Dutch General Public Reaction on Governmental COVID-19 Measures and Announcements in Twitter Data

Shihan Wang\textsuperscript{1}, Marijn Schraagen\textsuperscript{1}, Erik Tjong Kim Sang\textsuperscript{2}, and Mehdi Dastani\textsuperscript{1}

\textsuperscript{1} Intelligent Systems, Utrecht University, The Netherlands
\textsuperscript{2} Netherlands eScience Center

Abstract. Public sentiment (the opinion, attitude or feeling that the public expresses) is a factor of interest for government, as it directly influences the implementation of policies. Given the unprecedented nature of the COVID-19 crisis, having an up-to-date representation of public sentiment on governmental measures and announcements is crucial. While the staying-at-home policy makes face-to-face interactions and interviews challenging, analysing real-time Twitter data that reflects public opinion toward policy measures is a cost-effective way to access public sentiment. In this context, we collect streaming data using the Twitter API starting from the COVID-19 outbreak in the Netherlands in February 2020, and track Dutch general public reactions on governmental measures and announcements. We provide temporal analysis of tweet frequency and public sentiment over the past four months. We also identify public attitudes towards the Dutch policy on wearing face masks in a case study. By presenting those preliminary results, we aim to provide visibility into the social media discussions around COVID-19 to the general public, scientists and policy makers. The data collection and analysis will be updated and expanded over time.

Keywords: COVID-19 · sentiment analysis · Twitter data analysis · Dutch policy measures · natural language processing

1 Introduction

Public support is essential for the success of policy measures. Support can be measured from physical behaviours, but also from how people think and talk about these measures, which is known as public sentiment. As public sentiment (the opinion, attitude or feeling that the public expresses) can directly influence the implementation of policies \cite{1}, it is crucial for policy makers to know the public sentiment of chosen policies and to take this sentiment into account when deciding on new policies.

This preprint was last updated on June 12th, 2020.
Given the unprecedented nature of the COVID-19 crisis, having an up-to-date representation of Dutch public sentiment on governmental measures and announcements becomes even more important. However, it is uncertain how public sentiment evolves once the perceived urgency of policy measures changes over time, which is our main research focus in this paper.

The staying-at-home policy makes analysing public sentiment towards specific policies by means of face-to-face research methods like interviews and questionnaires challenging. Meanwhile, about 2.8 million users in the Netherlands use Twitter to share their opinions, making it a valuable platform for tracking and analysing public sentiment. It also allows for much more and frequent measurements and better indicating changes over time in public reactions. Therefore, aiming at understanding the variation of Dutch public sentiment during the COVID-19 outbreak period, we propose to analyse Twitter data using machine learning and natural language processing approaches \[\text{5,7}\]. By analysing real-time data through non-invasive methods, we aim to provide a cost-effective way to access public sentiment towards policy measures in a timely manner.

In this paper we present the preliminary results of our data analysis. First, we collected Dutch Twitter data and filtered the data based on pre-defined COVID-19 related keywords. Second, we analysed the temporal pattern of Dutch public sentiment during the period of COVID-19 epidemic in the Netherlands. Third, we conducted a case study to extract the public attitude towards governmental policy of wearing face masks. By presenting those findings, we expect to provide a sentiment-oriented overview of social media discussions around COVID-19 to the general public, scientists and policy makers.

2 Data Acquisition and Processing

Our main data set consists of Dutch tweets collected by the twiqs.nl service \[\text{6}\]. The analysis of this data is compared with comments from the discussion website Reddit and the Dutch news website nu.nl, where the general public write comments towards certain topics or articles.

