GEORGE MASON UNIVERSITY | College of Science

Ensemble Learning Interpretation of COVID-19 Simulation Models Through Clustering Latent Feature Representations and Aggregated Time-Series Forecasts with Deep Learning

Author(s):

RAYAN YU, ANDY CHEN, JOHN PESAVENTO, Taylor Anderson, Andreas Züfle, Hamdi Kavak, Joon-Seok Kim

Mentor(s):

Kavak, Computational and Data Sciences, George Mason University

Abstract:

Amidst the COVID-19 pandemic, there have been significant efforts to develop simulation models to forecast trends of the virus. However, there remains a lack of analysis between such models, resulting in uncertainty among policy-makers and the general population about the virus's future trends. This study develops two ensemble learning approaches of classification and regression to find agreement between prominent COVID-19 death forecast models for more comprehensive policy-making and judgement. To standardize uneven forecasts, we test imputation and normalization methods including, but not limited to linear interpolation, tensor factorization, and generative adversarial networks. We show that piecewise linear interpolation outperforms more complex approaches due their inability to exploit temporal autocorrelations. For classification, we apply a principal component analysis to extract latent feature representations. We employ the partitioning-around-medoids algorithm with a Manhattan distance metric in a k-medoids problem to classify the extracted representations into interpretable clusters. We name each medoid as a representative and quantify the range of deviation of other cluster members from the representative to easily interpret each cluster as a whole. For regression, we train a deep neural network (DNN) to predict ground truth COVID-19 deaths from a sliding window input of other models' aggregated predictions. We show that the DNN can adequately forecast the ground truth based only on these aggregated predictions while remaining robust against outliers, reaching a mean absolute error of under 200 when forecasting incidental deaths for a single day a week into the future. Our ensemble

models contribute a comprehensive method to analyze various consensus between current COVID-19 simulation models.

Presentation:

ASSIP 2020 - Rayan Yu, Andy Chen, John Pesavento - NSF Rapid