Ensemble Forecasting of the Zika Space-Time Spread with Topological Data Analysis

Marwah Soliman (marwahs@clemson.edu), Clemson University, USA

As per the records of the World Health Organization, the first formally reported incidence of Zika virus occurred in Brazil in May 2015. The disease then rapidly spread to other countries in Americas and East Asia, affecting more than 1,000,000 people. Zika virus is primarily transmitted through bites of infected mosquitoes of the species Aedes (Aedes aegypti and Aedes albopictus). The abundance of mosquitoes and, as a result, the prevalence of Zika virus infections are common in areas which have high precipitation, high temperature, and high population density. Nonlinear spatio-temporal dependency of such data and lack of historical public health records make prediction of the virus spread particularly challenging. In this paper we enhance Zika forecasting by introducing the concepts of topological data analysis and, specifically, persistent homology of atmospheric variables, into the virus spread modeling. The key rationale is that topological summaries allow for capturing higher-order dependencies among atmospheric variables that otherwise might be unassessable via conventional spatio-temporal modelling approaches based on geographical proximity assessed via Euclidean distance. We introduce a new concept of cumulative Betti numbers and then integrate the cumulative Betti numbers as topological descriptors into three predictive machine learning models: random forest, generalized boosted regression, and deep neural network. Furthermore, to better quantify for various sources of uncertainties, we combine the resulting individual model forecasts into an ensemble of the Zika spread predictions using Bayesian model averaging. The proposed methodology is illustrated in application to forecasting of the Zika space-time spread in Brazil in the year 2018.