Abstract

The COVID-19 pandemic was lived in real-time on social media. In the current project, we use machine learning to explore the relationship between COVID-19 cases and social media activity on Twitter. We were particularly interested in determining if Twitter activity can be used to predict COVID-19 surges. We also were interested in exploring features of social media, such as replies, to determine their promise for understanding the views of individual users. With the prevalence of mis/disinformation on social media, it is critical to develop a deeper and richer un- derstanding of the relationship between social media and realworld events in order to detect and prevent future influence op- erations. In the current work, we explore the relationship be- tween COVID-19 cases and social media activity (on Twitter) in three major United States cities with different geographical and political landscapes. We find that Twitter activity resulted in sta- tistically significant correlations using the Granger causality test, with a lag of one week in all three cities. Similarly, the use of replies, which appear more likely to be generated by individual users, not bots or public relations operations, was also strongly correlated with the number of COVID-19 cases using the Granger causality test. Furthermore, we were able to build prom- ising predictive models for the number of future COVID-19 cases using correlation data to select features for input to our models. In contrast, significant correlations were not identified when comparing the number of COVID-19 cases with mainstream media sources or with a sample of all US COVID-related tweets. We conclude that, even for an international event such as COVID-19, social media tracks closely with local conditions. We also suggest that replies can be a valuable feature within a machine learning task that is attempting to gauge the reactions of individual users.