

Emergency Disaster Framework: Utilizing Textual Analysis to Estimate Daily COVID Patients

Kang, Taeuk (School: Korea International School, Jeju Campus)

Existing methods for COVID-19 informational websites rely on multiple sources that are not clearly labeled which leads to a lack of transparency and trust due to the fragmentation of data between government websites, major media outlets, and unofficial, local sources. Also, in most cases, data is updated in batches, which prevents users (the general public) from accessing the latest data. This may create a large deviation from the actual data and cause more confusion to the general public. However, as COVID-19 spreads quickly across neighborhoods and as test results are usually already delayed, it is critical that the public gets the notice as soon as possible. In Korea, the cell broadcasting system sends a public safety message for live COVID-19 updates. These messages can be collected and parsed to calculate a live, daily estimate through text pattern matching algorithms and natural language processing. By applying this technique using actual data collected from October to November of 2020, on average, the estimates were off by about 6.57 cases from the official data which were reported the day after (percent error of 22.81%). Also, the estimated cases correctly represented the trend of how infections are spreading, allowing the public to recognize the disease's implications around their immediate area. This data can be used to visualize the live statistics on a COVID-19 status board web app for users to explore and determine their daily activity range.