. It is very simple to implement and can be solved efficiently by standard linear algebra methods. In addition, results obtained by spectral clustering often outperform the traditional approaches. However, the success of spectral clustering depends heavily on the choice of the similarity measure, but this choice is generally not treated as part of the learning problem. Thus, time-consuming manual feature selection and weighting is often a necessary precursor to the use of spectral methods[5].

Recently, several works have considered methods to relieve this burden by incorporating prior knowledge into the metric, either in the setting of K-means clustering[4][9] or spectral clustering[10][12]. In this paper, we consider a complementary approach to use ℓ^1 graph to construct the similarity matrix needed by spectral clustering. This method is based on the assumption that each instance can be reconstructed from the sparse linear combination of other instances. We calculate a ℓ^1 graph by solving ℓ^1 minimization problems. Weights of this graph

S. Boll et al. (Eds.): MMM 2010, LNCS 5916, pp. 761–766, 2010.

are then used as the weight matrix for spectral clustering. Experiments result shows that our spectral clustering method achieves excellent performance in image clustering.

The rest of this paper is organized as follows. Section 2 briefly introduces some related works. Section 3 presents our method to calculate the ℓ^1 graph. Section 4 reports the experiment results on synthetic data set. Section 5 concludes.

2 Related Works

2.1 Graph Construction

Given n instances $\mathbf{x}_1, ..., \mathbf{x}_N \in \mathbb{R}^d$, previous graph based algorithms usually construct a weighted graph with n nodes in the following way:(1)Constructing graph adjacency: Two nodes $\mathbf{x}_i$ and $\mathbf{x}_j$ are connected in the graph if they are considered to be close. (2)Calculating graph weight. Weights in the graph are used to reflect how strong two nodes are related.

An obvious drawback of this method is that the calculation of graph structure and weights is divided into two different steps. The structure of the graph has already been fixed after the first step. Therefore, the calculation of graph weights in the second step will be heavily constrained by this fixed graph structure.

2.2 Spectral Clustering

Spectral clustering[1][7][13] methods arise from concepts in spectral graph theory. The basic idea is to construct a weighted graph from the initial data set where each node represents a pattern and each weighted edge simply takes into account the similarity between two patterns. In this framework the clustering problem can be seen as a graph cut problem, which can be tackled by means of the spectral graph theory. The core of this theory is the eigenvalue decomposition of the Laplacian matrix of the weighted graph obtained from data. In fact, there is a close relationship between the second smallest eigenvalue of the Laplacian and the graph cut.

3 The Proposed Algorithm

Our proposed clustering algorithm is composed of three steps: (1) Constructing graph W for spectral clustering. (2) Solving the k smallest eigenvectors from W. (3) Using K-means to cluster the instances with eigenvectors as features. These three steps will be introduced in detail in the following subsections.

3.1 Constructing Graph

The method for constructing sparse graph that we adopted in this paper was firstly proposed in [11] for semi-supervised learning. We first introduce some notations: Given n instances $\mathbf{x}_1, ..., \mathbf{x}_N \in \mathbb{R}^d$ and $X = [\mathbf{x}_1, ..., \mathbf{x}_N]$ is the column matrix of all instances, our aim is to construct a sparse weighted graph G to characterize the relationship between instances. In order to obtain a sparse representation of a new instance $\mathbf{y}$, one obvious way is to solve the following ℓ^0-norm optimization problem:

$$\hat{\mathbf{a}} = \arg\min \|\mathbf{a}\|_0, s.t. X\mathbf{a} = \mathbf{y}, \tag{1}$$

where $\|\cdot\|_0$ counts the number of nonzero entries in a vector. However, the problem of optimizing equation(2) is NP-hard: there is no algorithm for solving it more efficiently than enumerating all subsets of $\mathbf{a}$, and it is difficult even to approximate as well[3][8]. To overcome this issue, Wright et al.[8] proposed to solve the following ℓ^1-minimization problem instead:

$$\hat{\mathbf{a}} = \arg\min \|\mathbf{a}\|_1, s.t. X\mathbf{a} = \mathbf{y}, \tag{2}$$

It has been proved that if the $\mathbf{a}_0$ sought is sparse enough, the solution of the ℓ^0-minimization problem is equal to the solution of equation(3), and equation(3) can be solved in polynomial time by standard linear programming methods[8].

