Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Xinyi Hu March 31, 2021

Open-domain Sentiment-based Event Detection for COVID-19

by

Xinyi Hu

Adam Glynn

Adviser

Roberto Franzosi

Adviser

Department of Quantitative Theory & Methods

Adam Glynn

Adviser

Roberto Franzosi

Adviser

Daniel Sinykin

Committee Member

2021

Open-domain Sentiment-based Event Detection for COVID-19

By

Xinyi Hu

Adam Glynn

Adviser

Roberto Franzosi

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Quantitative Theory & Methods

2021

Abstract

Open-domain Sentiment-based Event Detection for COVID-19

By Xinyi Hu

As one of the most popular social media platforms in recent years, Twitter has provided a database containing abundant information reflecting the public's reactions to various events and discussions. Many sociological researchers and news agencies have accustomed to collecting and processing Twitter data to achieve opinion-mining or detect significant social events. The importance of event detection has become even more remarkable during the special time of the global pandemic because it's crucial to keep the public informed timely about social subjects like change of policy and disease prevention strategies.

The main goal of this research is extracting major social events occurring in the bud stage of the coronavirus in the United States. The major focus of this research is to carefully examine whether sentiment-based event detection can be successfully implemented when the focal event is essentially negative. In this case, the pandemic is a worldwide public health emergency, which results in a large bias on the emotion polarity of tweets. This study employs a data set that covers more than a million English tweets that contains keywords about Covid-19 posted in a month span. Sentiment analysis tools such as Stanford CoreNLP and Hedonometer calculates the emotion score of tweets, enabling the researcher to apply mathematical models that define emotion spike to determine whether an event has occurred on certain day. In addition, after discovering that sentiment-based event detection, especially with Stanford CoreNLP, can efficiently discerns hot spot occurrences, this research utilizes Topic Modeling and NER (named-entity recognition) to draw out words and phrases to help summarize the possible social events.

Open-domain Sentiment-based Event Detection for COVID-19

By

Xinyi Hu

Adam Glynn

Adviser

Roberto Franzosi

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Quantitative Theory & Methods

2021

Acknowledgements

My thanks to Dr. Roberto Franzosi and Dr. Adam Glynn, my advisors, who provided great help to my research in my senior year and made this project experience truly rewarding. In addition, I want to thank Dr. Sinykin to inspire me with many great works of NLP in his class QTM 340, which sets the foundation for my study. Besides these great mentors, Zhengyang Qi and Zhengzhe Yang were absolutely supportive throughout this project and offered me guidance when I encountered technical difficulties, for which I am grateful.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As one of the most popular social media platforms in recent years, Twitter has provided a database containing posts that document the public's reactions and discussions of various events. Many sociological researchers and news agencies have accustomed to collecting and processing Twitter data to achieve opinion-mining or detect significant social events (Section 2.1). The importance of event detection has become even more remarkable during the special period of the pandemic because it's crucial to keep the public informed about urgent social subjects like high risk places, change of policy in testing and vaccines, and novel prevention or treatment methods.

The main goal of this research is applying sentiment-based event detection to extract major social occurrences during the impending stage of the coronavirus in the United States. Information from social media platforms like Twitter can be overwhelming and thus hinders people from having an

overview of the content. This event-detection model alleviates the need for close-reading. The major focus of this research is to carefully examine whether sentiment-based event detection can be successfully implemented when the focal event is essentially negative. In this case, the pandemic is a worldwide public health emergency, which results in a large bias on the emotion polarity of tweets. A major advantage of the pipeline built by this research is that the event-detection is open-domain, which bypasses the restriction on topics imposed by the popular closed-domain event detection and enables the user to simply select the token of interest and achieve event identification from social media posts, without any other domain specification [1].

This study employs a dataset containing more than a million English tweets having keywords about Covid-19 posted in a twelve-day span (March 1st to March 12th, 2020). Sentiment analysis tools such as Stanford CoreNLP and Hedonometer calculate the emotion score of tweets, enabling the researcher to apply mathematical models that define emotion spike [12] to determine whether an event has occurred on certain day. In addition, this research utilizes Topic Modeling and Named-entity Recognition (NER) to draw out words and phrases to help summarize the possible social events.

To begin with, Chapter 2 will describe the current use of social media

resources as the corpora for event detection, the impetus of choosing social media posts as the database, the advantages and drawbacks of the state-of-the-art event detection algorithms, the approaches of using topic clustering to extract words for event description, and the restriction of locations in certain event-detection applications. Any foundation work prior and related to this dataset will be discussed. A comparison between Covid-themed tweets with other datasets will be presented as well to show its distinctness.

Chapter 3 will give examples and statistics of the Covid-themed tweets dataset which prove the proposed Twitter data can serve as a comprehensive event-detection research resource. This chapter also discusses the rationale for the data collection method, the process of preliminary filtering of the dataset, and the quality assurance procedure, which all serve to substantiate the dataset's validity.

