Sentiment Analysis: What Social Media can tell us about COVID19

# Spatio-temporal Sentiment Analysis (Australian Tweets)

Following figure shows spatio-temporal distribution of positive sentiment ratio in tweets related to coronavirusl. The temporal dimension spans from 27 November 2019 to 7 April 2020 discretised by weeks (roughly 17 weeks). Spatial dimension is discretised by Australian Sates and major cities. Tweet user location that does not list city but list country as Australia is categorised as Australia (au). A small portion of tweet users are from foreign cites that we categorise as other (oth).

|  |
| --- |
|  |
| Spatio-temporal distribution of positive sentiment ratio in tweets related to coronavirus |

The figure shows that as soon as COVID19 hit the world, the distribution of positive sentiments in tweets related to coronavirus dropped sharply (from roughly 85% to roughly 48% on average). The percentage stayed there for up to around 12 weeks. Then it gradually changed for three weeks with a very marginal increment. For the final two weeks the increment was a bit more than the previous three weeks.

Possible explanation of the trend can be as follows. As soon as COVID19 hit the world, the online community got shocked by the news. It took some times for world leaders to come up with plans on how to combat COVID19. During this time period (12 weeks) people remained shocked. When the world leads explained their combat plans and ideas, they talked about those positive initiatives by world leaders in this time (three weeks). During the final two weeks, Australian government announced social safely plans, e.g. economic aids to organisations, business and individuals; it announced more strict rules for social distancing and the COVID19 infection curve started flattening. People started to become a bit comfortable and discussed these positive things in their tweets, and therefore number of positive tweets has increased. All these patterns show that by monitoring conversational dynamics on social media, we can can identify the how people are feeling about coronavirus, what initiatives are working or making people comfortable. In the following two sections, we decouple spatial dimension and temporal dimension to see how these observations matches.

# Temporal Sentiment Analysis

The following two figures show temporal distribution of sentiments in tweets related to coronavirus. Another figure following these two figures shows new and cumulative COVID19 cases in Australia by notification date. Time in the first figure is discretised by days and in the second figure by weeks. These two figures show that when COVID19 hit the world on 27 November 2019, there were not many discussion about it in Australian space. Noticeable number of coronavirus related tweets started to be posted after 60 days or around eight weeks, i.e. end of January. Next one week the number increased and then started to fall. The main burst of tweets started after another 30 days or 4 weeks, i.e. end of February. This might be because this time several people in Australia from overseas were identified COVID19 positive. The number exponentially increased for the next 20 days and reached to its peak by the third quarter of March. This exponential increase might have occurred because during this time many Australian were identified COVID19 positive and some of them died. Then it started to fall gradually. This might be because during this time government initiatives and strict social distancing worked and the COVID19 infection rate death from it decreased.

Roughly in any time in the space, among all the coronavirus related posts, only 50% of them were positive. We see two significant drop in the ratio of positive sentiments, one is at the beginning when the world was hit by COVID19 and the next one is by week seven or third quarter of January 2020. During this time period there we not many discussion of coronavirus in Australia. However, it looks the second drop actually triggered the increase in the number of coronavirus related posts. In other words, this second drop alerted the community about the upcoming danger of COVID19. We can assume that the small number of tweets related to coronavirus came from the people who might be Journalists, social workers, health care workers or people who are conscious of health issues.

During the period when noticeable number of posts were related to coronavirus, there are two small drops in the ratio of positive sentiments. One is in week 10 and another is in week 14. Both falls are followed by significant increase in the number of coronavirus related posts. Even though these two drops are small in ratio, but the drop in number of positive tweets were large enough to initiate triggers. It looks, monitoring the ratio of positive sentiment tweets can signal us the trigger in the increase of coronavirus related posts.

A comparison between the third figure and the other two figures show that the number of total coronavirus related tweets is strongly correlated with the number of new COVID19 cases by notification date.

|  |  |
| --- | --- |
|  |  |
| Temporal distribution of sentiments where time is discretised by days. | Temporal distribution of sentiments where time is discretised by weeks. |
|  | |
| New and cumulative COVID19 cases in Australia by Notification date [1] | |

# Spatial Sentiment Analysis

The following two figures show the relation between tweet count and COVID19 cases in states and major cities in Australia. The first figure shows number of tweet counts in states and major cities in Australia. The second figure shows the number of COVID19 cases in states in Australia. The two figures show that there is a strong relation between tweet count and COVID19 cases. The more the number of tweets in a state or in its cipital city, the more the number of COVID19 cases in the state. For example, highest number of coronavirus related tweets are observed in Sydney (syd), the capital city of New South Wells (NSW), and the highest number of COVID19 cases in NSW. The second highest number of coronavirus related tweets are in Melbourne (mel), the capital city of Victoria (VIC) and the third highest number in Victory (vic). The second highest number of COVID19 cases are also in VIC. The same is true for Queensland (QLD). Other cities follow similar pattern with minor order changes. The minor order change might be coming from the fact that we consider capital city and state separately in tweet count but for COVID19 case count the capital city and state are merged together.

|  |
| --- |
|  |
| Number of tweet counts in states and major cities |
|  |
| COVID19 cases in australia by state and source of transmission [1] |

The following figure shows count of coronavirus related tweets and positives tweets in states and major cities. Cities and states that have significant drop in positive tweet count are Sydney (syd), Melborne (mel), Victoria (vic), Queensland (qld) and Canberra (can), Perth (per), Tasmania (tas) and son on in descending order. The order follows the number COVID19 cases order.

Two interesting facts in this figure are observed in (qld, bne) and (nsw, syd) pairs. There is a significant drop in positive tweets in qld but not in bne. The majority of COVID19 cases in Queensland happened in Gold Coast not in Brisbane. There is a significant drop in positive tweet count in syd but not in nsw. The majority of COVID19 cases happened in Sydney not in other parts of New South Wells.

|  |
| --- |
|  |
| Count of coronavirus related tweets and positives tweets in states and major cities |

# Reference

[1] https://www.health.gov.au/news/health-alerts/novel-coronavirus-2019-ncov-health-alert/coronavirus-covid-19-current-situation-and-case-numbers