Yan and Wang [11] used equation(3) to represent each instance in the form of the linear combination of other instances. For each instance $\mathbf{x}_i$, set $X_i = X \setminus \mathbf{x}_i = [\mathbf{x}_1, ..., \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, ..., \mathbf{x}_N]$, then the reconstruction weight $\hat{\mathbf{a}}_i$ for $\mathbf{x}_i$ can be calculated by solving the following ℓ^1-minimization problem:

$$\hat{\mathbf{a}}_i = \arg\min \|\mathbf{a}_i\|_1, s.t. X_i\mathbf{a}_i = \mathbf{x}_i, \tag{3}$$

Let $\mathbf{a}_i^j$ denote the jth entry of $\mathbf{a}_i$, the ℓ^1 graph is then determined in the following way: a direct edge is placed from node $\mathbf{x}_i$ to $\mathbf{x}_j$, iff $\mathbf{a}_i^j \neq 0$, and the weight W_{ij} is set to $|\mathbf{a}_i^j|$. Due to the fact that ℓ^1- minimization automatically leads to a sparse representation, the graph obtained here is a sparse graph. It is also important to note that the graph obtained is a directed graph. In order to make it symmetric for spectral clustering, we transform the weight matrix W into $(W^T + W)/2$ and use it as the weight matrix for clustering.

3.2 Spectral Clustering

After calculating the sparse representation for each instance, we use the classical spectral clustering algorithm[1] with weight matrix W to discover the cluster structure. We first calculate the normalized matrix $L = D - W$, where D is a diagonal matrix with $D_{ii} = \sum_{j=1}^{n} W_{ij}$. Then the eigen-vectors of L is solved to obtain the first k eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_k$. Each row of eigenvectors is then normalized and regarded as a representation of the corresponding instance. K-means clustering algorithm is finally employed to cluster the rows of eigenvectors.

The main advantages of our algorithm are as follow: (1) Compared with many ℓ^2 based clustering algorithm, the ℓ^1 minimization employed in our algorithm can lead to a sparse representation. (2) Our clustering method can determine both the graph adjacency and weight in ℓ^1 optimization, while most of the

Algorithm 1. Image Clustering via Sparse Representation

Input:
X: A set of real valued instances $X = [\mathbf{x}_1, ..., \mathbf{x}_N], \mathbf{x}_i \in \mathbf{R}^d$.
Parameter:
k: The number of clusters
Process:

1. Normalize the instances to have unit ℓ^2 norm.
2. For each instance $\mathbf{x}_i$, solve equation(4) to obtain $\mathbf{a}_i$
3. If $\mathbf{a}_i^j \neq 0$, set the weight $W_{ij} = |\mathbf{a}_i^j|$, $1 \leq i, j \leq N$
4. $W = (W^T + W)/2$
5. $D_i \leftarrow \sum_{j=1}^N W_{ij}, \quad D \leftarrow diag\{D_i\}_i$
6. $L = D - W$
7. Compute the first k eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_k$ of L, then form these eigenvectors into $V = [\mathbf{v}_1, ..., \mathbf{v}_k] \in \mathbb{R}^{n \times k}$
8. Normalize each row of V: $V_{ij} = V_{ij}/(\sum_k v_{ik}^2)^{\frac{1}{2}}$
9. Each row of V is considered as an instance and cluster these instances into k clusters with K-means clustering algorithm.

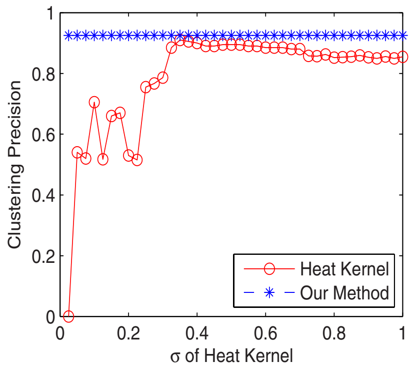
previous clustering algorithms separate them into two steps. (3) While many other clustering algorithms are very sensitive to the parameters, our algorithm is parameter free. The performance of traditional spectral clustering is heavily related to the choice of σ in Heat Kernel.

4 Experiment

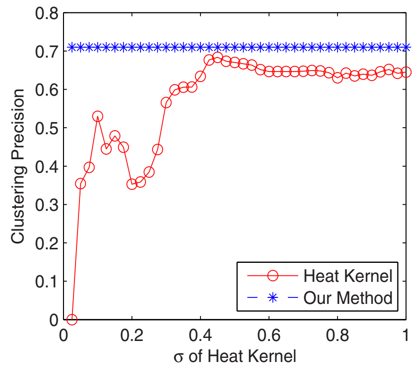
In this section, we present the results of applying our algorithm on different kinds of image data set. To validate that our algorithm can achieve better clustering performance and is very stable, we compare our algorithm with the classical clustering algorithm proposed in [1], which adopted Heat Kernel as the similarity measure.