Chapter 4 will explain different approaches (Sentiment Analysis with Stanford CoreNLP, Sentiment Analysis with Hedonometer, Topic Modeling with Gensim, and Named-entity Recognition) and specific adjustments of the chosen methods.

Chapter 5 will report the results for all approaches attempted: sentiment analysis by both Stanford CoreNLP and Hedonometer on all tweets and a

subset of tweets filtered by location, topic modeling with Gensim on the highlighted days' tweets and the news articles referred in tweets, and NER of the tweets and news articles. The results, including the dates extracted by the sentiment-based event detection model to label possible events and the key words corresponding to the events, will be presented in figures, charts and tables to visually demonstrate Twitter's potential as an open-domain event detection dataset.

Chapter 6 wraps up all the contributions I have made to the research community in NLP and the field of event detection in particular. Future paths on this dataset will also be suggested for researchers who share mutual interests.

Chapter 2

Background

2.1 Social Media Data Mining for Event Detection

Social media has long been utilized by researchers for event extraction because of its data size and lack of restrictions. Researchers have successfully implemented collective action from social media (CASM) on Sina Weibo, a popular Chinese social platform to extract more than 100,000 social events in a seven-year time span, and even some detected social occurrences are never reported by the mainstream news channels because of governmental censoring [20]. While the government is ramping up the surveillance of the public's sharing and exchange of information, a major characteristic of social media data stream is that people post so much content every second that the real-time supervision and censoring towards social media are much more

difficult than those of TV reports and newspaper.

2.2 Sentiment-based Event Detection with Token and Hashtag

Prior work mostly employ frequency-based event detection, specifically *Token Spikes* method, which calculates the frequency of a token appearing in social media posts in a time series and compare to see whether certain day's frequency exceeds the value of the threshold times the average frequency of the previous days [17]. For instance, Cataldi et.al noticed that normally, the frequency of the word "earthquake" appearing in tweets maintains at a stable and low level (the number of tweets containing "earthquake" usually makes up less than 0.5% of all tweets). However, when the catastrophic earthquake happened in Haiti on January 12th, 2010, the frequency escalated to about 2% and stayed at this level from Jan 12th to 14th, 2010. This peak helps the researcher to make a conjecture that a major event must have happened, causing people to suddenly use the word "earthquake" more frequently [2].

Building on the frequency-based event detection, Paltoglou proposed a refined sentiment-based detection which is able to achieve higher accuracy than frequency-based method [12]. For example, Paltoglou attempted to detect the

event of USA Census 2010 with both frequency-based and sentiment-based methods, and he discovered that frequency-based method would fail because the increase in the number of posts discussing this event was too insignificant to be captured by the model, whereas sentiment-based method successfully discerned the upcoming census, despite the fact that not a lot of posts talked about it [12]. Paltoglou's work relies on on the premise that the public would have collective emotion changes when a major event happens, and such sudden increase or decrease of sentiment is quantifiable and thus detectable by math formulas [12].

The research exploited a "ternary classification scheme" in which each tweet can be positive, neutral, or negative [12]. Then, the frequency of the positive/negative tweets of each day are put into the math formula below, a negative spike, which signifies a negative social occurrence is detected on day **d** in a time frame of **n** days if

$$\mathbf{negFreq}(\mathbf{d}) \geq threshold * \mathbf{avgNegFreq}(\mathbf{d-1}, \mathbf{n})$$

where

$$\mathbf{avgNegFreq}(\mathbf{d}, \mathbf{n}) = \frac{1}{n+1} \sum_{i=d-n}^{n} \mathbf{avgNegFreq}(\mathbf{i}) \text{ [12]}$$

In plain words, a negative event will be detected if the frequency of negative tweets on day **d** is so large that it exceeds a bar set by previous studies, which is the average frequency of negative tweets on days prior to day **d** times a threshold of 2 [12]. The same formula applies for positive spike detection, except the NegFreq values are replaced by PosFreq values in the equation above [12].

Here is an example of using the formula presented above to detect major social events: Paltoglou noticed that for tweets containing "tigerwoords", "tiger", and "woods", the frequency of negative tweets on February 19th 2010 was so high that the sentiment-based model flagged this date [12]. He concluded that this negative spike was caused by the public apology made by Tiger Woods on Feb 19th 2010 [12].

Categorizing the emotion of tweets into polarity significantly accelerates the model's classification process, because there are various available sentiment analysis tools that already adopt the three-fold classification system, such as Stanford CoreNLP, Valence Aware Dictionary for Sentiment Reasoning (VADER), SentiWordNet, and Hedonometer. Though ANEW (Affective Norms for English Words) is also a powerful sentiment analysis tool that evaluates words' valence, excitement, and level of control, it does not employ

a classification system that would categories tweets into positive, negative, or neutral. Thus, this study does not consider using ANEW to compute the sentiment scores of tweets.

