An Interactive Clustering-Based Visualization Tool for Air Quality Data Analysis

Frederick Kin Hing Phoa

Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

Abstract

Examining PM2.5 (atmospheric particulate matter with a maximum diameter of 2.5 micrometers) seasonal patterns is an important research motivation for environmental scientists. An improved understanding of PM2.5 seasonal patterns can help environmental protection agencies (EPA) make decisions and develop complex models for controlling their concentration in different regions. This work proposes an R Shiny App web-based interactive tool, namely "model-based time series clustering (MTSC) tool," for clustering PM2.5 time series data using spatial and population variables and their temporal features, such as seasonality. Our tool allows stakeholders to visualize important characteristics of PM2.5 time series data, including temporal patterns and missing values, and cluster by attribute groupings. We apply the MTSC tool to cluster Taiwan's PM2.5 time series data based on air quality zones and type of stations, and thereby reveal an improvement in Taiwan's air quality since 2017. Our analysis also isolates southern Taiwan from other air quality zones due to a large difference in seasonality patterns, allowing EPA experts to execute different but effective air quality policies for different zones. This is a joint work with my former postdoc Dr. Mahsa Ashouri (University of Michigan - Ann Arbor), my master student Mr. Bo-Wu Chen (National Taiwan University), and my collaborators Dr. Chun-Houh Chen (Institute of Statistical Science, Academia Sinica) and Dr. Galit Shmueli (National Tsing Hua University).