Robust Subspace Clustering by Learning an Optimal Structured Bipartite Graph via Low-rank Representation

Wei Chang, Feiping Nie, Rong Wang and Xuelong Li, *Fellow, IEEE*

hsomewei@gmail.com, feipingnie@gmail.com, wangrong07@tsinghua.org.cn, xuelong_li@ieee.org

Problem Formulation

Subspace clustering plays a very important role in clustering problem. At present, graph based methods have a good development in solving the problem of subspace clustering. The low-rank representation based method (LRR) proposed by Liu et al. [2] is one of the classical method, which can be described as

$$\min_{Z} \|Z\|_*, \ \ s.t. \ X = XZ. \tag{1}$$

Problem (1) presents the low-rank representation based model. In general, after obtaining the optimal solution Z, the graph is constructed by $(|Z^T| + |Z|)/2$ and spectral clustering is utilized as postprocessing on this graph. In this paper, we want to learn an optimal structured graph to avoid this postprocessing.

Contributions

We proposed a novel low-rank representation based method LOSBG via the optimal bipartite graph, and main contributions of this paper are summarized as follows.

- Different with classical subspace clustering methods which need spectral clustering as postprocessing on the constructed graph to get the final result, our method can directly learn a structural graph with k connected components so that different clusters can be obtained easily.
- We introduce a regularization term of error matrix to our model which makes the proposed algorithm more effective to learn an optimal graph under the circumstances of various noise.
- An efficient algorithm is designed to achieve the subspace clustering method, and extensive experiments are conducted to verify the effectiveness and superiority of our model.

LOSBG

In this work, based on the idea of co-clustering [3], we want to learn an optimal bipartite graph with k connected components which can avoid the postprocessing. Combining Theorem 1 with theoretical derivation, the final optimization problem can be described as

$$\min_{Z,E,S,F} \|Z\|_* + \lambda_1 \|E\|_{2,1} + \lambda_2 \|S - Z\|_F^2 + \lambda_3 tr(F^T \tilde{L}_G F)$$
$$s.t. \quad X = XZ + E, S \geq 0, S'\mathbf{1} = \mathbf{1}, F^T F = I, F \in R^{N\times k}. \tag{2}$$

Here, $\mathbf{1} = (1, 1, ..., 1)^T$, $\tilde{L}_G$ is the normalized Laplacian matrix of graph G. For problem (2), we introduce a regularization term of error matrix, which makes our model robust to noise. Combining with the idea of co-clustering, the bipartite graph G is constructed by the learned matrix S as follows:

$$G = \begin{bmatrix} 0 & S \\ S^T & 0 \end{bmatrix}. \tag{3}$$

Optimization

For the objective function (2), there are four variables needed to be updated. When fixing the variables S and F, problem (2) can be further transformed into the following problem

$$\min_{Z,E,J} \|J\|_* + \lambda_1 \|E\|_{2,1} + \lambda_2 \|S - Z\|_F^2$$
$$s.t. \quad X = XZ + E, Z = J. \tag{4}$$

Hence, we can utilize the Augmented Lagrange Multiplier problem of problem (4) to update the variables Z, E, J.

Updating the matrix J, problem (4) becomes

$$\arg\min_{J} \frac{1}{\mu}\|J\|_* + \frac{1}{2}\|J - (Z + \frac{1}{\mu}Y_2)\|_F^2. \tag{5}$$

The reference [1] has proved that problem (5) has an analytical solution.

Updating the matrix Z, we can get the closed form

$$Z = [(1 - \frac{2\lambda_2}{\mu})I + X^T X]^{-1}[X^T X - \frac{2\lambda_2}{\mu}S - X^T E + J + \frac{1}{\mu}(X^T Y_1 - Y_2)]. \tag{6}$$

Updating the error matrix E, we have

$$\arg\min_{E} \frac{\lambda_1}{\mu}\|E\|_{2,1} + \frac{1}{2}\|E - (X - XZ + \frac{1}{\mu}Y_1)\|_F^2. \tag{7}$$

Lin et al. [2] have given the closed-form solution for this problem in Lemma 3.2.