2.1 Twitter Data Collection

Twiqs.nl is a service from the Netherlands eScience Center and Surf which collects Dutch tweets and provided aggregated analysis to the research community \[\text{6}\]. The service has been available since 2013 and harvests about 500,000 tweets written in Dutch per day. We use the tweets from February 2020, the first month a COVID-19 patient was found in The Netherlands, and later. We rely on Twitter's lang feature to determine the language that the tweet is written in. Tests with tweet replies indicate that the corpus covers about 55% of the tweets written in Dutch. An overview of the corpus size can be found in Table \[\text{1}\].
| Month         | Number of tweets | Per day | Per hour |
|--------------|------------------|---------|----------|
| February 2020| 14,853,294       | 512,182 | 21,340   |
| March 2020   | 21,190,815       | 683,575 | 28,521   |
| April 2020   | 18,722,608       | 624,087 | 26,004   |
| May 2020     | 18,064,275       | 582,719 | 24,280   |

Table 1. Total number of Dutch tweets collected in the study

| Category | Keyword   | Translation            |
|----------|-----------|------------------------|
| Disease  | corona    |                        |
|          | covid     |                        |
| Health care | huisarts | doctor                |
|          | mondkapje | face mask              |
| Government | rivm    | national health organization |
| Social   | flattenthecurve |          |
|          | blijfthuis | stay home              |
|          | houvol   | hang in there          |

Table 2. Terms used for keyword filtering

| Month         | Number of tweets | Per day | Per hour |
|--------------|------------------|---------|----------|
| February 2020| 279,251          | 9,629   | 401      |
| March 2020   | 3,152,638        | 101,698 | 4243     |
| April 2020   | 2,115,728        | 70,524  | 2939     |
| May 2020     | 1,264,644        | 40,795  | 1700     |

Table 3. Size of corpus used for our analysis

### 2.2 COVID-19 Keyword Filtering

We filtered the general corpus to obtain only tweets that contain disease keywords or related topic words and hashtags. The list of filter words is shown in Table 2 which we selected based on four important categories related to our research focus. Keyword filtering was done case-insensitive and longer words containing one of these keywords as substrings were also selected. The size of the selected corpus can be found in Table 3.

The differences between the numbers of tweets in the various months are striking. We explored the reason of this diversity in the following section 3.1. We also observe some topics drift among our selected tweets, for example in February the discussion was often about China and Wuhan while in March that shifted to the measures (maatregelen) taken by the Dutch government against the pandemic.
3 General Public Reaction and Sentiment

We perform an analysis of COVID-19 related tweets by examining their frequency and their average sentiment. We take the weekly press conference of the Dutch government into account to analyse the correlation between public reaction and government policies over time period.

3.1 Temporal Frequency of COVID-19 Related Tweets

The daily frequencies of Dutch COVID-19 tweets can be found in Figure 1. We define the COVID-19 topic as listed in Table 2 in the previous section: any tweet written in Dutch that contains any of the eight keywords corona, covid, huisarts, mondkapje, rivm, flattenthecurve, blijfthuis and houvol (case-insensitive and possibly as substring). We also found the most popular four keywords (i.e. corona, covid, rivm and mondkapje) and also plotted frequencies of tweet containing one of them respectively.

![Daily frequencies of Dutch tweets with pandemic words](image)

**Fig. 1.** Frequencies of Dutch tweets in the COVID-19 related topics and some of the participating keywords. The most frequent topic keyword is corona. The line of ‘topic’ covers tweets with any of the keywords in Table 2. The general COVID-19 related topics were most popular around the date of a pandemic press conference of the Dutch government (March 12th, 2020).

In Figure 1 the COVID-19 related tweets peak around a press conference of the Dutch government on Thursday March 12th 2020 (in which the government announced the first lockdown measures to stop spread of corona virus in the Netherlands) and reach the top after March 15th 2020 (when the prime minister
addressed the nation about the corona virus, the first national crisis address since the oil crisis in 1973). After that, the COVID-19 related tweets keep decreasing in frequency. As noted in Table 3, the number of tweets containing the selected keywords decreases over time. There are two possible reasons for the decrease, either the topic became less popular or people continue to talk about COVID-19 in tweets but without using the keywords in Table 2, such as corona or covid.