We tested our algorithm on the famous COREL image data set. 6 classes of images are picked out from Corel image data set, where each image class is composed of 100 images. Among these 6 classes, we chose 2, 3, 4, 5 and 6 classes seperately to form six data sets with different number of clusters. Each image is represented in the combination of 64 dimensional color histogram and 64 dimensional Color Texture Moment [6]. Clustering Accuracy is used to measure the the clustering performance[2]. Various σ for Hear Kernel are used to carry out the experiment, and the accuracy is calculated by repeating the experiment 100 times, then averaging the accuracy. The results are presented in Figure 1.

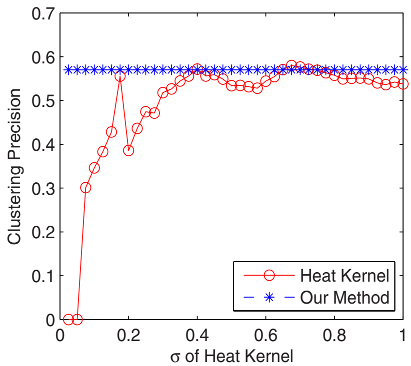
We can see from Figure 1 that our method, which is parameter free , performs better than the spectral clustering method using Gaussian function, which is sensitive to the parameter σ.



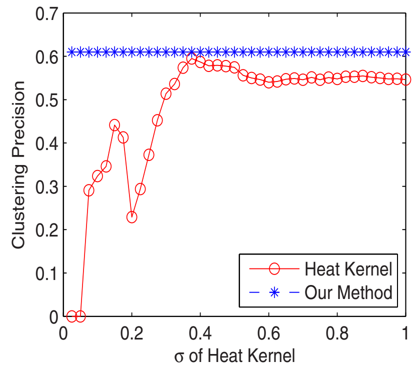
(a) 2 clusters



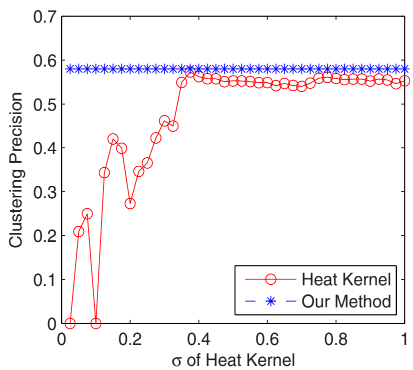
(b) 3 clusters



(c) 4 clusters



(d) 5 clusters



(e) 6 clusters

Fig. 1. Results on Corel data set

5 Conclusion

In this paper, we proposed a novel image clustering approach based on a sparse representation model. Similarity matrix are obtained from neighborhoods information by solving ℓ^1-minimization programming problems, then spectral clustering algorithm is applied to the similarity matrix obtained. Our method is parameter free and can lead to a sparse representation of similarity matrix. In the experiments, we show that our algorithm can be used to cluster images effectively.

References

1. Ng, A., Jordan, M., Weiss, Y.: On spectral clustering: analysis and an algorithm. In: Advances in Neural Information Processing Systems, vol. 14, pp. 849–856 (2002)
2. Cai, D., He, X., Han, J.: Document clustering using locality preserving indexing. IEEE Transactions on Knowledge and Data Engineering 17, 1624–1637 (2005)
3. Amaldi, E., Kann, V.: On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. Theoretical Computer Science 209, 237–260 (1998)
4. Xing, E.P., Ng, A.Y., Jordan, M.I., Russell, S.: Distance metric learning, with application to clustering with side-information. In: Proceedings of Neural Information Processing Systems, vol. 15, pp. 505–512 (2002)
5. Bach, F.R., Jordan, M.I.: Learning Spectral Clustering, With Application To Speech Separation. Journal of Machine Learning Research 7, 1963–2001 (2006)
6. Yu, H., Li, M.J., Zhang, H.J., Feng, J.F.: Color texture moments for content-based image retrieval. In: Proc. International Conference on Image Processing, vol. 3, pp. 929–932 (2002)
7. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 888–905 (2000)
8. Malik, J., Ganesh, A., Yang, A., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 210–227 (2008)
9. Wagstaff, K., Cardie, C., Rogers, S., Schroedl, S.: Constrained K-means clustering with background knowledge. In: Proc. the Eighteenth International Conference on Machine Learning, pp. 577–584 (2001)
10. Kamvar, S.D., Klein, D., Manning, C.D.: Spectral learning. In: Proc. the Eighteenth International Joint Conference on Artificial Intelligence, pp. 561–566 (2003)
11. Yan, S.C., Wang, H.: Semi-supervised Learning by Sparse Representation. In: Proc. SIAM Data Mining Conference, pp. 792–801 (2009)
12. Yu, S.X., Shi, J.: Grouping with bias. In: Advances in Neural Information Processing Systems (2002)
13. Weiss, Y.: Segmentation using eigenvectors: a unifying view. In: Proc. of the Seventh International Conference on Computer Vision, vol. 2, p. 975 (1999)