
SHAN GAO, University of Alberta

Outbreak or Not? A Framework for Detecting Infectious Disease Outbreaks

Forecasting the occurrence and absence of novel disease outbreaks is essential for disease management, yet existing methods are often context-specific, require a long preparation time, and non-outbreak prediction remains understudied. To address this gap, we propose a novel framework using a feature-based time series classification (TSC) method to forecast outbreaks and non-outbreaks. We tested our methods on synthetic data from a Susceptible–Infected–Recovered (SIR) model for slowly changing, noisy disease dynamics. Outbreak sequences give a transcritical bifurcation within a specified future time window, whereas non-outbreak (null bifurcation) sequences do not. We identified incipient differences, reflected in 22 statistical features and 5 early warning signal indicators, in time series of infectives leading to future outbreaks and non-outbreaks. Classifier performance, given by the area under the receiver-operating curve (AUC), ranged from 0.99 for large expanding windows of training data to 0.7 for small rolling windows. The framework is further evaluated on four empirical datasets: COVID-19 incidence data from Singapore, 18 other countries, and Edmonton, Canada, as well as SARS data from Hong Kong, with two classifiers exhibiting consistently high accuracy. Our results highlight detectable statistical features distinguishing outbreak and non-outbreak sequences well before potential occurrence, in both synthetic and real-world datasets presented in this study.