Thai Journal of Mathematics. Vol. 2022, Issue Special Issue, pages 197 - 211

Clustering selected Terengganu's rainfall stations based on persistent homology

Abstract

Topological Data Analysis (TDA) is an emerging technique rooted from Algebraic Topology that reveals the geometrical structure of high-dimensional data sets. The approach in TDA is twofold; i.e. Persistent homology (PH) which quantifies topological invariants of a given data set, and Mapper which represents the high-dimensional data set into a 1D graph with nodes and edges. In this work, we employ PH as a tool to quantify the first dimensional holes (H_1) in the daily rainfall data set collected between 2012 to 2017 from six rainfall stations located in Terengganu, Malaysia. We divided the rainfall data based on one year (365 days) resulting in each station having five sets of rainfall point clouds. Since a rainfall point cloud consists of 1D data set, direct comparison of rainfalls between stations may not show a clear pattern. Thus, we first embed them into point clouds of 10D with time delay $\tau = 13$, using Takens embedding, preserving its original dynamical state. Next, we employ PH to generate persistence diagram to quantify 1D holes (H_1) in the rainfall point clouds and record its maximum persistence (H_1 lifespan), as its topological feature to characterize the distribution and intensity of rainfall. The first result is; based on past flood events, flood occurred when the year's average persistence score exceeds 13. The second part of this work involves clustering the stations using two approaches; the standard dynamic time warping (DTW) method which matches the rainfall frequency before computing its dissimilarity distance; and the PH approach using five years maximum H_1 lifespan as its distance matrix. The dendrograms produced by both clustering approaches are different, in which DTW has three distinct clusters, but dissimilar to its rainfall distribution. However, PH neatly ranks based on its annual rainfall intensity and recurrence, hence outperforming DTW approach..

Keywords

Clustering; Dynamic Time Warping; Persistent Homology; Rainfall; Time Series