AMRITA SCHOOL OF ENGINEERING, CHENNAI

RCK Nagar-SH 50A, Vengal, Tamil Nadu-601103

21MAT204 MATHEMATICS FOR INTELLIGENCE SYSTEM

Regulation: 2022

PROBLEM STATEMENT- IMPLENTATION OF SPECTRAL CLUSTERING

SUBMITTED BY: TEAM_9

APARNA S	CH.EN.U4AIE21107
G V ROHAN	CH.EN.U4AIE21113
HARSHITHA M	CH.EN.U4AIE21115
KORRAYI SAITEJA	CH.EN.U4AIE21122
M SRAVANI	CH.EN.U4AIE21125

ACKNOWLEDGEMENT

We would like to offer our sincere pranams at the lotus feet of Universal guru, MATA AMRITANANDAMAYI DEVI who blessed us with her grace to make this a successful major project.

We express our deep sense of gratitude to Dr.PRASANNA KUMAR, Chairperson, for his constant help, suggestions, and inspiring guidance. We are grateful to our guide Dr.R.Ishwariya ,Department of MATHEMATICS, ASE, Chennai for his invaluable support and guidance during the major project work.

We would also like to extend our gratitude to our director Shri. Manikandan, Principal Dr. V JAYA KUMAR who has always encouraged us. We are also thankful to all our classmates who have always been a source of strength, for always being there and extending their valuable bits of help to the successful completion of this work.

STUDENT DECLARATION

I hereby declare that my Project Report titled SPECTRAL CLUSTERING record for the course **21MAT204 – MATHEMATICS FOR INTELLIGENCE SYSTEMS** which I have submitted to Department of Computer Science and Engineering (AIE), Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai in partial fulfillment of the credit requirements for the B.Tech. degree, is my authenticwork done under the guidance of **Dr.R.Ishwariya**. This project report has not been copied, duplicated or plagiarised from any other paper, journal, document or book and has not been submitted to any educational institute, course, department or otherwise for the award of any credit, certificate, diploma, degree or recognition. This is an authentic piece of work and in case there is any query regarding the same, I shall be held responsible for answering any queries in this regard.

Date: 23.12.22

Student name and Register no.:

Supervisor’s Name:

SPECTRAL CLUSTERING

Problem statement- To implement spectral clustering

Abstract

The term "spectral" refers to the clustering outcomes are obtained by analyzing the graph Laplacian's spectrum. Spectral clustering and Laplacian eigenmaps algorithm are very effective dimensionality reduction techniques. Spectral graph theory is also the foundation of the Laplacian eigenmaps concept. An adjacency matrix must be built, and the eigen-decomposition of the related Laplacian matrix must be determined, for the general spectral clustering approach.

There are still a number of unresolved problems with spectral clustering algorithms, despite their many practical triumphs, which cluster points using matrices generated from the data. The eigenvectors are used in a wide range of algorithms in various ways, which is the first point to make. The second problem is that many of these algorithms lack evidence that they would produce accurate clustering. The spectral clustering approach we provide in this paper is simple and Python-based. On a variety of difficult clustering issues, we also provide remarkably effective experimental outcomes. In this research, we propose a modified NJW method that is simple to implement. Furthermore, the method can estimate the value of k to produce acceptable clusters. It performs well, as evidenced by experimental findings on a range of difficult clustering situations.

Keywords: Spectral clustering, Laplacian graph, segmentation, NJW Algorithm

1. Introduction

The common and effective method of dividing a graph's nodes into clusters with few outside connections relative to volume (sum of degrees) is known as spectral clustering.

One of the most often used clustering methods in the current era is spectral clustering. It frequently outperforms established clustering techniques like the k-means algorithm and is straightforward to construct, can be solved effectively by common linear algebra software, and is less computationally intensive. Spectral clustering first seems a little strange, and it is not immediately clear why it even functions and what it actually accomplishes.

