Boosting Spectral Clustering on Incomplete Data via Kernel Correction and Affinity Learning

Fangchen Yu[1], **Runze Zhao**[1], Zhan Shi[1], Yiwen Lu[1]
Jicong Fan[1,2], Yicheng Zeng[2], Jianfeng Mao[1,2], Wenye Li[1,2(✉)]

[1]The Chinese University of Hong Kong, Shenzhen, China
[2]Shenzhen Research Institute of Big Data, Shenzhen, China

37th Conference on Neural Information Processing Systems

November 19, 2023

1. Introduction: Spectral Clustering on Incomplete Data

- **Spectral Clustering**'s performance highly relies on the quality of the similarity matrix, such as the kernel matrix and affinity matrix.

  Step 1. Construct A Similarity Matrix: $S = K$ or C
  - Calculate a (Gaussian) kernel $K \in \mathbb{R}^{n \times n}$ from $X \in \mathbb{R}^{d \times n}$.
  - Learn a self-expressive affinity matrix $C \in \mathbb{R}^{n \times n}$ from K.

  Step 2. Perform Normalized Cut [1]

Research Question is how to estimate a high-quality kernel or affinity matrix for incomplete data, benefiting spectral clustering tasks.

2. Related Work

Traditional methods to deal with incomplete data:

- **Data Imputation**: to impute missing values with estimated values.
  - Statistical imputation: zero, mean, kNN, regression, ...
  - Matrix completion: e.g., $\min \|UV^\top - X\|$
- **Distance Calibration**: to calibrate a non-metric distance to a metric.
- **Limitation**: no guarantee on the quality of the kernel matrix.

Research Goal is to propose new imputation-free algorithms based on properties of kernel matrices with a theoretical guarantee.

3.1 Kernel Correction Algorithm

Motivation:

- A valid kernel is a symmetric matrix being *positive semi-definite* [2].

Formulation:

- Estimation: estimate a naive kernel K^0 on incomplete data [3].
- Optimization: $\min_{K \in \mathbb{R}^{n\times n}} \|K - K^0\|_F^2 \; s.t. \; K \succeq 0, \; k_{ij} = k_{ji} \in [l, u], \; \forall \; i, j$
- Obtain the solution $\hat{K}$ by Dykstra's projection algorithm [4].

Guarantee:

- $\|K^* - \hat{K}\|_F \leq \|K^* - K^0\|_F$, where K^* is the unknown ground-truth.

3.2 Kernel Self-expressive Learning Algorithms

- **K**ernel **S**elf-expressive **L**earning with **S**chatten **p**-norm (**KSL-Sp**)
  - Schatten p-norm: $\|C\|_{S_p} := (\sum_{i=1}^{n} \sigma_i^p(C))^{1/p}$ with flexible sparsity.

$$\boxed{\min_{C \in \mathbb{R}^{n \times n}} \|\phi(X) - \phi(X)C\|_F^2 + \lambda \|C\|_{S_p}, \ s.t. \ 0 \leq c_{ij} \leq 1, \ \forall 1 \leq i, j \leq n.}$$

- **A**daptive **K**ernel **L**east-**S**quares **R**epresentation (**AKLSR**)
  - Learn kernel and affinity matrices iteratively via KLSR and KC:

$$\boxed{\min_{K \succeq 0, \ C} \|K - K^o\|_F^2 + \mathrm{Tr}(K - 2KC + C^\top KC) + \lambda \|C\|_F^2.}$$

  - Solve it by ADMM [5] based on the augmented Lagrange function.

4. Performance on Spectral Clustering

- Dataset: Yale64 images (missing completely at random, 80% missing).
- Relative error (**RE** $= \frac{\|\hat{K}-K^*\|_F}{\|K^*\|_F}$): measures the accuracy of estimation.
- Performance on **Kernel Estimation** & **Standard Spectral Clustering**:
  Our KC method achieves smallest estimation error and highest NMI performance.

Metric	Naive	ZERO	MEAN	kNN	EM	SVT	GR	KFMC	DC	TRF	EE	KC
RE↓	0.113	0.382	0.195	0.381	0.195	0.380	0.376	0.335	0.180	0.112	0.097	**0.089**
NMI↑	0.588	0.246	0.429	0.257	0.428	0.269	0.264	0.297	0.551	0.587	0.593	**0.596**

- Performance on **Self-expressive Affinity Learning**: Both employ corrected kernels to yield dependable affinity matrices, elevating SC performance (NMI values).

Method	Naive	ZERO	MEAN	kNN	EM	SVT	GR	KFMC	DC	TRF	EE	KC
KSSC	0.219	0.215	0.167	0.173	0.177	0.218	0.208	0.259	0.588	0.210	0.209	**0.616**
KLSR	0.606	0.311	0.604	0.320	0.609	0.321	0.327	0.318	0.597	0.603	0.604	**0.616**
KSL-Sp	0.370	0.315	0.581	0.303	0.579	0.305	0.304	0.295	0.555	0.364	0.599	**0.619**
AKLSR	0.452	0.327	0.606	0.338	0.605	0.308	0.338	0.312	0.570	0.464	0.575	**0.614**

5. Conclusion

How to boost spectral clustering on incomplete data?

- Proposed an imputation-free framework:
- Learned a high-quality *kernel* matrix by the kernel correction algorithm.
- Learned a high-quality *affinity* matrix by the kernel self-expressive affinity learning algorithms.
- Experiments show the effectiveness of kernel correction method, compared to existing data imputation and distance calibration approaches.

References

[1] Andrew Ng, Michael Jordan, and Yair Weiss. "On spectral clustering: Analysis and an algorithm". In: *Advances in Neural Information Processing Systems* 14 (2001).

[2] Rafic Nader et al. "On the positive semi-definite property of similarity matrices". In: *Theoretical Computer Science* 755 (2019), pp. 13–28.

[3] Fangchen Yu et al. "Online estimation of similarity matrices with incomplete data". In: *Uncertainty in Artificial Intelligence*. PMLR. 2023, pp. 2454–2464.

[4] Heinz H Bauschke and Jonathan M Borwein. "Dykstra's alternating projection algorithm for two sets". In: *Journal of Approximation Theory* 79.3 (1994), pp. 418–443.

[5] Stephen Boyd et al. "Distributed optimization and statistical learning via the alternating direction method of multipliers". In: *Foundations and Trends® in Machine learning* 3.1 (2011), pp. 1–122.