Since researchers have been increasingly applying sentiment analysis tools with the ternary categorization scheme, several studies have examined and cross-compared the performance of multiple state-of-the-art SA methods. After testing the tools with corpora of various genre, researchers discovered that the level of performance depends on the specific corpus [14]. Reagan and his fellow researchers concluded that ANEW always had the least satisfying accuracy, regardless of the type of the corpora, while other tools do not display a distinct advantage because all of them perform poorly on certain texts, but better accuracy is achievable when the input corpus is in other genre [14].

In sentiment analysis, we will compute the relative frequency of positive and negative tweets, rather than their absolute value. Utilizing the relative frequency rather than the absolute count of tweets ensures that the efficacy of the detection model is not affected by the varying amount of social media posts on different days [12].

Paltoglou tests the detection models on ten major social events of multiple genres, like movie awards, train collision, and public figure's apology [12].

He compared the accuracy of Frequency-based method (*Token Spikes* and *BNgram*) and sentiment-based method, and eventually found out that the models' performances are dependent on the size of the dataset (how many tweets are employed) [12]. However, when we make datasize as the control variable, the two methods' success rate depends on the length of time frame we allow for the methods to detect events: if we define "success recall (of event)" as the method capable of detecting an event on the exact day when the event takes place, sentiment-based event detection model can recall 60% of targeted events, while frequency-based method can only recall 20% of events; If we define "success recall (of event)" as the method capable of detecting an event on the day when the event occurs or the next day, sentiment-based method can recall 90% of targeted events, whereas frequency-based method can only recall 60% of the events [12].

In addition, Paltoglou utilized social media posts with various quantity to test whether the accuracy rate is related to the size of the dataset, and he found out that frequency-based detection model would show an increase in accuracy when the dataset increases to about 100k, and the accuracy rate stops increasing after hitting the threshold [12]. On the other hand, the performance of sentiment-based detection continues to improve until the dataset

reaches 750k, which indicates that feeding more data to sentiment-based model can help to yield better result, while frequency-based model receives no benefit from additional data [12].

2.3 Topic Clustering on Social Media Posts

Researchers have been attempting to achieve event detection with standard latent Dirichlet allocation (LDA) models [19]. Vavliakis and fellow researchers discovered that topic modeling and Named-entity recognition (NER) can be successfully applied to extract major social events from Web Social Media, specifically, billions of blog posts and data from an event detection data contest [19]. They determined that LDA models provide accurate and relatively complete representations of their corpora [19]. Moreover, NER, a process of the machine categorizing nouns as persons' name, location, organization, etc., also helps to identify the words related to the particular event, because NER efficiently draws out the nouns that are associated with the actions, such as when, who, where, and what [19].

2.4 Location-specific Event Detection

Researchers have experimented with event detection when placing a restriction on the events' place of occurrences [18, 5]. Feng and fellow researchers presented their study of taking posts from Sina Weibo as raw data and utilzing Part-of-Speech tagging as an effective location-specification method which clusters Sina Weibo posts that have mentioned similar places into groups for higher-accuracy event detection [5]. Unankard proposed a novel understanding of certain social occurrences: "hotspot events", because they receive an intense popularity in limited locations. He proposed a Location Sensitive Emerging Event Detection (LSED) model that helps researchers to quickly discern social events, especially unexpected social emergencies [18]. The LSED model takes social media posts (especially short texts from micro-blogs) as the input, use both NER and POStagging to find nouns of location names, and apply keyword co-occurrence search to describe the events [18].

2.5 Sentiment-based Event Detection for Tokens with Negative Connotation

One of the major challenges of employing sentiment-based event detection is determining the definition of good and bad events, which can undermine the accuracy of sentiment spike detection algorithm if people perceive the same event differently.

Moutidis and Williams have mentioned that researchers should be cautious about defining positive or negative events, because the sentiment associated with many social occurrences are created by the media or the government and then internalized by the public [11]. They hypothesized that some sentiment polarity of event might be products of social construct and they validated that the media did have an impact on the public's perception of certain social events [11]. Nevertheless, some events are fundamentally "bad", such as the COVID pandemic. Ever since the outbreak becomes a global emergency, there have been more than 120 millions of infected people around the world and over 2.65 million cases of death (all figures are collected up until March 15th, 2021). The situation is urgent across the world. The U.S. has 29.5M people testing positive and 535k death cases. In China, though the pandemic is now under control after months of quarantine and frequent testing, there still have

been more than 90k cases of infections and about 4.6k deaths. For UK, there are 4.26M positive cases and 126k deaths. In Europe, there have been more than 36M reported infections and 984k deaths. The pandemic also halts many nations' economic growth and causes public health crisis. Therefore, covid-19 and relevant keywords carry fundamentally-negative meanings. However, there has been limited work on exploring whether sentiment-based event detection can be applied to events with negative context, which inspires this article about covid-related event detection using sentiment spike method.

