

Spectral clustering

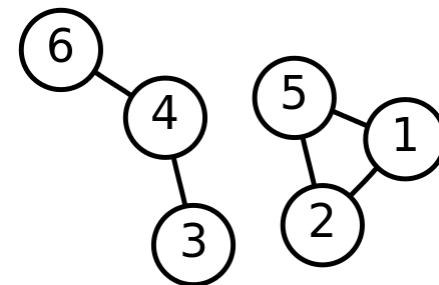
From Wikipedia, the free encyclopedia

In multivariate statistics and the clustering of data, **spectral clustering** techniques make use of the spectrum (eigenvalues) of the similarity matrix of the data to perform dimensionality reduction before clustering in fewer dimensions. The similarity matrix is provided as an input and consists of a quantitative assessment of the relative similarity of each pair of points in the dataset.

In application to image segmentation, spectral clustering is known as segmentation-based object categorization.

Contents

- 1 Algorithms
- 2 Relationship with *k*-means
- 3 Measures to compare clusterings
- 4 See also
- 5 References



An example of two connected graphs

Algorithms

Given an enumerated set of data points, the similarity matrix may be defined as a symmetric matrix \mathbf{A} , where $A_{ij} \geq 0$ represents a measure of the similarity between data points with indices i and j . The general approach to spectral clustering is to use a standard clustering method (there are many such methods, *k*-means is discussed below) on relevant eigenvectors of a Laplacian matrix of \mathbf{A} . There are many different ways to define a Laplacian which have different mathematical interpretations, and so the clustering will also have different interpretations. The eigenvectors that are relevant are the ones that correspond to smallest several eigenvalues of the Laplacian except for the smallest eigenvalue which will have a value of 0. For computational efficiency, these eigenvectors are often computed as the eigenvectors corresponding to the largest several eigenvalues of a function of the Laplacian.

One spectral clustering technique is the **normalized cuts algorithm** or *Shi–Malik algorithm* introduced by Jianbo Shi and Jitendra Malik,^[1] commonly used for image segmentation. It partitions points into two sets $(\mathbf{B}_1, \mathbf{B}_2)$ based on the eigenvector \mathbf{v} corresponding to the second-smallest eigenvalue of the symmetric normalized Laplacian defined as

$$L^{\text{norm}} := I - D^{-1/2} A D^{-1/2},$$

where D is the diagonal matrix

$$D_{ii} = \sum_j A_{ij}.$$

A mathematically equivalent algorithm [2] takes the eigenvector corresponding to the largest eigenvalue of the random walk normalized Laplacian matrix $\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$. The Meila–Shi algorithm has been examined in the context of diffusion maps which is found to be related to computational quantum mechanics.[3]

Another possibility is to use the Laplacian matrix defined as

$$\mathbf{L} := \mathbf{D} - \mathbf{A}$$

rather than the symmetric normalized Laplacian matrix.

Partitioning may be done in various ways, such as by computing the median m of the components of the second smallest eigenvector \mathbf{v} , and placing all points whose component in \mathbf{v} is greater than m in \mathbf{B}_1 , and the rest in \mathbf{B}_2 . The algorithm can be used for hierarchical clustering by repeatedly partitioning the subsets in this fashion.

If the similarity matrix \mathbf{A} has not already been explicitly constructed, the efficiency of spectral clustering may be improved if the solution to the corresponding eigenvalue problem is performed in a matrix-free fashion (without explicitly manipulating or even computing the similarity matrix), as in the Lanczos algorithm.

For large-sized graphs, the second eigenvalue of the (normalized) graph Laplacian matrix is often ill-conditioned, leading to slow convergence of iterative eigenvalue solvers. Preconditioning is a key technology accelerating the convergence, e.g., in the matrix-free LOBPCG method. Spectral clustering has been successfully applied on large graphs by first identifying their community structure, and then clustering communities.[4]

Spectral clustering is closely related to nonlinear dimensionality reduction, and dimension reduction techniques such as locally-linear embedding can be used to reduce errors from noise or outliers.[5]

Free software to implement spectral clustering is available in large open source projects like Scikit-learn,[6] MLlib for pseudo-eigenvector clustering using the power iteration method,[7] and R.[8]

