Tang, Francesca; Feng, Yang; Chiheb, Hamza; Fan, Jianqing

The interplay of demographic variables and social distancing scores in deep prediction of U.S. COVID-19 cases. (English) J. Am. Stat. Assoc. 116, No. 534, 492-506 (2021).

Summary: With the severity of the COVID-19 outbreak, we characterize the nature of the growth trajectories of counties in the United States using a novel combination of spectral clustering and the correlation matrix. As the United States and the rest of the world are still suffering from the effects of the virus, the importance of assigning growth membership to counties and understanding the determinants of the growth is increasingly evident. For the two communities (faster versus slower growth trajectories) we cluster the counties into, the average between-group correlation is 88.4% whereas the average within-group correlations are 95.0% and 93.8%. The average growth rate for one group is 0.1589 and 0.1704 for the other, further suggesting that our methodology captures meaningful differences between the nature of the growth across various counties. Subsequently, we select the demographic features that are most statistically significant in distinguishing the communities: number of grocery stores, number of bars, Asian population, White population, median household income, number of people with the bachelor’s degrees, and population density. Lastly, we effectively predict the future growth of a given county with a long short-term memory (LSTM) recurrent neural network using three social distancing scores. The best-performing model achieves a median out-of-sample $R^2$ of 0.6251 for a four-day ahead prediction and we find that the number of communities and social distancing features play an important role in producing a more accurate forecasting. This comprehensive study captures the nature of the counties’ growth in cases at a very micro-level using growth communities, demographic factors, and social distancing performance to help government agencies utilize known information to make appropriate decisions regarding which potential counties to target resources and funding to.

MSC:

92D30 Epidemiology
68T07 Artificial neural networks and deep learning
62H30 Classification and discrimination; cluster analysis (statistical aspects)

Keywords:
bloc model; community detection; COVID-19; learning; neural network; spectral clustering; stochastic machine

Software:
COVID-Net

Full Text: DOI arXiv

References:

[1] Abbe, E., “Community Detection and Stochastic Block Models: Recent Developments,” The Journal of Machine Learning Research, 18, 6446-6531 (2017).

[2] Abbe, E.; Fan, J.; Wang, K.; Zhong, Y., “Entrywise Eigenvector Analysis of Random Matrices With Low Expected Rank,” Annals of Statistics, 48, 1452-1474 (2020) - Zbl 1450.62066

[3] Balakrishnan, S.; Xu, M.; Krishnamurthy, A.; Singh, A., “Noise Thresholds for Spectral Clustering,” Advances in Neural Information Processing Systems, 24, 954-962 (2011)

[4] Betensky, R. A.; Feng, Y., “Accounting for Incomplete Testing in the Estimation of Epidemic Parameters,” International Journal of Epidemiology, 49, 1419-1426 (2020) - doi:10.1093/ije/dyaa116

[5] Brownlees, C.; Gudmundsson, G. S.; Lugosi, G., “Community Detection in Partial Correlation Network Models,” Journal of Business & Economic Statistics (2020)

[6] Chen, J.; Yuan, B., “Detecting Functional Modules in the Yeast Protein-Protein Interaction Network,” Bioinformatics, 22, 2283-2290 (2006) - Zbl 1119.83331 · doi:10.1093/bioinformatics/btl370

[7] Chen, Y., Chi, Y., Fan, J., and Ma, C. (2020), “Spectral Methods for Data Science: A Statistical Perspective,” arXiv preprint
Li, Y.; Ye, F.; Guan, W.; Yang, Y.; Li, F.; and Yuqi Xie, S. L.; Liu, B.; Wang, Z.; Zhang, S.; Wang, Y.; Zhong, N.; He, J., “Modified Seir and AI Prediction of the Epidemics Trend of Covid-19 in China Under Public Health Interventions, Journal of Thoracic Disease, 12, 165-174 (2020)· doi:10.21037/jtd.2020.02.64

[36] Zheng, C., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., Liu, W., and Wang, X. (2020), “Deep Learning-based Detection for Covid-19 From Chest CT Using Weak Label,” medRxiv.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.