Chapter 3

The Corpus

In this chapter, the generation of Covid-themed tweets dataset will be discussed in details. The source of data (Section 3.1), the mechanism and word choice for tweet scraping ((Section 3.1), the rationale for choosing data produced in the twelve-day span (Section 3.2), the preliminary filtering process (Section 3.3), and string removal (Section 3.4) will be elaborated to demonstrate our dataset's integrity. To ensure the quality of the data, we additionally apply quality assurance procedures (Section 3.5) with a hope to convince readers that Covid-themed Tweets Dataset could serve as a valid and rich event detection research resource in NLP community.

3.1 Twitter Dataset

As shown in Figure 3.1, twelve days' tweets (in total 1,160,591 tweets) are scraped via the Twitter API using keyword search.

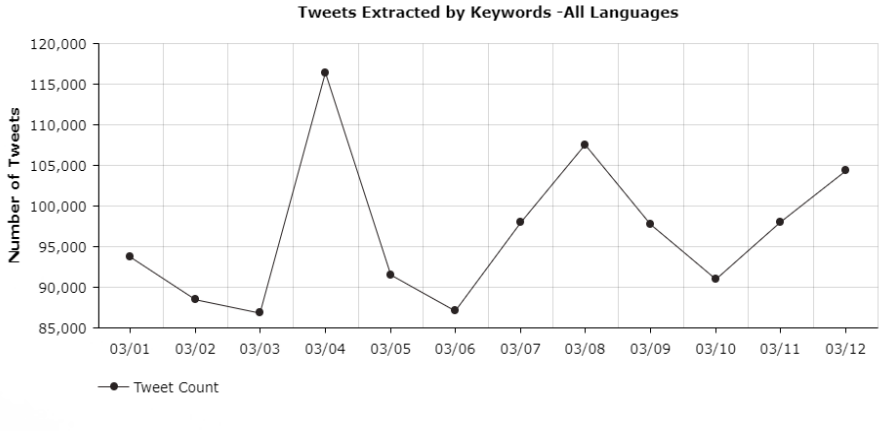

Figure 3.1: Number of Covid-19-related tweets collected by keyword search, posted from March 1st to March 12th, 2020, without any language filtering.

To ensure the dataset covers as many coronavirus-related posts as possible, we manually read many tweets discussing the pandemic and collect relevant words that frequently show up in tweets under this topic. As a result, multiple keywords are included in the keyword search process: *coronavirus, koronavirus, corona, wuhancoronavirus, wuhanvirus, kungflu, epidemic, covid-19, covid19, corona virus, covid, chinavirus*, and *pandemic*.

Many prior studies rely on hashtag extraction to collect tweets containing the token of interest [12]. To explain the difference between hashtag extraction and keyword extraction, we would first need to understand the function of hashtags in Twitter. The hashtag symbol "#" is put before words or phrases to help Twitter users indicate the relevant topics of their tweets. Additionally,

when people are interested in certain subject, they can simply tap on the hashtagged word or phrase to see all other tweets that have used the same hashtag.

Here is an example of event detection using hashtag search: Paltoglou notices that people will use the hashtag "#oscars2010" to discuss related matters, because it's an annual event and the use of hashtags indicates that the event is not only attractive to the public but more importantly, relatively rare [12]. However, since hashtags are usually used to highlight sudden heated topics, as the epidemic escalates in the U.S., people are less likely to use hashtags when discussing Covid-19-related subjects. In this context, hashtag extraction method increases the risk of undersized sampling. Therefore, to avoid any bias caused by the normalization of the pandemic, this study employs keyword search rather than hashtag extraction. In addition, some chosen keywords are in their misspelled form (e.g., *koronavirus*) or conveying racist connotations (e.g., *kungflu* and *chinavirus*). Although these words rarely appear in mainstream news reports, they are frequently used by social media especially in the early stage of coronavirus pandemic. The use of these racist words has become even more prevalent after President Trump blames China for not restricting international flights and eventu-

ally infecting the whole world, inciting the ethnic hatred towards Asians, especially Chinese. Thus, this study incorporates them into the keyword search to improve the completeness of data collection. On average, about one hundred thousand tweets are scraped for each day in this twelve-day span.

3.2 Choice of Time Span

This study concentrates on detecting and extracting possible social events, especially the unexpected public emergencies, when the Covid-19 outbreak is still incipient in the United States, because many later studies have shown that if the U.S. government and the public have taken faster response to the disease, many infection and death cases could have been avoided [15]. Nonetheless, determining which time period should be classified as the early stage of the pandemic is a challenging task for this study. On March 13th, 2020, then-president Donald J. Trump made an announcement to declare Covid-19 pandemic as a national emergency and passed the travel restriction between the U.S. and Europe, which marks that the U.S. government has officially reached a consensus with C.D.C that Covid-19 is a public health crisis and implemented several policies to prevent further infections [4]. Therefore,

this study takes this president's declaration as a time stamp where the Covid outbreak ends its budding stage in the United States.