2. *Literature survey*

- In previous research on spectral clustering, it is applied in many areas, including image processing.
- In recent years, it is implemented to solve efficiently by standard linear algebra software.
- Study on different graph Laplacians and their basic properties, present the most common spectral clustering algorithms and those derived algorithms from scratch by different approaches.
- Survey paper on methods like statistics, machine learning, pattern recognition, data mining, and image processing are discussed in detail and later introduces the relationship between spectral clustering and k-means clustering by Nystrom methods.
- Study on spectral clustering which depend on the eigenstructure of a similarity matrix to partition points into disjoint clusters with points in the same cluster.
- In previous studies, they investigated the consistency of the popular family spectral clustering algorithms, with the help of graph laplacian matrices for increasing the sample size, those eigenvectors converge to the eigenvectors of certain limit operators.
- Study on Diffusion Maps, Spectral clustering and eigenfunctions of Fokker-Planck operators.

3. Theory

a. Spectral Clustering

In spectral clustering, we build a graph out of the data and divide it up by analyzing its connectivity. This differs from some of the more well-known methods, which depend on the presumption that clusters are concentrated in terms of typically Cartesian distance, such as k-means or developing a mixed model using EM. When the clusters are complicated or have an unknown shape, they typically tend to not function effectively.

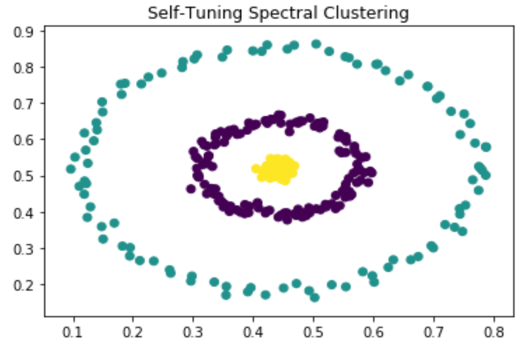


Fig1.1 Self-tuning spectral clustering in concentric circles

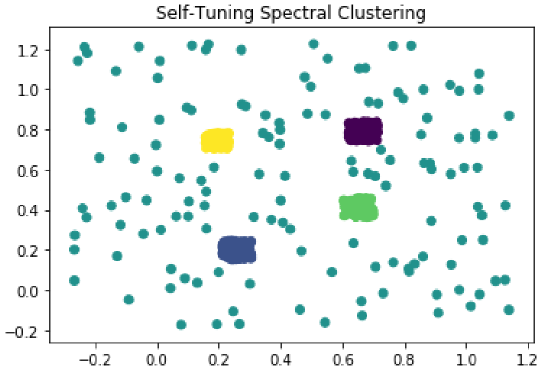


Fig1.2 Self-tuning spectral clustering in hubs and background in two different densities

4. Methodology

b. How spectral clustering is calculated?

- Create a distance matrix
- Convert the distance matrix into an affinity matrix A
- Calculate the Laplacian matrix L= D-A and the degree of matrix D.
- Find the eigenvalues and eigenvectors of L.
- The matrix is formed using the eigenvectors and the k-largest eigenvalues calculated in the preceding phase.
- Normalize the vectors.
- Cluster the data points in k-dimensional space.

c. Graph Laplacian

The graph Laplacian has a useful characteristic that enables the clustering algorithm. A graph G with n nodes Laplacian matrix is defined as

$$L=D-A$$

L→ nxn matrix with elements given by

$$L_{i,j} = \begin{cases} \deg(v_i) & \text{if } i=j \\ -w_{i,j} & \text{if } i \neq j \text{ and there is an edge} \\ & \text{between I and j} \\ 0 & \text{otherwise} \end{cases}$$

