

Statistics and Data Science Seminar

Change Point Inference for Non-Euclidean Data Sequences using Distance Profiles

Paromita Dubey (University of Southern California)

Abstract: We introduce a powerful scan statistic and the corresponding test for detecting the presence and pinpointing the location of a change point within the distribution of a data sequence with the data elements residing in a separable metric space (Ω, d) . These change points mark abrupt shifts in the distribution of the data sequence as characterized using distance profiles, where the distance profile of an element $x \in \Omega$ is the distribution of distances from x as dictated by the data. This approach is tuning parameter free, fully non-parametric and universally applicable to diverse data types, including distributional and network data, as long as distances between the data objects are available. We obtain an explicit characterization of the asymptotic distribution of the test statistic under the null hypothesis of no change points, rigorous guarantees on the consistency of the test in the presence of change points under fixed and local alternatives and near-optimal convergence of the estimated change point location, all under practicable settings. To compare with state-of-the-art methods we conduct simulations covering multivariate data, bivariate distributional data and sequences of graph Laplacians, and illustrate our method on real data sequences of the U.S. electricity generation compositions and Bluetooth proximity networks.

Wednesday, October 16 at 4:00 PM in 636 SEO