3.3 Language Selecting and Filtering

This data collected from keyword search first goes through a language filtering step, as shown in Figure 3.2. In the Twitter database, there's an item called "lang" (short for "language") for each tweet that indicates the primary language that the tweet is written in. For English tweets, they have "en" for "lang". To reduce noise caused by words appearing in different forms of multiple languages, this study filters the collected tweets by language, only keeping those written in English. The number of tweets kept in the final dataset is shown in Figure 3.3. Statistics in Figure 3.3 confirm that the filtered dataset still provides an enormous amount of social media data, ranging from 54k tweets/day to 86k tweets/day.

As shown in Figure 3.3, there are some striking increases of tweet count on March 4th, March 8th, and March 12th. Although our research focuses on applying sentiment-based event detection, intuitively, these sudden increase in the number of tweets also indicate possible COVID events. Thus, we add

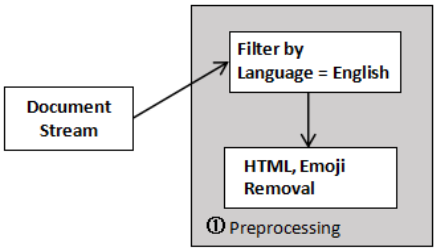

Figure 3.2: Proposed Methodology of Preliminary Data Filtering.

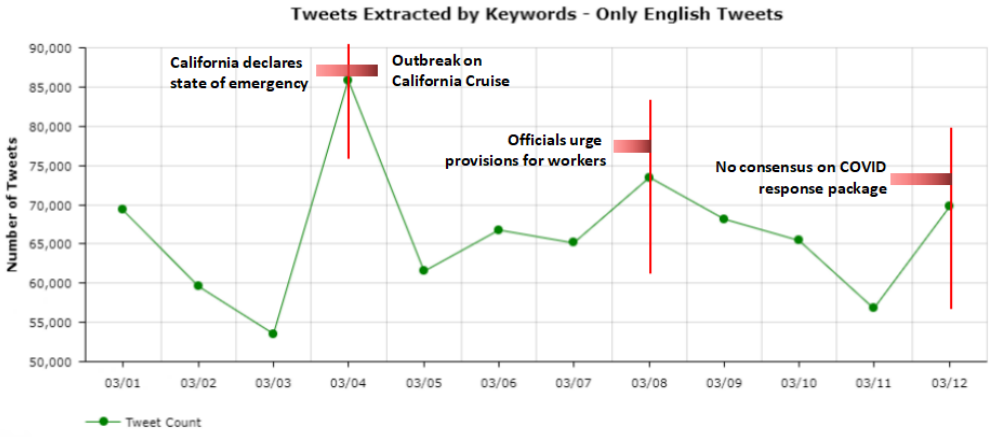

Figure 3.3: Number of Covid-19-related English tweets posted from March 1st to March 12th, 2020.

vertical lines and summaries of events for the three dates in Figure 3.3, and the specific events are presented in Table 3.1. This step is performed for references in later work where we test if the sentiment-based detection can discern these events.

Date	Events
March 4th	1. California governor declares state of emergency over coronavirus 2. 11 passengers and 10 crew members on California cruise ship are "symptomatic"
March 8th	Pelosi and Schumer urge Trump to include paid sick leave and other provisions in any proposal to confront coronavirus impact
Marth 12th	No deal reached tonight on coronavirus response package

Table 3.1: Major COVID Events on March 4th, 8th, and 12th

3.4 Particular Content Removal

In addition, within these tweets' content, there are often some HTML links. Because URL links provide little help to the upcoming sentiment analysis and topic modeling, and may even cause topic modeling to produce clustering with no substantive meaning, we decide to remove them. In Twitter database, every tweet has an item called "url_list" which contains the HTML link mentioned in the tweet. We use what's displayed in the "usl_list" to remove the strings from the tweets' textual content. Using Python script (NLTK package), we successfully remove HTML links from the tweets' full text.

We perform an additional step where we examine if the website addresses lead to news reports by testing if the the string in "usl_list" fit into certain major news websites' URL format. Firstly, we need to determine what news

websites that will be included in this study as our "major news websites". We look through multiple online rankings and eventually select 13 credible news websites, which are CNN, New York Times, FOX, Guardian, Bloomberg, Wall Street Journal, The Epoch Times, Washington Post, BBC, NBC, White House, ABC, and Yahoo News.

After examining the 13 news websites, we find out that they always have a fixed format for the URL links of articles posted on their sites. For instance, CNN news articles always have "www.cnn.com" at the beginning of their URL links. If they pass this filtering process, these HTML links of news articles are stored in a separate dataset where they are grouped based on their source website for further news article scraping. We also discover that among all the HTML links included in tweets, links that lead to news report make up a large proportion. For instance, on March 5th, there are 17,641 links appearing in covid-related tweets, and 3,302 (18.7%) of them lead to the selected 13 news websites.

