

CLUSTER ANALYSIS OF HAZE EPISODES BASED ON TOPOLOGICAL FEATURES

NUR FARIHA SYAQINA ZULKEPLI

Persistent homology is a tool used in topological data analysis (TDA) to extract essential topological features from data. Cluster analysis is a technique that is used for grouping objects in data sets into different clusters such that the members that are placed in the same cluster are similar with each other compared to the members in other clusters. Commonly, cluster analysis is applied based on available information of data without considering topological information. Thus, this study aims to apply cluster analysis based on topological features and the effectiveness of this approach is observed by comparing with original clustering approach. This is achieved by extracting topological features (connected components and holes) of particulate matter (PM_{10}) which is the major pollutant during haze episodes in Malaysia and the cluster members (months with and without haze) are observed. We apply Hierarchical Agglomerative Clustering Analysis (HACA), which is a standard technique in air quality studies, on its own (baseline) and the results are compared with combination of HACA and topological features (proposed). HACA process is initiated by calculating dissimilarity distance between objects (months) and two objects with minimum distance is merged forming a single cluster. For the next cluster, new set of distance is calculated and again clusters with minimum distance merged and form a cluster. This process is continued in hierarchical way until one single cluster containing all objects is produced. Based on the results, proposed approach is able to cluster months with and without haze correctly compared with baseline approach.

REFERENCES

- [1] B.S. Everitt, S. Landau, M. Leese, and D. Stahl: *Cluster analysis*, John Wiley & Sons Ltd., (2011).
- [2] C.M. Pereira and R.F.D. Mello: *Persistent homology for time series and spatial data clustering*, *Expert Syst. Appl.*, **42**, No. 15-16 (2015), 6026–6038.
- [3] J.C.M. Pires, S.I.V. Sousa, M.C. Pereira, M.C.M. Alvim-Ferraz and F.G. Martins : *Management of air quality monitoring using principal component and cluster analysis–Part I: SO₂ and PM₁₀*, *Atmospheric Environment*, **42**, No. 6 (2008), 1249-1260.
- [4] N.Otter, M.A.Porter, U. Tillmann, P. Grindrod and H.A. Harrington: *A roadmap for the computation of persistent homology*, *EPJ Data Science*, **6**, No. 1 (2017), 17.

(NUR FARIHA SYAQINA ZULKEPLI) SCHOOL OF MATHEMATICAL SCIENCES, FACULTY OF SCIENCE AND TECHNOLOGY,
UNIVERSITI KEBANGSAAN MALAYSIA 43600 BANGI, SELANGOR, MALAYSIA
E-mail address: farihasyaqina@yahoo.com