Relations Between Adjacency And Modularity Graph Partitioning: Principal Component Analysis vs. Modularity Component Analysis

Hansi Jiang

Carl Meyer

North Carolina State University

October 27, 2015

Application and Methodology

- Application
  - Graph partitioning or data clustering
- Methodology
  - Spectral methods based on modularity components or principal components
  - Relation between dominant eigenvectors of unnormalized modularity and similarity matrices
  - Comparing modularity components and principal components

Definitions

- Given a graph or data, we have:
  - Modularity matrix: $\mathbf{B} = \mathbf{A} - \mathbf{d}\mathbf{d}^T/(2m)$
  - $\mathbf{A}$ is an adjacency matrix (graph) or similarity matrix (data)
  - $\mathbf{d} = \mathbf{A}\mathbf{e}$ is the degree vector
  - $m = \mathbf{d}^T\mathbf{e}$ is the number of edges in the graph

Definitions

- Modularity: $Q(\mathbf{s}) = \mathbf{s}^T \mathbf{B} \mathbf{s}$
- Invented by Newman and Girvan (2004)
- Goal: Pick $\mathbf{s}$, $\|\mathbf{s}\|_2 = 1$ s.t. Q is maximized
- Dominant eigenvector of $\mathbf{B}$ maximizes Q

Properties

- Modularity matrix $\mathbf{B}$
  - Symmetric, not semi-definite
  - $(0, \mathbf{e})$ is an eigenpair of $\mathbf{B}$
  - The eigenvector corresp. to the largest eigenvalue is used in partitioning
- Spectral clustering uses the eigenvector corresp. to the second smallest eigenvector of $\mathbf{L} = \mathbf{D} - \mathbf{A}$

Dominant Eigenvectors of Modularity and Similarity Matrices

- Modularity matrix: $\mathbf{B} = \mathbf{A} - \mathbf{d}\mathbf{d}^T/(2m)$
- We want to give an explicit expression of $\mathbf{b}_1$ in terms of eigenvalues and eigenvectors of $\mathbf{A}$.

Theorem 1

The dominant eigenvector of $\mathbf{B}$ is $\mathbf{b}_1 = \frac{1}{\|\mathbf{d}\|_2} \sum_{i=1}^{n} \frac{\mathbf{v}_i^T \mathbf{d}}{\alpha_i - \beta_1} \mathbf{v}_i$.

- $(\alpha_i, \mathbf{v}_i)$'s are eigenpairs of $\mathbf{A}$
- The proof is based on Cauchy's Interlacing Theorem.

Dominant Eigenvectors of $\mathbf{B}$ when $\mathbf{A} = \mathbf{X}^T\mathbf{X}$

- Suppose $\mathbf{X}_{p \times n} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is the uncentered data matrix
- Consider a special case: $\mathbf{A} = \mathbf{X}^T\mathbf{X}$
- Suppose $\mathbf{A}$ has k positive eigenvalues and they are simple
- The first $k-1$ largest eigenvalues of $\mathbf{B}$ are simple by the interlacing theorem
- The result in theorem 1 can be extended

Dominant Eigenvectors of $\mathbf{B}$ when $\mathbf{A} = \mathbf{X}^T\mathbf{X}$

Theorem 2

Suppose the largest $k-1$ eigenvalues of $\mathbf{B}$ are $\beta_1 > \beta_2 > \cdots > \beta_{k-1}$ and the nonzero eigenvalues of $\mathbf{A} = \mathbf{X}^T\mathbf{X}$ are $\alpha_1 > \alpha_2 > \cdots > \alpha_k$. Further suppose that for $1 \leq i \leq k-1$ we have $\beta_i \neq \alpha_i$ and $\beta_i \neq \alpha_{i+1}$. Then the $k-1$ dominant eigenvectors of $\mathbf{B}$ can be written by

$$\mathbf{b}_i = \sum_{j=1}^{k} \gamma_{ij}\mathbf{v}_j,$$

where

$$\gamma_{ij} = \frac{\mathbf{v}_j^T\mathbf{d}}{(\alpha_j - \beta_i)\|\mathbf{d}\|_2}.$$

Application: Modularity Components

Definition

Let

$$\mathbf{m}_i^T = \mathbf{b}_i^T \mathbf{X}^\dagger = \sum_{j=1}^{k} \frac{\gamma_{ij}}{\sigma_j} \mathbf{u}_j^T,$$

then the *i*-th modularity component is defined to be

$$\mathbf{c}_i = \frac{\mathbf{m}_i}{\|\mathbf{m}_i\|_2}.$$

- σ_j is the j-th singular value of $\mathbf{X}$, $\mathbf{u}_j$ corresp. singular vector
- Can help to explain why using several eigenvectors of $\mathbf{B}$ to cluster data is reasonable

Review: Properties of Principal Components

- Principal Components derived from **X** are orthogonal to each other
- Require centering data
- Projection of the centered data onto to the span of principal components gives clusters
- The first principal component has maximal variance
- Each succeeding principal component has maximal variance with the constraint that it is orthogonal to all prior principal components

Properties of Modularity Components

- Modularity Components derived from $\mathbf{X}$ are orthogonal to each other
- Does not require centering
- Projection of the uncentered data onto to the span of modularity components gives clusters
- The first modularity component has maximal modularity
- Each succeeding modularity component has maximal modularity with the constraint that it is orthogonal to all prior modularity components

Significance of Modularity Components

- Analogous to the principal components
- Does not require centering
- Gives reason to use multiple eigenvectors of **B** to cluster data

Example: PenDigit Dataset (Subset of MNIST)

- ~12,000 data points
- Data points: vectors converted from a grey scale image
- Subset used: Digits 1, 5 and 7

	MCA Result		
	1	5	7
1	4487	196	1
5	203	3526	66
7	870	101	3430

Table 1: MCA Result on PenDigit Data

	PCA Result		
	1	5	7
1	4480	203	1
5	429	3290	76
7	787	112	3502

Table 2: PCA Result on PenDigit Data

Conclusion

- The exact linear relation between the dominant eigenvectors of $\mathbf{B}$ in terms of the eigenvectors of $\mathbf{A}$ is given
- The definition of modularity components and their properties are given
- The comparison between modularity components and principal components is given
- An example comparing the results from MCA and PCA is given

Thanks

Thank you! Questions and Answers

[1] Hansi Jiang and Carl Meyer.
Modularity component analysis versus principal component analysis.
submitted(http://arxiv.org/pdf/1510.05492v1.pdf), 2015.

[2] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner.
Gradient-based learning applied to document recognition.
Proceedings of the IEEE, 86(11):2278–2324, 1998.

[3] Mark EJ Newman and Michelle Girvan.
Finding and evaluating community structure in networks.
Physical review E, 69(2):026113, 2004.

[4] Ulrike Von Luxburg.
A tutorial on spectral clustering.
Statistics and computing, 17(4):395–416, 2007.