When fixing the variables Z and E, problem (2) is equivalent to the following problem

$$\min_{S,F} \|S - Z\|_F^2 + \lambda tr(F^T \tilde{L}_G F)$$
$$s.t. \quad S \geq 0, S'\mathbf{1} = \mathbf{1}, F^T F = I, F \in R^{N\times k}, \tag{8}$$

here, $\lambda = \lambda_3/\lambda_2$. Nie et al. [3] present an iterative algorithm which can solve the problem (8) effectively. Due to the limitation of space, we omit this optimization process which can be seen in our paper.

Theoretical Support

Theorem 1 ([3]). *The multiplicity k of the eigenvalue 0 of the normalized Laplacian matrix $\tilde{L}_G$ is equal to the number of connected components in the graph associated with G.*

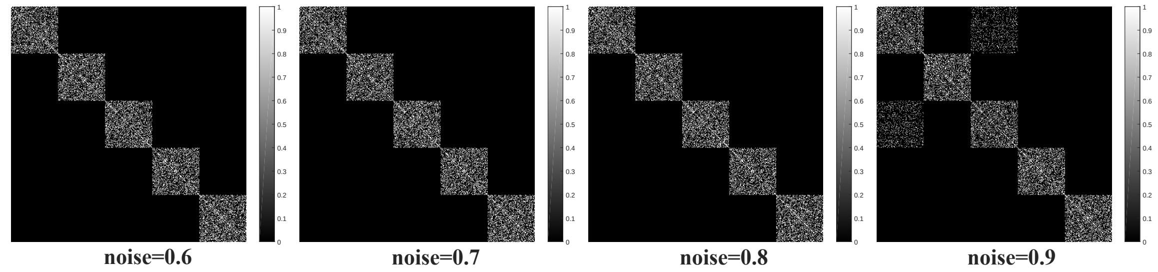
In reference [1], it has proved that the following problem has an closed form solution.(Here, USV^T is the SVD of W.)

$$U\mathcal{S}_\epsilon[S]V^T = \arg\min_{X} \epsilon\|X\|_* + \frac{1}{2}\|X - W\|_F^2, \ where \quad \mathcal{S}_\varepsilon[x] = \begin{cases} x - \varepsilon, & if\, x > \varepsilon, \\ x + \varepsilon, & if\, x < -\varepsilon, \\ 0, & otherwise. \end{cases} \tag{9}$$

Lemma 1 (the Lemma 3.2 in reference [2]). *Let $Q = [q_1, q_2, ..., q_i, ...]$ be a given matrix and $\|.\|_F$ be the Frobenius norm. The following problem has an optimal solution W^*.(Here, $W^*(:,i)$ represents the i-th column of W^*.)*

$$W^* = \arg\min_{W} \lambda\|W\|_{2,1} + \frac{1}{2}\|W - Q\|_F^2, \ where \quad W^*(:,i) = \begin{cases} \frac{\|q_i\|_2 - \lambda}{\|q_i\|_2} q_i, & if\, \lambda < \|q_i\|_2, \\ 0, & otherwise. \end{cases} \tag{10}$$

Visualization Experiment Results



We apply LOSBG to a high-dimensional synthetic dataset as a sanity check, which contains five 50-dimensional subspaces. In order to verify the robustness of LOSBG, we add Gaussian noise to this dataset and set the proportion of noise to be $r = 0.6, 0.7, 0.8, 0.9$ respectively. The above figures show the learned structured graph S by LOSBG under different levels of noise. The clustering accuracies are 100%, 100%, 100% and 79.80% respectively from left to right.

Algorithm Description

Input: data matrix X, the cluster number k.
Initialize: Randomly initialize the matrix S to satisfy the constraint condition in problem (2).
while *not converge* **do**

1. Fix others, update J by solving problem (5).
2. Fix others, update Z by formula (6).
3. Fix others, update E by solving problem (7),
4. Fix others, update S and F, the matrices S and F can by obtained effectively by optimize the problem (8) with an iterative algorithm proposed by Nie et al. [3].
5. Update multipliers Y_1, Y_2 and parameter μ.

Output: the learned bipartite graph G and the cluster label.

Reference

[1] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.

[2] Guangcan Liu, Zhouchen Lin, and Yong Yu. Robust subspace segmentation by low-rank representation. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 663–670, 2010.

[3] Feiping Nie, Xiaoqian Wang, Cheng Deng, and Heng Huang. Learning a structured optimal bipartite graph for co-clustering. In *Advances in Neural Information Processing Systems*, pages 4129–4138, 2017.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China grant under number 61772427 and 61751202.