D→ Degree matrix

A→ Adjacency matrix of graph G

(G is an unweighted graph)

d. Implementing the NJW algorithm

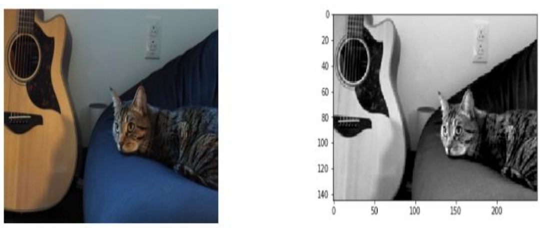
- Determine affinity matrix A.
- Create Laplacian $L = D^{-1/2}AD^{-1/2}$ by defining D as the orthogonal matrix whose entries are (i, i) elements is the sum of A's i-th row.
- To create matrix $X = [x_1x_2\ldots x_k]$ find the k-largest distinct eigenvectors $x_1,x_2,..,x_k$ by arranging them in columns.
- Each row in X should be renormalized to have a unit length.
  This algorithmic phase, which deals with connectivity variations within a cluster, is cosmetic.
- Using other algorithm, such as k-means, group each row of Y into k clusters.
- Assign the original point. We reuse the clusters variable in terms of implementation.

e. Determining the number of clusters

The number of clusters is currently a requirement for all the algorithms we've discussed. However, the ideal cluster count is frequently confusing in real-world situations. In order to determine the number of groups, the approach of Zelnik-Manor and Perona (2014) suggests using non-maximal suppression after rotating the eigenvector.

f. Segmentation of image

The following is one of the use case of clustering on segmentation of the image



(a) Original Image (b) Greyscale image used in algorithm

Fig1.3 Segmentation of image

5. Results and Discussion

The boundaries of the guitar and portion of the cat are indicated with darker lines. Other modes of segmentation are represented by other modes of the eigenvectors. The following Fig 1.4, 1.5, 1,6 represents the segmentation of the image different eigenvalues.

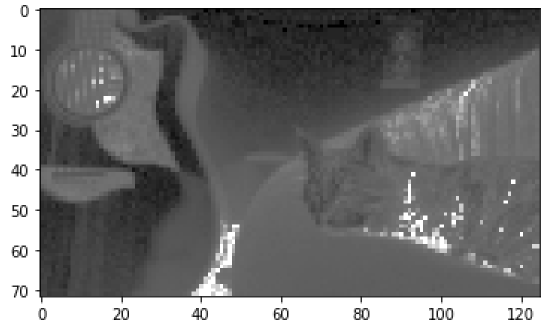


Fig 1.4

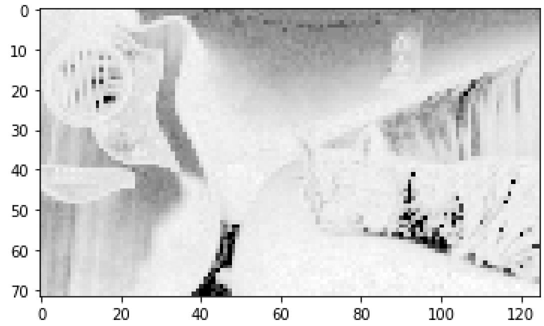


Fig 1.5

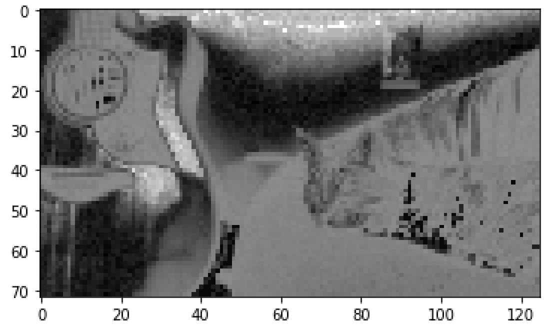


Fig 1.6

Here are some segmentation results of the image when the value of eigenvectors are different.