To check these hypotheses a similar analysis is performed on data from Reddit. This social network is organized in a large number of high-level topics called subreddits. Users can post a message in a subreddit in order to start a thread, and other users can post comments in reply to either the original message or an existing comment. There are a large number of threads related to COVID-19 on reddit, most of which receive only a few comments. There are however some larger threads with many active participants that extend over a large period of time. One of these is a thread called “Megathread Coronavirus COVID-19 in Nederland” in the subreddit r/thenetherlands, which has been active since March 2020. This thread has been used in a comparison analysis.

![Daily frequency of “Megathread Coronavirus COVID-19 in Nederland”](image)

**Fig. 2.** Comment frequency on an example Reddit thread

Figure 2 shows the frequency of comments per day for this thread, which is consistent with the Twitter data. In this case the decrease may be caused either by people finding other threads to post comments, or a general loss of interest in the topic. Further analysis shows that the relative frequency of keywords in this thread remains rather stable, as shown in Figure 3. Because all the messages in the thread are about COVID-19 by default, the relative frequencies are indicative of the way people talk about the topic. This data combined with the tweet analysis therefore seems to support the hypothesis that the topic is actually less
Fig. 3. Weekly average of keyword frequency on an example Reddit thread.

popular in June 2020 compared to the preceding period, and the hypothesis that people use different words to discuss the topic is not supported.

Interestingly, Figure 3 also shows that a large majority of the messages in the Reddit thread did not use any of the four popular keywords in Figure 1 while all messages are about COVID-19. One explanation for this may be that within long comment threads the messages may become less self-contained than, e.g., a new tweet, which could influence the usage of keywords. However, it is also possible that the Twitter analysis suffers from a large false negative rate of COVID-19 related tweets that are not selected by the keyword filter, even in the peak period around March 15th where keyword frequencies are the highest.

3.2 Temporal Sentiment of COVID-19 Related Tweets

We used the sentiment module of pattern [4] to automatically assign a sentiment score to each tweet. The scores are based on pattern’s Dutch sentiment lexicon which contains 3918 words, mostly adjectives. Two-thirds of the tweets were assigned a non-zero score. An overview of the average sentiment on Dutch Twitter can be found in Figure 4.

In Figure 4 the y-axis represents the sentiment score calculated by our sentiment analysis approach, where a higher sentiment score means more positive attitude. We can observe that the average sentiment score of Dutch tweets related to COVID-19 was always lower than the general sentiment, which indicates that, compared with general topics discussed in Twitter, the public is more negative towards COVID-19 related topics. Also, some interesting links can found between press conferences and trends of public sentiment. For instance, sentiment reached a low peak point on the date of the third press conference of the Dutch
government about the pandemic, March 12th, when the first lockdown measures were announced. Most recently, the COVID-19 related sentiment reached a high peak point on the date of May 19th, when the government announced the first release measures. According those findings, we think the trend of average daily sentiment shown in Twitter can be influenced by governmental measures and announcements. However, further analysis is needed to explore the unclear correlation between some press conferences and public sentiment. For instance, check whether tweets with a strong negative or positive sentiment represent particular subtopics, either using a keyword analysis or by investigating the general context of such tweets.

In addition, we studied the temporal pattern of sentiment for all Dutch tweets on March 12th in detail to investigate the short-term influence of governmental announcements. We found that the impact of the press conference is visible in the evolution of the overall public sentiment measured per hour. As shown in Figure 5, the general public sentiment is relatively constant during the day but suddenly drops around the time of the press conference. The result indicates that the reaction of public sentiment towards governmental measures and announcements can be captured in a very short time period from Twitter data.

