A Simple Approach to Automated Spectral Clustering

Jicong Fan[1,2], Yiheng Tu[3], Zhao Zhang[4], Mingbo Zhao[5], Haijun Zhang[6]

[1]The Chinese University of Hong Kong, Shenzhen
[2]Shenzhen Research Institute of Big Data, Shenzhen
[3]Chinese Academy of Science, Beijing
[4]Hefei University of Technology, Hefei
[5]Donghua University, Shanghai
[6]Harbin Institute of Technology, Shenzhen

Motivation

- Spectral Clustering (SC)
  - step 1: construct a similarity matrix
  - step 2: perform normalized cut [Shi and Malik, 2000]

Motivation

- Spectral Clustering (SC)
  - step 1: construct a similarity matrix
  - step 2: perform normalized cut [Shi and Malik, 2000]
- Limitations of SC
  - performance heavily relies on the quality of affinity matrix
  - difficult to do model and hyperparameter selection

Automated Spectral Clustering

- Relative-Eigen-Gap (REG)

$$\mathrm{reg}(\boldsymbol{L}) := \frac{\sigma_{k+1}(\boldsymbol{L}) - \frac{1}{k}\sum_{i=1}^{k}\sigma_i(\boldsymbol{L})}{\frac{1}{k}\sum_{i=1}^{k}\sigma_i(\boldsymbol{L}) + \varepsilon} \tag{1}$$

$\boldsymbol{L}$ is the laplacian matrix of the affinity matrix $\boldsymbol{A}$

Automated Spectral Clustering

- Relative-Eigen-Gap (REG)

$$\mathrm{reg}(\boldsymbol{L}) := \frac{\sigma_{k+1}(\boldsymbol{L}) - \frac{1}{k}\sum_{i=1}^{k}\sigma_i(\boldsymbol{L})}{\frac{1}{k}\sum_{i=1}^{k}\sigma_i(\boldsymbol{L}) + \varepsilon} \tag{1}$$

  $\boldsymbol{L}$ is the laplacian matrix of the affinity matrix $\boldsymbol{A}$

- REG guided search

$$\begin{aligned} \underset{(f,\theta)\in\mathcal{F}\times\Theta}{\text{maximize}} \quad & \mathrm{reg}(\boldsymbol{L}), \\ \text{subject to} \quad & \boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2},\ \boldsymbol{A} = f_\theta(\boldsymbol{X}) \end{aligned} \tag{2}$$

  $\mathcal{F}$ is a set of pre-defined functions and Θ is a set of hyperparameters.

Automated Spectral Clustering

Table: A few examples of f and its θ for affinity matrix construction

f	K-NN	ϵ -neighborhood	Gaussian kernel	SSC	LRR	LSR	KSSC	AASC
θ	K	ϵ	σ	λ	λ	λ	λ, σ	$\sigma_1, \sigma_2, \ldots$

SSC: [Elhamifar and Vidal, 2013]; LRR: [Liu et al., 2013]; AASC: [Huang et al., 2012]

- AutoSC

$$\begin{aligned} \underset{(f,\theta)\in\mathcal{F}\times\Theta}{\text{maximize}} \quad & \text{reg}(\boldsymbol{L}), \\ \text{subject to} \quad & \boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}, \ \boldsymbol{A} = f_\theta(\boldsymbol{X}) \end{aligned}$$

Automated Spectral Clustering

Table: A few examples of f and its θ for affinity matrix construction

f	K-NN	ϵ -neighborhood	Gaussian kernel	SSC	LRR	LSR	KSSC	AASC
θ	K	ϵ	σ	λ	λ	λ	λ, σ	$\sigma_1, \sigma_2, \ldots$

SSC: [Elhamifar and Vidal, 2013]; LRR: [Liu et al., 2013]; AASC: [Huang et al., 2012]

- AutoSC

$$\underset{(f,\theta)\in\mathcal{F}\times\Theta}{\text{maximize}} \ \ \text{reg}(\boldsymbol{L}),$$
$$\text{subject to} \ \ \boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}, \ \boldsymbol{A} = f_\theta(\boldsymbol{X})$$