When we download the tweets data from Twitter API, they come in as json files, which contain data objects consisting of attribute–value pairs and array data types. However, for sentiment analysis and topic modeling, our input file needs to be text files so that NLP Suite [6] can process them

Original Tweet Content	Tweet Content after File Type Transformation
RT @sophieperez: this corona virus is getting in the way of my summer travels 😒	RT @sophieperez_: this corona virus is getting in the way of my summer travels \ud83d\ude13
RT @sweetromance: BREAKING: This New Coronavirus Chart Would End ALL Hysteria If The Media Actually Reported It 👀🌹	RT @sweetromance: BREAKING: This New Coronavirus Chart Would End ALL Hysteria If The Media Actually Reported It \n\ud83d\udc40\ud83c\udf39\n
@potatosoupstan @Koreaboo They were using the letters 'BTS' to attract people to their cult and it was spreading coronavirus 😡	@potatosoupstan @Koreaboo They were using the letters 'BTS' to attract people to their cult and it was spreading coronavirus \ud83d\ude21

Table 3.2: Garbled Characters Produced by Emojis in Tweets

properly. Therefore, emojis are removed because when transform from json files to txt files, some emojis become garbled characters that corrupt the continuity of textual data as shown in Table 3.2.

3.5 Quality Assurance

Some tweets contain words in various languages but still get categorized as English tweets by Twitter, and languages like Japanese characters usually result in garbled texts after file type transformation. Thus, the tweets are manually examined and go through a series of special character/math symbol/Greek letter searching and deleting to make sure the textual data contains no garbled characters.

Chapter 4

Approach

To prove the potential of Covid-related Twitter Dataset, multiple steps of Natural Language Processing methods including Sentiment Analysis (Section 4.1), which is based on Stanford CoreNLP (Section 4.1.1) and Hedonometer (Section 4.1.2), and Keyword Extraction (Section 4.3), which is based on Topic Modeling with Gensim (Section 4.3.1 and 4.3.2) and NER (Section 4.3.3 and 4.3.4) are used to evaluate our dataset as a practical resource for detecting major social occurrences and building advanced deep learning models.

The proposed pipeline is shown in Figure 4.1, which involves a six-step process of data cleaning, calculating emotion score, keyword generation, and human evaluation. Building on this model, our research discovers that the state-of-the-art sentiment analysis performs poorly because the tweets are posted across the U.S., which results in a diluting effect on the sentiment spike caused by local news events. Therefore, this study proposes a refined

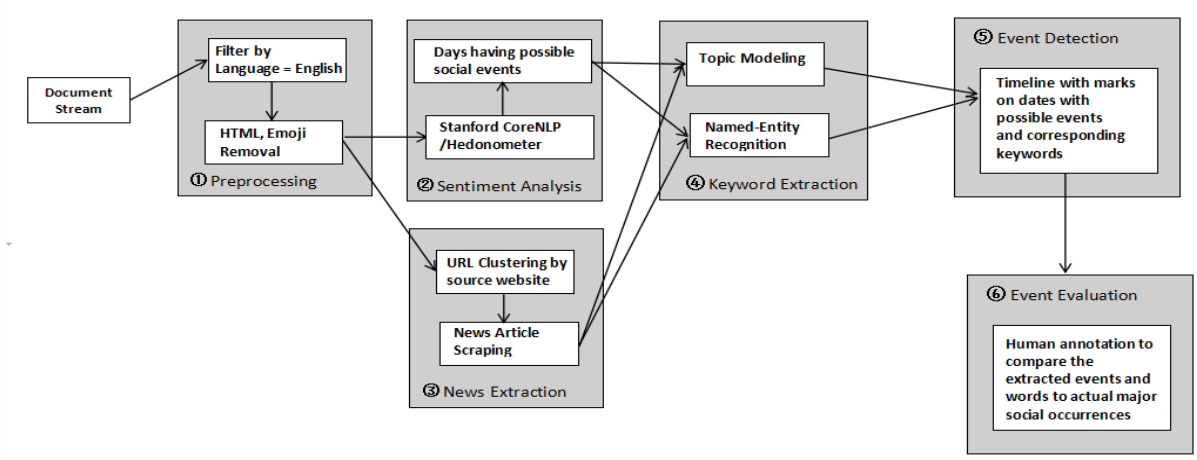

Figure 4.1: Proposed Methodology of Data Processing and Human Evaluation

event detection model which incorporates a location-specification step, as shown in Figure 4.2. The complete process of location-specification will be explained in details in Section 4.1.3.