4 Case Study: Public Attitude towards Specific Policies

Besides the above analysis of sentiment analysis in broad topics, we are also interested in public opinion towards certain topics (e.g. a specific policy measure), known as stance analysis [3]. While development of automatic stance identification methods is ongoing, we start by manually indicating stance on a sample of
tweets. Due to time limitations we focus on the specific Dutch policy measure on wearing a face mask (Dutch: *mondkapjes*), and only annotated one week of tweets related to this topic. In this case study, the analysis is tailored towards the policy question *What is the opinion about the April 2020 RIVM policy on advising against the use of face masks by the general public?*. Tweets are annotated manually by our researchers with one of four labels: agree, neutral, reject and irrelevant, where the fourth label indicated out-of-topic tweets. The result of the analysis can be found in Figure 6.

![Sentiment scores of Dutch data on Thursday 12 March 2020](image)

**Fig. 5.** Average daily sentiment of all Dutch tweets on the date of the third Dutch pandemic press conference about the first lockdown measures (March 12th, 2020).

We found that most of the tweets (90%) rejected the policy on advising against the use of face masks by the general public. Two examples of frequent opinions (translated in English) are as follows:

![Public attitudes towards the official face mask policy measures based on data from Twitter (left) and Nu.nl (right)](image)

**Fig. 6.** The public attitudes towards the official face mask policy measures based on data from Twitter (left) and Nu.nl (right).
Face masks are useless: corona droplets do not travel more than 1.5m. Is this a scientific conclusion of the RIVM? Or is it a fake argument because of shortages? 2008 RIVM scientist says that face masks reduce infections and have effect.

Broadly I agree. But with respect to face masks I refer to scientists in our neighboring countries. They have another opinion than the RIVM. So yes... who is right?

However, according to our previous experience on Twitter data analysis, we noticed that Twitter contains a large amount of negativity about the government. Therefore, in order to validate our analysis on this specific topic, we conducted a similar analysis using data from a Dutch news website named Nu.nl. In this website, the general public is allowed to write comments under news articles and reply to each other. We selected the first news article introducing this policy (i.e. advising against the use of face masks by the general public) and measured public stance on all comments of the article. Although the data collection time of Nu.nl is slightly different from that of Twitter, the comparison is reasonable as the governmental measure kept the same over this time period. We found the majority of commenters on Nu.nl also rejected the policy, however the differences between support and rejection were a lot smaller. A possible reason of this phenomenon is that the comments on Nu.nl are actively moderated while tweets are not. These findings, while validating the results on Twitter data, also indicate the importance of capturing the public reactions across different sources.

5 Discussion and Future Work

We present a first analysis of Dutch tweets related to the COVID-19 pandemic. We concentrate on online public reactions towards the measures taken by the government against the spread of the corona virus. Our analysis shows that the pandemic generated large interest on Dutch Twitter with tens of thousand tweets on the topic each day. The sentiment of the COVID-19 related topic tended to be more negative than that of Dutch tweets on average. We also found that national press conferences about governmental measures and announcements (e.g. the first national press conference about the lockdown policy) can influence public sentiment, and furthermore the impact on tweet sentiment can be captured in a very short time period. However, we did not find the direct correlation between all press conferences and trends of public sentiment. We thus would like to investigate the influence of governmental measures and announcements in more detailed analysis. In future work, we also see the potential to explain temporal trends of the public sentiment by linking it to daily numbers of infections and deaths in the Netherlands.

In a case study, we assessed the stance of Dutch tweets on the national advice not to wear face masks. The analysis shows that people largely reacted negatively to this advice. Since the position of the government on this topic has changed recently, it would be interesting to continue the analysis for later dates and study the temporal evolution of public sentiment against changing policies. Moreover,
the current manual annotation on public stance is very time consuming, so we would like to develop machine learning approaches to automatically perform the analysis. Additionally we are interested in performing stance analysis regarding other policy measures, like compulsory school closing, social distancing and infection testing. In order to be able to tackle these with machine learning, some initial manual annotation needs to be done for each of these new topics. Active learning [2] could be an interesting approach for this task. This method identifies the parts of the data for which annotation would be most beneficial to the machine learner and thus limits the effort required for performing the manual annotation process.

