Orthogonal and Nonnegative Graph Reconstruction for Large Scale Clustering

Junwei Han, Kai Xiong, Feiping Nie

Northwestern Ploytechnical University, Xi'an, 710072, P. R. China

Motivations & Contributions

◆ Two Issues of Spectral Clustering

- Scalability Issue: Spectral clustering suffers from high computational cost. It takes $O(n^3)$ for eigen-decomposition with n denoting the number of data points.
- Post-processing: Spectral clustering relys on post-processing. Kmeans is a common way to obtain the final cluster labels, while kmeans itself is sensitive to initialization.

◆ Two Major Ways for Solving The Scalability Issue

- Reduce the cost of eigen-decomposition. The Nyström method is a popular technique for finding an approximate solution.
- Reduce the data size by sampling some representative points beforehand.

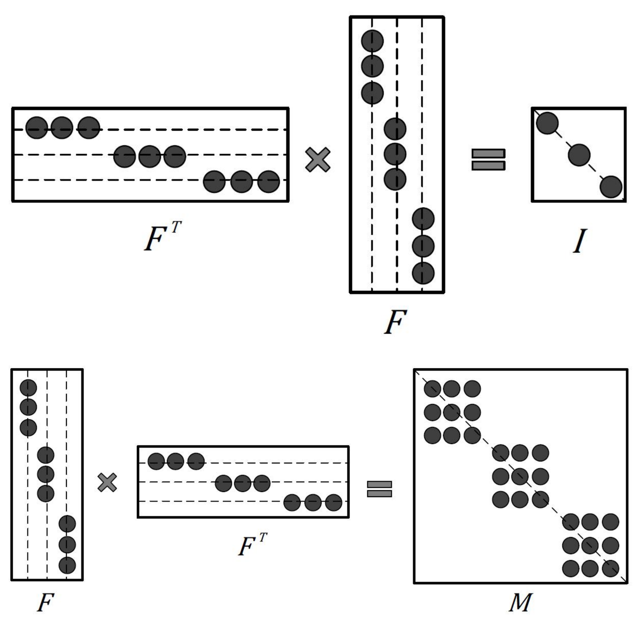
◆ Our Contributions

- The proposed method scales linearly with the data size, handling the scalability issue from the viewpoint of graph reconstruction.
- No post-processing. The interpretability is offered to obtain the cluster labels directly.
- Due to the orthogonal and nonnegative constraints, the reconstructed graph naturally has clear structure about the clusters.

Formulation & Illustration

- The relaxed problem of Ncut.
  $$\min_{F^TF=I_k} Tr(F^TLF)$$
- Add a non-negative constraint to get discrete indicator matrix.
  $$\min_{F\geq 0,\, F^TF=I_k} Tr(F^TLF)$$
- The viewpoint of graph reconstruction. Here W is doubly-stochastic.
  $$\min_{F\geq 0,\, F^TF=I_k} \left\| W - FF^T \right\|_F^2$$
- Introduce a slack variable G, which is called label matrix.
  $$\min_{G\geq 0,\, F^TF=I_k} \left\| W - FG^T \right\|_F^2 + \lambda \left\| F - G \right\|_F^2$$

- Under the two constraints, F has only one non-zero entry in each row, and the L_2-norm of each column is 1.
- The nonnegativity offers interpretability of F, and the reconstructed graph FF^T is naturally structured.
- In a sense, the interpretability of F is passed on to G.



Optimization

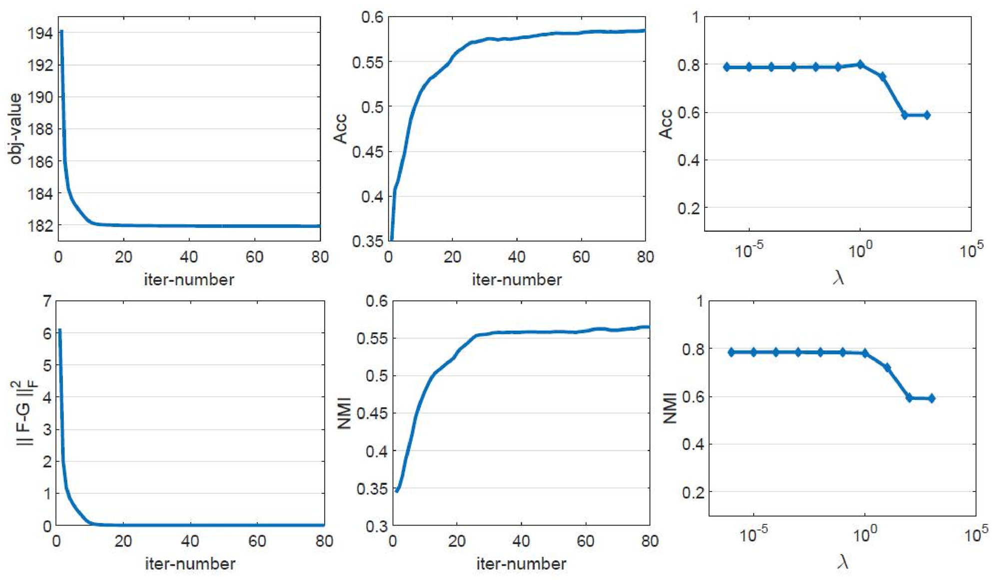
Update G with F fixed ⇒ $\min_{G\geq 0} \left\| G - \frac{WF + \lambda F}{1+\lambda} \right\|_F^2$ ⇒ $G = \left(\frac{WF + \lambda F}{1+\lambda} \right)_+$, $(x)_+ = \max(x, 0)$