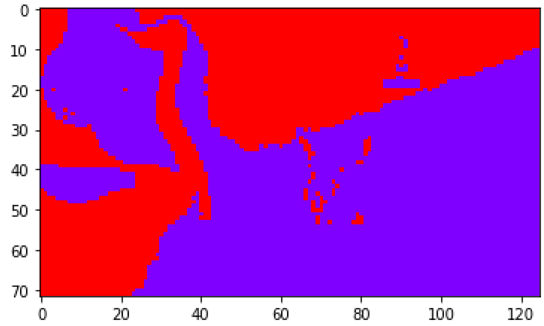


Fig 1.7 Segmentation of image when K=2

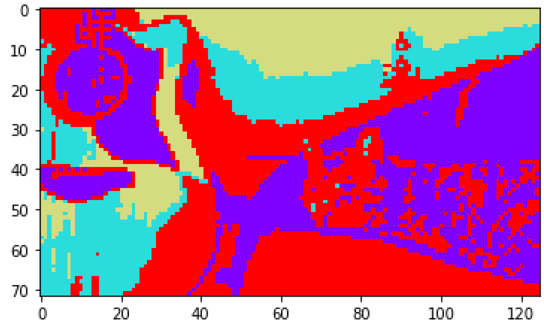


Fig 1.8 Segmentation of image when K=4

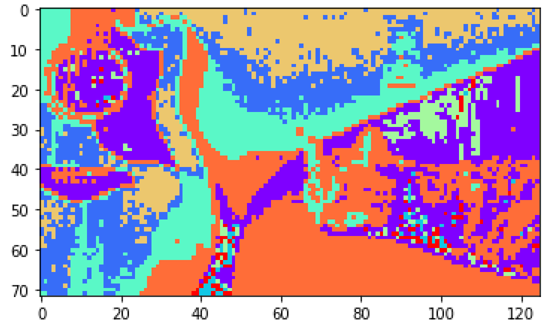


Fig 1.9 Segmentation of image when K=8

6. Applications

- Text image segmentation
- Entity resolution
- speech separation

- spectral clustering of protein sequences
- educational data mining
- image segmentation
- As spectral clustering is the graph-based theory, this approach is used to identify communities of vertices in a graph based on the edges connecting them.

7. Limitations of spectral clustering

- In general, the K-means method implies that the clusters be round or spherical, that is, located within a k-radius of the cluster centroid.
- The cluster centroid in k-means must be determined after numerous rounds.
- Points that are far apart but connected belong to the same cluster, while points that are closer to one another but not connected could belong to different clusters.
- Clusters in the spectrum do not have a set form or pattern.
- Even though computing eigenvalues and eigenvectors before clustering is necessary for large datasets, it may be expensive to do so.

8. Conclusion

Here, we discussed spectral clustering, how we calculated the Laplacian graph, and how the NJW algorithm is implemented in spectral clustering. The RGB-colored image is converted into greyscale and then it undergoes segmentation image and segmentation of the image is done based on the values of eigenvectors. Here, the eigenvector indicates the mode of segmentation of the image when k=2, 4, 8.

9. References

[1] Meila, M. (2016). Spectral Clustering: a Tutorial for the 2010's. Handbook of cluster analysis, 1-23.

[2] Xiang, T., & Gong, S. (2008). Spectral clustering with eigenvector selection. Pattern Recognition, 41(3), 1012-1029.

[3] Amghibech, S. (2003). Eigenvalues of the discrete p-Laplacian for graphs. *Ars Combin.*, *67*, 283--302.

[4] Liu, J., Wang, C., Danilevsky, M., & Han, J. (2013, June). Large-scale spectral clustering on graphs. In *Twenty-Third International Joint Conference on Artificial Intelligence*.

[5] Andrew Ng, Michael Jordan and Yair Weiss. "On spectral clustering: Analysis and an algorithm" NeurIPS 2001.

[6] Suryanarayana, S. V., Rao, G. V., & Swamy, G. V. (2015). A Survey: Spectral Clustering Applications and its Enhancements. *International Journal of Computer Science and Information Technologies*, *6*(1), 185-189.

[7] Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, *17*(4), 395-416.

[8] Yan, Donghui, Ling Huang, and Michael I. Jordan. "Fast approximate spectral clustering." In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 907-916. 2009.

[9] Von Luxburg, U., Belkin, M., & Bousquet, O. (2008). Consistency of spectral clustering. *The Annals of Statistics*, 555-586.