References

1. Burstein, P.: The impact of public opinion on public policy: A review and an agenda. Political research quarterly 56(1), 29–40 (2003)
2. Dasgupta, S.: Two faces of active learning. Theoretical Computer Science 412, 1756–1766 (2011)
3. Kunneman, F., Lambooij, M., Wong, A., Van Den Bosch, A., Mollema, L.: Monitoring stance towards vaccination in twitter messages. BMC Medical Informatics and Decision Making 20(1), 1–14 (2020)
4. Smedt, T.D., Daelemans, W.: Pattern for Python. Journal of Machine Learning Research pp. 2063–2067 (2012)
5. Tan, S., Li, Y., Sun, H., Guan, Z., Yan, X., Bu, J., Chen, C., He, X.: Interpreting the public sentiment variations on Twitter. IEEE transactions on knowledge and data engineering 26(5), 1158–1170 (2013)
6. Tjong Kim Sang, E., van den Bosch, A.: Dealing with big data: The case of Twitter. Computational Linguistics in the Netherlands Journal 3, 121–134 (2013)
7. Wang, S., Terano, T.: Detecting rumor patterns in streaming social media. In: 2015 IEEE International Conference on Big Data (Big Data). pp. 2709–2715. IEEE (2015)
## A Press conference topics

| Date                      | Measures                                                                                                                                                                                                 |
|---------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sunday 1 March 2020       | People with respiratory issues should stay at home                                                                                                                                                    |
| Monday 9 March 2020       | Wash hands regularly. Sneeze in elbow. Use paper handkerchiefs. Stop shaking hands. Brabant only: work at home if possible. Spread working hours.                                                       |
| Thursday 12 March 2020    | No events with more than 100 people. Museums and theaters close. Sporting events cancelled. Work at home when possible. Vulnerable people avoid social contact.                                          |
| Sunday 15 March 2020      | Schools close. Bars, restaurants and sport clubs close. Keep a distance of 1.5 metres from each other.                                                                                                    |
| Tuesday 17 March 2020     | Government will provide financial aid for small and large corporations and for hospitals.                                                                                                               |
| Thursday 19 March 2020    | No visits allowed in carehomes.                                                                                                                                                                          |
| Monday 23 March 2020      | No meetings of more than two people in the public space punishable by fines. Whole household self-quarantaines if one member has a fever.                                                                 |
| Tuesday 31 March 2020     | Measure deadline extended to 28 April (was 6 April).                                                                                                                                                     |
| Tuesday 7 April 2020      | Government thinks about a contact tracing app.                                                                                                                                                         |
| Wednesday 15 April 2020   | No new measures.                                                                                                                                                                                        |
| Tuesday 21 April 2020     | Primary schools partially reopen 11 May. Children may do sport training outside from 28 April. Events before 1 September are cancelled.                                                               |
| Wednesday 29 April 2020   | No new measures.                                                                                                                                                                                        |
| Wednesday 6 May 2020      | Avoid crowded places. Testing will be available to everybody from 1 June. Road map announced: 11 May: contact professions (for example barbers and driving instructors) may start working. Grownups may start sporting outside. Libraries reopen on. 1 June: Compulsory wearing of face masks in public transport. High schools reopen partially. Bars, restaurants, museums and cinemas reopen on: max 30 guests inside, no limit outside. 1 July: limit is raised to 100 people inside, also for churches and weddings. Public toilets reopen. 1 September: Sporting inside is allowed. |
| Wednesday 13 May 2020     | No new measures.                                                                                                                             |
| Tuesday 19 May 2020       | Primary schools reopen completely on 8 June. Partial reopening of universities on 15 June. There will be contact tracing of positive cases. Visitors allowed in care homes from 15 June.                        |
| Wednesday 27 May 2020     | All measures of 1 September move to 1 July. Spread holiday times. Only travel to corona-safe countries.                                                |