4.1 Sentiment Analysis

Researchers have long been incorporating sentiment analysis into NLP studies because when machines transform the emotions embedded in texts into statistics, researchers can apply various measurements to keep track of people's level of happiness and extend these data to use in other domains such as reputation monitoring [14]. For this research, two major sentiment analysis

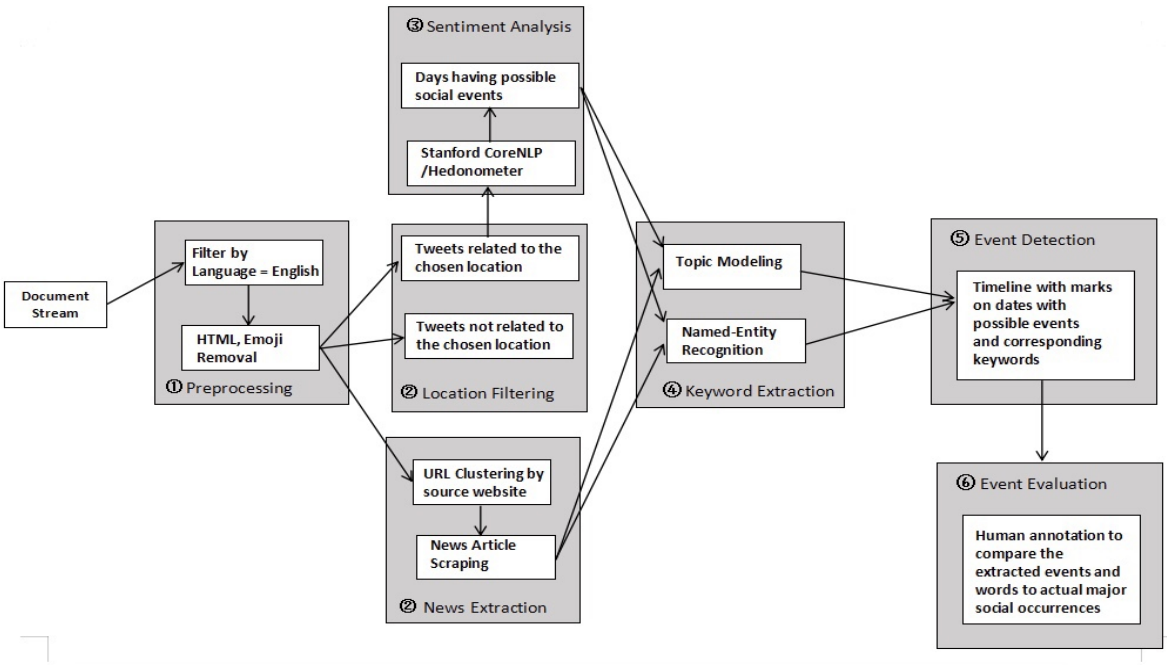

Figure 4.2: Refined Methodology of Data Processing and Human Evaluation

algorithms are employed: Stanford CoreNLP and Hedonometer.

4.1.1 Sentiment Analysis with Stanford CoreNLP

Stanford CoreNLP is a Java-based pipeline of Natural Language Processing which takes raw text as input and produces annotated text as output [8]. As part of the Stanford CoreNLP algorithms, Sentiment Analysis is enabled by deep learning models that assess the emotions conveyed by each sentence, give scores to quantify every sentence's emotion, and mark each sentence's sentiment level as positive, neutral, or negative [8, 16]. A significant advantage

of this approach is its measurement of the context. Unlike other sentiment analysis tools like Hedonometer which is only capable of referring to the dictionary to gauge each word's sentiment, Stanford CoreNLP considers both the word's dictionary meaning and the context to obtain a relatively accurate evaluation of the sentence's emotion level.

In this study, each day's tweets data are annotated by Stanford CoreNLP embedded in NLP Suite [6]. Subsequently, the researcher collects the statistics concerning the frequency of negative tweets and positive tweets for further analysis.

4.1.2 Sentiment Analysis with Hedonometer

Hedonometer is employed in NLP to gauge the scale of happiness conveyed by texts. Hedonometer computes the weighted average magnitude of joyfulness by a frequency-based measurement [3]. Similar to Stanford CoreNLP, Hedonometer gives a score for the sentence's scale of happiness, as well as categorizing it into positive, neutral, or negative.

The Hedonometer measurement model is not context-dependent, but rather relies on a combination of texts collected from Google Books, New York Times, Lyrics, and Tweets to train the model to assign scores of hap-

piness to new corpora, which may fail to yield an accuracy as high as that of Stanford CoreNLP. Nevertheless, when Dodds and his fellow researchers developed Hedonometer, they make this model highly efficient when applied to social media texts like blogs and tweets [3]. Therefore, Hedonometer is incorporated in this study to help assess the level of pleasure conveyed in Covid-themed tweets.

4.1.3 Location-Specific Sentiment Analysis with Stanford CoreNLP

After preliminary computations and comparisons of tweets' sentiment score, the researcher notices that sentiment-based event detection model would always fail.

Considering that the dataset contains tweets posted across the United States, the researcher attributes the ineffectiveness of the model to the excessive locations covered by the dataset. Feng's study validates that social event detection should sometimes be restricted to certain geographic areas to avoid the diluting effect of posts made in other places [5], and thus a location-specification of the dataset is introduced. The location filtering process is added to the pipeline and inserted between Step1. Language Filtering and Step3. Sentiment Analysis.

