Multivariate Clustering of Nonparametric Time Trends

Marina Khismatullina

Abstract In this paper, we propose a new clustering procedure for uncovering hidden group structure in multivariate time series based on the trends they exhibit. While clustering methodologies for univariate time series have been extensively studied, relatively few rigorous statistical methods have been developed for clustering multivariate time series data. However, the increasing volume of temporal data presents an opportunity to reveal meaningful hidden structures by clustering entities across multiple time series recorded simultaneously for each of the entities. The proposed clustering procedure is based on a new dissimilarity measure that stems from a recently developed multiscale testing method for comparison of univariate nonparametric trends [1]. Building on this method, we extend the methodology to accommodate multivariate time series and propose a new distance measure that is able to capture the difference between the trends. We show that the clustering algorithm has the desired asymptotic properties and complement the project with a simulation study. For illustration purposes, we conduct an empirical study on the dataset that comprises information for 73 weather stations across Spain. We observe substantial differences when applying univariate versus multivariate clustering techniques, a finding that underscores the importance of adopting multivariate techniques to fully capture the complex dynamics inherent in multivariate time series data.

Keywords: multivariate time series, time series clustering, nonparametric statistics

References

1. Khismatullina, M., Vogt, M.: Multiscale comparison of nonparametric trend curves. arXiv preprint (2022) doi: 10.48550/arXiv.2209.10841

Marina Khismatullina
Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam,
e-mail: khismatullina@ese.eur.nl