PI:

Md Yusuf Sarwar Uddin, PhD
Assistant Professor
Department of Computer Science and Electrical Engineering (CSEE)
University of Missouri-Kansas City
570E Flarsheim Hall, 5110 Rockhill Rd, Kansas City, MO 64110
816-235-6788 | muddin@umkc.edu
Website: http://m.web.umkc.edu/mu9b4/

Title:

Characterizing pandemic waves: Analysis of COVID-19 spread patterns across countries

Planned project duration

April 2021 – October 2021

Brief Description:

The spread of the COVID-19 pandemic (denoted by daily new confirmed cases or deaths) is observed to follow the shape of a wave (i.e., rise and fall of cases with time). Different geographic areas of the world (e.g., countries, states, or counties) have experienced different position and span of these "waves" over time. These waves are colloquially referred to as the first wave, the second wave, and so on. We propose that the presence and progress of these waves in a certain area can be represented by a sequence of five elements: R (rise), P (peak), S (steady), F (fall), and T (trough). With that, we obtain a 5-symbol string representation to denote the COVID-19 "spread pattern" of a target area. Despite a wide variation of pandemic wave patterns, we believe that different countries of the world can be grouped into a small number of heterogeneous classes each of which would have a representative COVID spread pattern. The key objective of this proposal is to analyze these spread patterns across countries. More specifically, we intend to—(a) characterize the spread patterns, (b) group geographical areas into a set of clusters based on their spread patterns, socio-demographics, and mobility indicators, and (c) develop a prediction model for each cluster that would forecast the spread into the future. We believe that this way of pattern recognition and prediction of COVID-19 spread would yield better results than a single model for all geographical areas and thus would help in taking effective prevention measures accordingly.

Keywords:

COVID-19 pandemic waves, spread patterns, cluster analysis, prediction models

Research Question(s) to be answered

(a) Characterizing COVID-19 spread patterns of different geographic areas in terms of the pandemic waves and their span and evolution over time. The areas can be countries in the international context or can be counties for the US context.
(b) Grouping geographic areas into a set of clusters based on the similarity/homogeneity of their COVID spread patterns.
(c) Developing a prediction model for each cluster that would forecast the progress and evolution of COVID-19 spread for that cluster. Unlike a single model for all areas, we argue that per-cluster model will ensure a better fitting of data and consequently will deliver better precision for future prediction.

What are the methods and approaches you're proposing to develop and use?

(a) For clustering, we will use clustering and pattern recognition techniques from the literature of machine learning and the popular Latent Class Analysis (LCA) for the joint estimation of class membership and class characteristics.
(b) For prediction modeling, we will apply the time-series analysis and time-dependent deep learning models such as recurrent neural networks and gated recurrent units. The model will be constructed for a specific spread pattern so that the training time of the model remains low as well as the accuracy of the model be high.

Geographic scope of the project

- The scope of the project is international as well as in the US.

Datasets to be used

- John Hopkins COVID-19 tracking dataset
- Google's community mobility report
- University of Maryland Transportation Institute (MTI) COVID-19 Impact Analysis Platform data
- Worldmeter datasets for population, area, GDP, data for countries
- US Census Bureau and Bureau of Labor Statistics for US population, employment, and county-level data

How will this project involve and engage patients and the general public?

Local policymakers can use of the developed models to track the spread of COVID-19 in their community and can take necessary action items to curve and contain the spread.

Timeline for project delivery and key milestones

- Apr 1 – starts the project
- Jun 30 – two months' report (initial spread pattern identification and characterization)
- Aug 30 – four months' report (clustering of spread patterns)
- Sep 30 – six months' report (the final report) (prediction modeling of top three patterns)

Willingness to contribute as exemplar/ Driver Project

YES

Critical Resource Requirements

The datasets we've listed are the datasets that we have experiences to work on in a recent COVID-19 project. The availability of those datasets Is very critical for the success of this project as well. Additional

datasets are also be investigated befitting to our goal, provided that we get the exploration permission on ICODA Workbence.

Risks
N/A

If you are working with Co-Principal Investigators (PIs), please add their names and institutions
NONE

Core research team
Rezwana Rafiq, Assistant Project Scientist, Institute of Transportation Studies, UCI
Expertise: transportation data science, activity-travel sequence analysis, human mobility pattern
Tools: STATA and R

Adnan Arefeen: PhD student, CSEE, University of Missouri-Kansas City.
Expertise: machine learning models, data science, high-performance computing.
Tools: Python, R, and Spark.

Sumaiya Tabassum Nimi: Phd student, CSEE, University of Missouri-Kansas City.
Expertise: Deep learning models
Tools: TensorFlow and PyTorch (Python and C++).

Budget Suggestion

(a) Two-month summer salary for PI at the rate of $100,000 per month scale ($11,111 per month)
(b) One postdoc for six months will be supported to conduct the research on pattern analysis and clustering ($4,000 per month).
(c) One full time graduate student support is requested for six months to conduct the research of the model construction for the prediction of pandemic waves ($2,000 per month)

Budget suggested
Total requested budget is **$58,222/-** (the split is shown below)
PI (2 summer month): 22,222/-
One postdoc (6 months): 24,000/-
One graduate student (6 months): 12,000/