Relationship with k -means

The kernel k -means problem is an extension of the k -means problem where the input data points are mapped non-linearly into a higher-dimensional feature space via a kernel function $k(\mathbf{x}_i, \mathbf{x}_j) = \varphi^T(\mathbf{x}_i)\varphi(\mathbf{x}_j)$. The weighted kernel k -means problem further extends this problem by defining a weight w_r for each cluster as the reciprocal of the number of elements in the cluster,

$$\max_{\{C_s\}} \sum_{r=1}^k w_r \sum_{x_i, x_j \in C_r} k(x_i, x_j).$$

Suppose F is a matrix of the normalizing coefficients for each point for each cluster $F_{ij} = w_r$ if $i, j \in C_r$ and zero otherwise. Suppose K is the kernel matrix for all points. The weighted kernel k -means problem with n points and k clusters is given as,

$$\max_F \text{trace}(KF)$$

such that

$$F = G_{n \times k} G_{k \times n}^T \\ G^T G = I$$

such that $\text{rank}(G) = k$. In addition, there are identity constraints on F given by,

$$F \cdot \mathbb{1} = \mathbb{1}$$

where $\mathbb{1}$ represents a vector of ones.

$$F^T \mathbb{1} = \mathbb{1}$$

This problem can be recast as,

$$\max_G \text{trace}(G^T G).$$

This problem is equivalent to the spectral clustering problem when the identity constraints on F are relaxed. In particular, the weighted kernel k -means problem can be reformulated as a spectral clustering (graph partitioning) problem and vice versa. The output of the algorithms are eigenvectors which do not satisfy the identity requirements for indicator variables defined by F . Hence, post-processing of the eigenvectors is required for the equivalence between the problems.^[9] Transforming the spectral clustering problem into a weighted kernel k -means problem greatly reduces the computational burden.^[10]

Measures to compare clusterings

Ravi Kannan, Santosh Vempala and Adrian Vetta in the following paper^[11] proposed a bicriteria measure to define the quality of a given clustering. They said that a clustering was an (α, ϵ) -clustering if the conductance of each cluster (in the clustering) was at least α and the weight of the inter-cluster edges was at most ϵ fraction of the total weight of all the edges in the graph. They also look at two approximation algorithms in the very same paper.

See also

- Affinity propagation
- Kernel principal component analysis
- Cluster analysis
- Spectral graph theory

References

1. Jianbo Shi and Jitendra Malik, "Normalized Cuts and Image Segmentation" (<http://www.cs.berkeley.edu/~malik/papers/SM-ncut.pdf>), IEEE Transactions on PAMI, Vol. 22, No. 8, Aug 2000.
2. Marina Meilă & Jianbo Shi, "Learning Segmentation by Random Walks (<http://www.citeulike.org/user/mpotamias/article/498897>)", Neural Information Processing Systems 13 (NIPS 2000), 2001, pp. 873–879.
3. Scott, T.C.; Madhusudan Therani; Xing M. Wang (2017). "Data Clustering with Quantum Mechanics" (<http://www.mdpi.com/2227-7390/5/1/5>). *Mathematics*. **5** (1): 1–17. doi:10.3390/math5010005 (<https://doi.org/10.3390/math5010005>).
4. Zare, Habil; P. Shooshtari; A. Gupta; R. Brinkman (2010). "Data reduction for spectral clustering to analyze high throughput flow cytometry data" (<http://www.biomedcentral.com/1471-2105/11/403>). *BMC Bioinformatics*. **11**: 403. PMC 2923634 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2923634>) . PMID 20667133 (<https://www.ncbi.nlm.nih.gov/pubmed/20667133>). doi:10.1186/1471-2105-11-403 (<https://doi.org/10.1186/1471-2105-11-403>).
5. Arias-Castro, E. and Chen, G. and Lerman, G. (2011), "Spectral clustering based on local linear approximations.", *Electronic Journal of Statistics*, **5**: 1537–1587, doi:10.1214/11-ejs651 (<https://doi.org/10.1214/11-ejs651>)
6. <http://scikit-learn.org/stable/modules/clustering.html#spectral-clustering>
7. <http://spark.apache.org/docs/latest/mllib-clustering.html#power-iteration-clustering-pic>
8. <https://cran.r-project.org/web/packages/kernlab>
9. Dhillon, I.S. and Guan, Y. and Kulis, B. (2004). "Kernel *k*-means: spectral clustering and normalized cuts". *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 551–556.
10. Dhillon, Inderjit; Yuqiang Guan; Brian Kulis (November 2007). "Weighted Graph Cuts without Eigenvectors: A Multilevel Approach". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **29** (11): 1–14. doi:10.1109/tpami.2007.1115 (<https://doi.org/10.1109/tpami.2007.1115>).
11. Kannan, Ravi; Vempala, Santosh; Vetta, Adrian. "On Clusterings : Good, Bad and Spectral". *Journal of the ACM*. **51**: 497–515. doi:10.1145/990308.990313 (<https://doi.org/10.1145/990308.990313>).

Retrieved from "https://en.wikipedia.org/w/index.php?title=Spectral_clustering&oldid=794221036"

-
- This page was last edited on 6 August 2017, at 17:28.

- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.