- Solve AutoSC
  - grid search
  - Bayesian optimization [Jones et al., 1998]

AutoSC with better affinity matrix construction method

- LSR (least squares representations) with thresholding

$$\underset{\boldsymbol{C}}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{C}\|_F^2 + \frac{\lambda}{2}\|\boldsymbol{C}\|_F^2 \tag{3}$$

  - $\text{diag}(\boldsymbol{C}) = \mathbf{0}$, $\boldsymbol{C} \leftarrow |\boldsymbol{C}|$
  - keep only the largest τ elements of each column of $\boldsymbol{C}$
  - $\boldsymbol{A} = (\boldsymbol{C} + \boldsymbol{C}^\top)/2$.

AutoSC with better affinity matrix construction method

- LSR (least squares representations) with thresholding

$$\underset{\boldsymbol{C}}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{C}\|_F^2+\frac{\lambda}{2}\|\boldsymbol{C}\|_F^2 \tag{3}$$

  - $\text{diag}(\boldsymbol{C})=\mathbf{0}$, $\boldsymbol{C}\leftarrow|\boldsymbol{C}|$
  - keep only the largest τ elements of each column of $\boldsymbol{C}$
  - $\boldsymbol{A}=(\boldsymbol{C}+\boldsymbol{C}^\top)/2$.
- KLSR with thresholding

$$\underset{\boldsymbol{C}}{\text{minimize}} \quad \frac{1}{2}\|\phi(\boldsymbol{X})-\phi(\boldsymbol{X})\boldsymbol{C}\|_F^2+\frac{\lambda}{2}\|\boldsymbol{C}\|_F^2 \tag{4}$$

  - The post-processing is similar to LSR

AutoSC with better affinity matrix construction method

- LSR (least squares representations) with thresholding

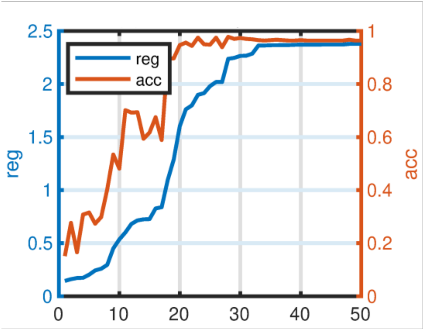
$$\underset{\boldsymbol{C}}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{C}\|_F^2 + \frac{\lambda}{2}\|\boldsymbol{C}\|_F^2 \tag{3}$$

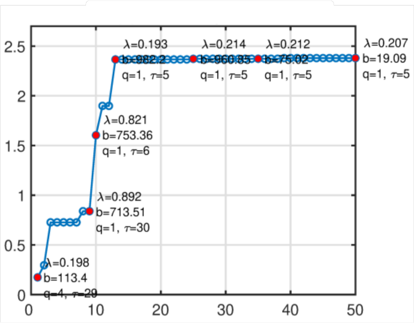
  - $\text{diag}(\boldsymbol{C}) = \mathbf{0}$, $\boldsymbol{C} \leftarrow |\boldsymbol{C}|$
  - keep only the largest τ elements of each column of $\boldsymbol{C}$
  - $\boldsymbol{A} = (\boldsymbol{C} + \boldsymbol{C}^\top)/2$.
- KLSR with thresholding

$$\underset{\boldsymbol{C}}{\text{minimize}} \quad \frac{1}{2}\|\phi(\boldsymbol{X}) - \phi(\boldsymbol{X})\boldsymbol{C}\|_F^2 + \frac{\lambda}{2}\|\boldsymbol{C}\|_F^2 \tag{4}$$

  - The post-processing is similar to LSR
- Determine τ by AutoSC
- Theoretical guarantee of LSR and KLSR (see the paper)

- AutoSC-BO with KLSR on the first 10 subjects of YaleB [Kuang-Chih et al., 2005] Face dataset





More numerical results can be found in the paper