MIT Connection Science
the technology of innovation
connection.mit.edu

ENDOR
www.endor.com

COVID-19 Outbreak Week-by-Week Prediction

Dr. Yaniv Altshuler and Prof. Alex Pentland

MIT and Endor.com

Summary

We demonstrate ability to predict which medical centers will show surges in COVID-19 cases up to 14 days in advance. Such capability can be used in order to provide better allocation of critical resources (e.g. testing, personnel) and to allow authorities to enforce local social distancing policies in areas that are about to experience an outbreak.

Outbreak Prediction: Main Findings

1. <u>Covid-19 outbreak locations are surprisingly dynamic:</u> Only 40% of neighborhoods (Census Block Groups) that had an unusual number of residents visiting hospitals at a given week also had an unusual number the following week. Consequently, using current neighborhood conditions to predict future neighborhood conditions gives poor results, so a more comprehensive, regional approach is required.
2. <u>Using regional patterns to predict next-week surges in hospital visits</u>: Using regional pattern detection, the 20% of neighborhoods predicted to be most "at risk" produce 52% of the following weeks' hospital visits. Theoretical best performance is top 20% account for 77% of cases. In addition, the 40% of neighborhoods predicted to be "safest" produce only 6% of the following weeks' visits.
3. <u>Two-week advance prediction</u>: For two-week advance prediction the top 20% most "at risk" neighborhood prediction accounts for 47% of hospital visits.

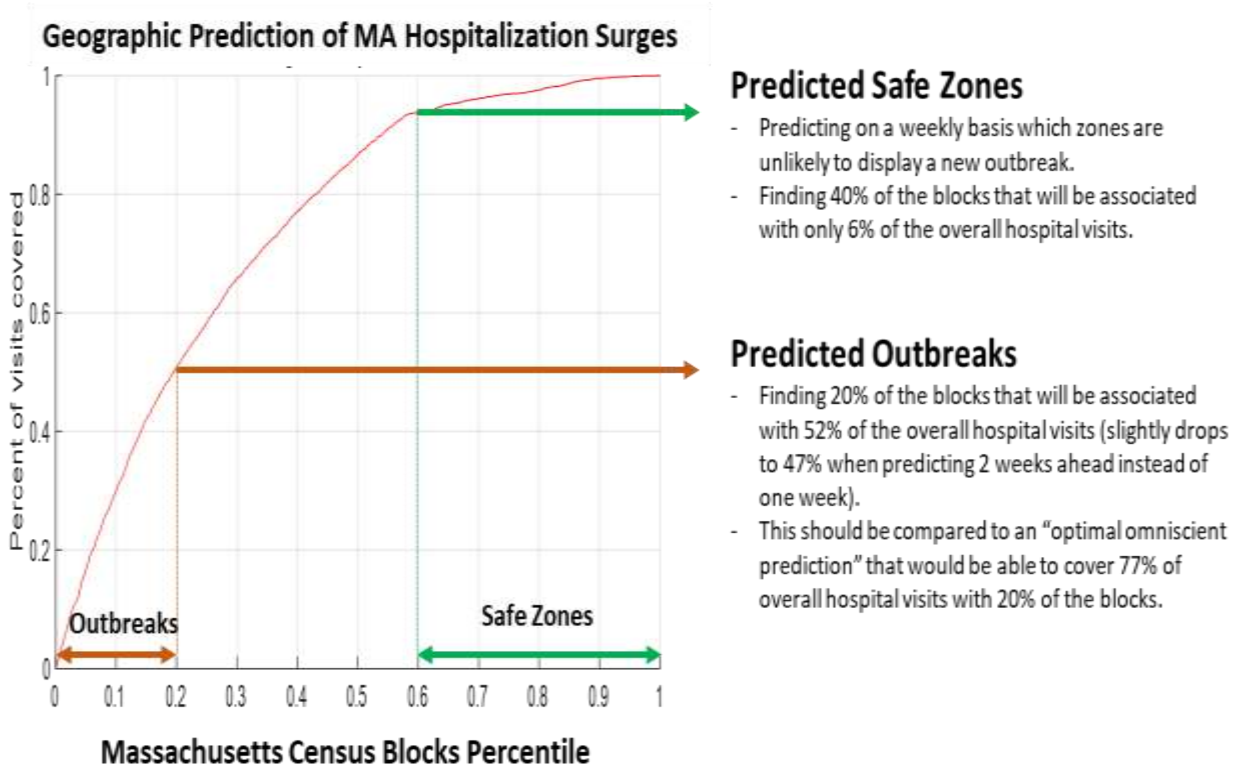

Figure 1. Operating curve for hospital surge prediction based on previous weeks' neighborhood (Census Block Group) data.

Data

This data used for this research is commercially available anonymized mobile device location data that is aggregated to the Census Block Groups ("neighborhoods") in order to preserve users' privacy. Such data has been used for detection of "hotspots" for infection exposure (e.g., grocery stores) in New York City (see https://tinyurl.com/Social-Distancing-NYC). Data covers January 1st to May 1st 2020. Mobility data is derived from users who opted in to share their data anonymously through a GDPR and CCPA compliant framework.

Methodology

This study focused on the state of Massachusetts, encompassing 12,384 Census Block Groups ("neighborhoods") of various sizes and locations. From these, 9,021 neighborhoods had statistically sufficient reported data for use in the prediction task. Neighborhoods without sufficient data mostly have very low population density.

Within this location and mobility data, we have isolated the location of 1,642 hospitals and medical centers located in 233 cities with the state of Massachusetts. For each of these locations we have tracked the number of unique visitors from each neighborhood on a weekly basis, and aggregated these visits in order to measure the total number of weekly hospital visits per neighborhood.

We have then provided the Endor Prediction Engine with a weekly list of neighborhood that have the highest increase in hospital visits in the previous week, and used the Prediction Engine to find predictive behavioral patterns. These patterns were then used to compile a list of the neighborhoods *most likely* to show a surge in expected hospitalization in both the next week and the week after that. In addition, a list of the neighborhoods that are *most unlikely* to show a surge in hospitalizations in the coming 14 days. We have assessed the accuracy of our predictions using location data that was provided following the generation of the predictions.

Background

The idea of finding statistical patterns in large-scale human behavior was classically called Social Physics (https://en.wikipedia.org/wiki/Social_physics). Today mathematical tools inspired by physics can be used to understand the behavior of human crowds. Developed at MIT at the Human Dynamics research group under Prof. Alex "Sandy" Pentland this technique can detect predictive behavioral patterns using very little data, in highly dynamic environments. This statistical technology behind the *Endor Prediction Engine*, that can efficiently analyze dynamic encrypted data-streams and provide accurate forecasts while maintaining data privacy (see https://www.endor.com/social-physics for additional information). Endor is currently employing this technology for finding predictive patterns in encrypted data in order to help governments around the world mitigate the Covid-19 threat.

Research at the MIT *Connection Science* under Prof. Pentland had recently demonstrated the use of data from mobile carriers in order to model the dynamics of the Covid-19 pandemic, demonstrating the effectiveness and impact of Social Distancing in the New York City metropolitan area and locating "hotspots" of likely infection (see details in http://connection.mit.edu/covid19).

Endor

MIT Media Lab spin-off (located in NYC and Tel Aviv, Israel). Serving government agencies worldwide, offering the world's first Federated Learning prediction engine that operates on encrypted datasets. World Economic Forum "*Technological Pioneer*" , Gartner "*Cool Vendor*" (www.endor.com).

Contact: yanival@mit.edu, yaniv@endor.com, pentland@mit.edu