Robust Subspace Clustering by Learning an Optimal Structured Bipartite Graph via Low-rank Representation
Weiqiang Chang, Feiping Nie, Rong Wang and Xuelong Li, Fellow, IEEE

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 method, some 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 method which makes the proposed model 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.

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:

$$\min_{E,Z,J} \|Z\|_F + \lambda_1 \|E\|_{2,1} + \lambda_2 \|S - Z\|_F^2$$

s.t. $X = ZX + E, Z = J$.

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

$$\min_{J} \|J\|_{2,1} + \frac{1}{2}\|J - (Z + \epsilon I)\|_F^2.$$

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

The updating matrix $Z$ can be obtained from the closed form

$$Z = \left(1 - \frac{2\lambda_2}{\rho}I + X^T X\right)^{-1}X^T X - \frac{2\lambda_2}{\rho}S - X^T E$$

$$+ J + \frac{1}{\rho} (X^T Y - Y^T Z).$$

Updating the error matrix $E$, we have

$$\min_{E} \frac{1}{2}\|E\|_{2,1} + \frac{1}{2}\|E - (X - XZ + \epsilon I)\|_F^2,$$

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:

$$\min_{S,F} \|S - Z\|_F^2 + \lambda \text{str}(F^T L_0 F)$$

s.t. $S \geq 0, S^T = 1, F^T F = I, F \in R^{N \times k}$.

LoseBG
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 2 with theoretical derivation, the final optimization problem can be described as

$$\min_{E,Z,F} \|Z\|_F + \lambda_1 \|E\|_{2,1} + \lambda_2 \|S - Z\|_F^2 + \lambda \text{str}(F^T L_0 F)$$

s.t. $X = ZX + E, S \geq 0, S^T = 1, F^T F = I, F \in R^{N \times k}$.

Here, $I = (1,1,1)^T$, $L_0$ 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:

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

Theoretical Support
Theorem 1 ([3]). The multiplicity $k$ of the eigenvalue 0 of the normalized Laplacian matrix $L_0$ 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 a closed form solution. Here, $USV^T$ is the SVD of $W$:

$$US^T[S]V^T = \arg \min_{X}s.t. \|X\|_F + \frac{1}{2}\|X - W\|_F^2$$

where $S_\epsilon = \begin{cases} x - \epsilon, & \text{if } x \geq \epsilon, \\ x + \epsilon, & \text{if } x < \epsilon, \\ 0, & \text{otherwise.} \end{cases}$

Lemma 1 (the Lemma 3.2 in reference [2]). Let $Q = [q_1, q_2, ..., q_l]$ be a given matrix and let $S$ 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 $W^*(:, i) = \begin{cases} \frac{q_i}{\sqrt{\sum q_i^2}}, & \text{if } |q_i| > \epsilon, \\ 0, & \text{otherwise.} \end{cases}$

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 $\rho = 0.6, 0.7, 0.8, 0.9$ 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 be 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 $\mu$.
Output: The learned bipartite graph $G$ and the cluster label.

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

Reference