Update F with G fixed ⇒ $\max_{F^TF=I_k} Tr\left(F^T(U\Lambda V^T)\right)$ ⇒ $F = U[I_k; 0]V^T$, $\mathrm{svd}(WG + \lambda G) = U\Lambda V^T$

- Graph Construction: $W = ZZ^T$, where Z is of size $n \times m$ ($m << n$) [1]. m is the number of anchors, which can be selected randomly (ONGR-R) or by kmeans (ONGR-K).
- Speed Up: $WF \to Z(Z^TF)$, $WG \to Z(Z^TG)$.
- Initialization: F can be initialized by the corresponding singular vectors of the sparse regression matrix Z.

Experimental Results

- Comparison Methods: (1) Nyström [2], (2) KNN-SC [3], (3) LSC-R, LSC-K [4], (4) SSSC [5].



(a) COIL-20 (b) MINIST-extend (c) USPS

Figure 1. Curves of objective, norm of (F-G), performance and parameter effect

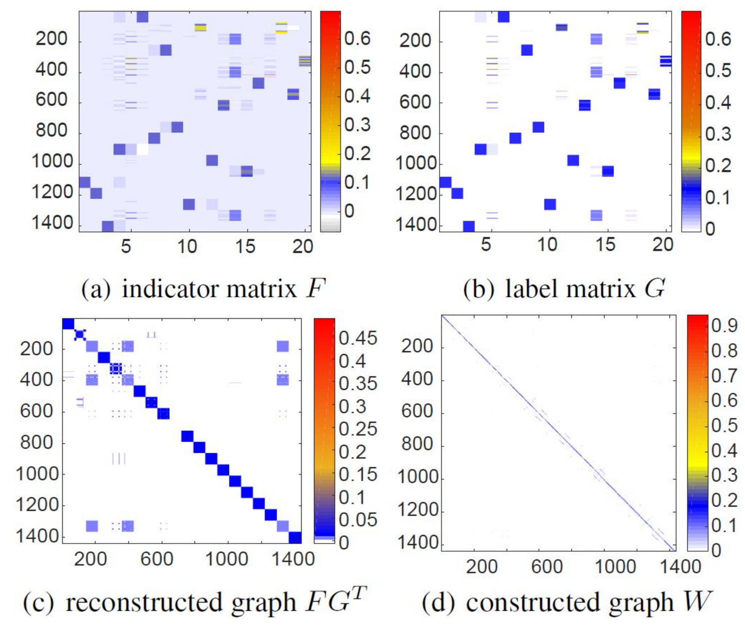


Figure 2. Practical illustrations

Table 1 Running time (s)

Dataset	KNN-SC	Nyström	SSSC	LSC-R	ONGR-R
USPS	9.65	5.84	89.76	2.19	**1.40**
PenDigits	28.04	33.03	110.98	21.67	**19.29**
MINIST	1401.41	40.88	217.68	**31.95**	39.40
CoverType	–	**168.55**	463.22	235.61	202.46
MINIST-extend	–	178.24	1095.78	166.55	**147.41**

