

UNIVERSITY OF YORK

MLAP

MACHINE LEARNING AND APPLICATIONS

Open Examination

Examination number:

Y3606797

Contents

1	Task 2	2
1.1	Relationship to random walks on graph	2
1.2	Commute time embedding advantages	2
1.3	Commute time embedding criticisms	3

1 Task 2

1.1 Relationship to random walks on graph

The prevailing idea in the papers is to preserve the locality of the data points across the mapping, graphs are constructed and it is useful to consider measures of distance between data on the manifolds which relate to a random walk model on the graphs:

The Laplacian eigenmap method [1] does so by using the weighted Laplacian of the similarity graph to approximate the Laplace-Beltrami operator. The weights are defined by a kernel (the heat kernel) which can be seen as the probability of a random walker making a step from x to y in the graph.

The diffusion map method [2] introduces a Diffusion operator which is also related to the weighted Laplacian as in [1]. The spectral properties of a Markov chain are used to describe the geometry of the data. A diffusion distance is defined relating the probability of a random walk starting at x and arriving at y in t steps (intuitively the rate of connectivity of the points x and y in the data), this is computed by the using the spectral decomposition of the weighted Laplacian.

The commute times method [3] uses a similar approach to [1] in that it relates to the random walk by the heat kernel of the weighted graph. It defines the commute time as the expected time for the random walk to travel from node u to reach node v and then return.

The methods of Laplacian eigenmap [1] and Commute times [3] are related because they both attempt to minimise a similar objective function, related to the heat kernel and associated random walk model.

The Laplacian eigenmap [1] method is a special case of Diffusion map [2] which handles manifolds where the data is sampled uniformly.

1.2 Commute time embedding advantages

Commute times reflect the combined effect of all possible weighted paths between a pair of nodes [3]. This means they are more robust against edge insertions and deletions. This combined effect results in pairs nodes that are clustered together by commute time methods having a higher affinity (affinity meaning the chosen measure of similarity) and having lower affinity with nodes outside the cluster.

This is also observed in the affinity matrix, there is a stronger block structure present, which has been observed as a key insight in the success of clustering methods.

More evidence of commute times advantageous clustering ability comes from the fact that the smallest eigenvector in the commute time matrix has a more stable distribution over it's components, this means that there is a stronger discrimination between data clusters.

The commute time method conveys more information about a cluster due to the fact that it takes into account the full Laplacian eigenspectrum when being computed [3].

These results are important as they allow the use of commute time embedding to work with data that is inherently difficult due to contamination with noise and outlier data points. Similar objects will have smaller commute times and outliers and noise will be spread sparsely which allows a clustering method such as K-Means to effectively differentiate between the data.

1.3 Commute time embedding criticisms

The paper explores the commute times for graphs which are large (have a large number of nodes). The main conclusion is that for large graphs the commute time does not provide any useful information about the global structure of the graph. Instead of providing information about clustering within a graph, it is shown that for large graphs the commute time can be approximated accurately with the following simple equation, which is just a local density measure for the nodes:

$$1/d_i + 1/d_j$$

Where d_i and d_j are the degrees of the nodes i and j respectively. It is stated that graphs with nodes in the 1000s are susceptible to this effect and that for high dimensional data, such as that in most machine learning problems, the effect is even stronger [4].

References

- [1] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003. DOI: 10.1162/089976603321780317. eprint: <https://doi.org/10.1162/089976603321780317>. [Online]. Available: <https://doi.org/10.1162/089976603321780317>.
- [2] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, “Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 21, pp. 7426–7431, 2005, ISSN: 0027-8424. DOI: 10.1073/pnas.0500334102. eprint: <http://www.pnas.org/content/102/21/7426.full.pdf>. [Online]. Available: <http://www.pnas.org/content/102/21/7426>.
- [3] H. Qiu and E. R. Hancock, “Clustering and embedding using commute times,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1873–1890, Nov. 2007, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2007.1103.
- [4] U. von Luxburg, A. Radl, and M. Hein, “Hitting times, commute distances and the spectral gap for large random geometric graphs,” *CoRR*, vol. abs/1003.1266, 2010. arXiv: 1003.1266. [Online]. Available: <http://arxiv.org/abs/1003.1266>.