Table 2 Clustering Performance

Metric	Dataset	KNN-SC	Nyström	SSSC	LSC-R	LSC-K	ONGR-R	ONGR-K
Acc	USPS	66.84 ± 3.05	69.52 ± 2.13	53.85 ± 0.69	75.67 ± 5.05	77.00 ± 7.20	78.82	**80.59**
	PenDigits	64.15 ± 0.15	72.33 ± 2.49	74.92 ± 0.00	79.16 ± 3.21	79.97 ± 6.11	87.30	**88.02**
	MINIST	68.72 ± 0.03	55.38 ± 3.13	53.01 ± 0.35	69.82 ± 5.23	76.21 ± 6.20	70.75	**78.59**
	COIL-20	82.22 ± 0.00	63.50 ± 3.00	61.38 ± 1.74	71.19 ± 4.79	72.89 ± 6.66	87.08	**87.92**
	COIL-100	59.81 ± 0.49	46.66 ± 1.54	43.94 ± 1.29	51.60 ± 1.59	57.45 ± 2.59	54.60	**67.13**
	Connect-4	42.68 ± 0.15	36.43 ± 0.05	**65.82 ± 0.00**	40.79 ± 2.80	40.03 ± 2.82	55.57	52.61
	Seismic	67.69 ± 0.01	67.21 ± 0.00	66.52 ± 0.00	67.58 ± 0.44	67.81 ± 0.12	**68.54**	68.42
	RCV1	–	16.94 ± 0.72	14.22 ± 0.00	16.47 ± 0.38	–	**17.49**	–
	CoverType	–	27.00 ± 1.06	44.06 ± 0.00	41.87 ± 2.01	–	**53.30**	–
	MINIST-extend	–	47.25 ± 2.47	55.74 ± 0.00	58.72 ± 5.09	–	**59.26**	–
	mean	(64.59 ± 0.55)	50.22 ± 1.66	53.35 ± 0.41	57.29 ± 3.06	(67.34 ± 4.53)	**63.27**	**(74.75)**
NMI	USPS	80.45 ± 1.31	65.19 ± 0.93	55.93 ± 0.56	77.48 ± 2.86	80.64 ± 2.34	78.48	**82.76**
	PenDigits	78.93 ± 1.27	66.65 ± 1.09	73.51 ± 0.00	79.84 ± 2.26	81.85 ± 2.74	83.50	**84.42**
	MINIST	76.60 ± 0.07	48.04 ± 1.27	53.55 ± 0.11	66.73 ± 2.29	77.33 ± 2.36	69.05	**79.50**
	COIL-20	91.15 ± 0.00	76.50 ± 1.39	78.09 ± 1.15	90.31 ± 2.89	90.90 ± 2.37	95.18	**96.35**
	COIL-100	83.80 ± 0.17	76.15 ± 0.58	69.11 ± 0.48	77.29 ± 0.53	82.96 ± 0.67	79.27	**88.16**
	Connect-4	0.18 ± 0.00	0.24 ± 0.01	0.24 ± 0.00	0.25 ± 0.09	0.22 ± 0.10	**0.58**	0.32
	Seismic	27.60 ± 0.02	27.52 ± 0.01	25.12 ± 0.00	29.85 ± 0.45	29.93 ± 0.83	31.90	**32.20**
	RCV1	–	**25.81 ± 0.27**	17.85 ± 0.00	23.65 ± 0.21	–	24.19	–
	CoverType	–	13.94 ± 0.00	20.58 ± 0.00	19.56 ± 0.84	–	**21.05**	–
	MINIST-extend	–	36.22 ± 0.88	54.75 ± 0.00	55.51 ± 1.61	–	**56.39**	–
	mean	(62.67 ± 0.41)	43.63 ± 0.64	44.87 ± 0.23	52.05 ± 1.40	(63.40 ± 1.63)	**53.96**	**(66.24)**

Conclusion & Outlook

We have proposed an approach for large scale clustering based on graph reconstruction. The reconstructed graph is structured, and the interpretability is provided to get rid of the post-processing.

- Since the noise and outliers are always there in real applications, a robust version is needed to better deal with the case.
- The original graph should better be structured. Hence, a structured and doubly-stochastic W needs to be designed efficiently.
- The value range of the original graph and the reconstructed graph may differ a lot. We can introduce a scale fator to fit them more properly.
- ONGR has close relationship with NMF. It is of great value to explore their underlying connections and develope fast NMF methods.

References & Acknowledgments

[1] Wei Liu, Junfeng He, and Shih-Fu Chang. Large graph construction for scalable semi-supervised learning, ICML, 2010.

[2] Charless Fowlkes, Serge Belongie, Fan Chung, and Jitendra Malik. Spectral grouping using the nystrom method, TPAMI, 2004.

[3] Wen-Yen Chen, Yangqiu Song, Hongjie Bai, Chih-Jen Lin, and Edward Y Chang. Parallel spectral clustering in distributed systems, TPAMI, 2011.

[4] Xi Peng, Lei Zhang, and Zhang Yi. Scalable sparse subspace clustering, CVPR, 2013.

[5] Deng Cai and Xinlei Chen. Large scale spectral clustering via landmark-based sparse representation, IEEE Transactions on Cybernetics, 2014.

◆**Acknowledgements:** This work was supported in part by the National Science Foundation of China under Grants 61522207 and 61473231.

Poster Presenter: Kai Xiong (bearkai1992@gmail.com). Research Interest: ML.

I am supposed to get my master's degree in March next year, and I'm looking for a job now.