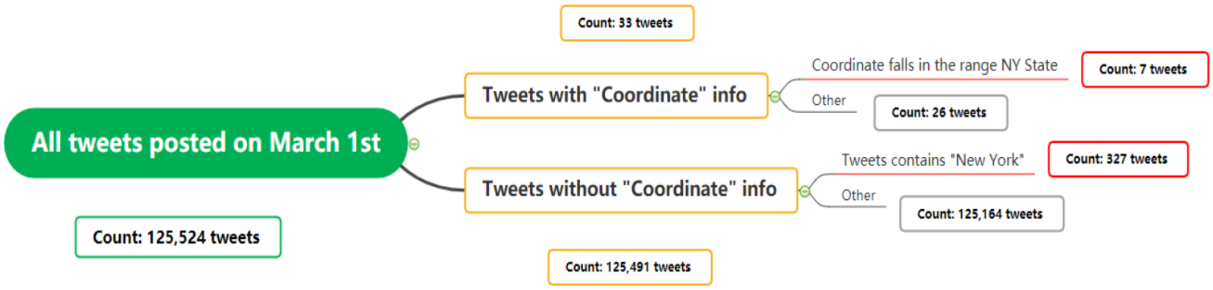

Figure 4.3: Selection of Tweets about New York

Ever since the state of New York had its first confirmed case on March 1st, 2020, both confirmed cases and death cases have been escalating in the state and NY soon becomes the epicenter of the nation. In early April, New York even has more confirmed cases than any other country in the world. Therefore, hoping to achieve event detection in the state facing the most severe threats of the outbreak, this research focuses on tweets related to New York. As shown in Figure 4.3, tweets will be filtered by their location information.

To begin with, since some tweets have coordinate data embedded if the Twitter user chooses to attach their location to the tweet, all data will be categorized into a) have coordinate information, and b) not have coordinate information. Secondly, group (a) will be broken into two subgroups 1) Coordinates falling into the range of New York State (40°30 N to 45°1

N 71°51 W to 79°46 W), 2) Coordinates falling outside of New York State; and group (b) will also be broken into two subgroups 1) Content contains the keyword "New York", and 2) Content does not have the keyword "New York". Lastly, tweets from a(1) and b(1) will be merged into a single file, which contains all the New-York-related tweets for the day.

4.2 URL Link Groupings

The researcher documents the possible compositions of the selected 13 major news websites' HTML links and then calculates the frequency of each news agencies' links appearing in tweets' content. This process enables the researcher to rank the popularity of news websites and choose the most frequently-referred sites for further web scraping.

4.2.1 News Article Scraping

All news articles are scraped in python (package: Goose3) to extract the news content from HTML source code to text files. With Goose3, we parse the metadata of HTML documents and fetch the full texts of news report articles [7].

4.3 Key Word Extraction

Besides generating the timestamps of possible social occurrences, another principal component of event detection is to collect a group of words to describe the reporting events. The common approaches to achieve word assembly are Topic Modeling and Named-entity recognition [19]. Section 4.3.1 and 4.3.2 will respectively present the algorithms of Topic Modeling and NER.

4.3.1 Topic Modeling with Gensim

When there are a considerable volume of texts, Topic Modeling is often employed to generate word clustering that represents hidden topics. Although computer scientists nowadays rarely use Gensim and Mallet, these Topic Modeling algorithms still maintain popular in social science studies [9, 10]. Gensim utilizes Variational Bayes sampling method which yields an advantageous processing speed, whereas Mallet uses Gibbs Sampling, which has a higher precision than Gensim [13]. Since this research needs to process a dataset of more than one million tweets, between the trade off between processing time and precision, the former is more essential. In addition, Gensim provides an interactive visualization that presents how the extracted topics overlap, the

words summarizing each topic, and the proportion of the topic.

To improve the efficiency of Topic Modeling which performs better when there are as many separate files as possible, each tweet is stored as a single file. For instance, we have 85,862 tweets for March 4th, and thus we will have 85,862 files for Topic Modeling. Gensim will be applied to each dataframe for extracting topic clustering.

4.3.2 Named-Entity Recognition

NER is a text annotation method that extracts entities like organization, person, religion, and nationality [19]. In this research, NER is utilized as a supplement to Topic Modeling to help determine the individuals and organizations involved in major events. Therefore, NER is assigned to extract the entities having tags of “person” and “organization”.

Chapter 5

Experiments

5.1 Sentiment Analysis

For our experiments, both results genenated by Stanford CoreNLP and Hedonometer are put into the sentiment-based detection formula proposed by Paltoglou:

$$\mathbf{negFreq(d)} \geq threshold * \mathbf{avgNegFreq(d-1, n)}$$

where

$$\mathbf{avgNegFreq(d, n)} = \frac{1}{n+1} \sum_{i=d-n}^{n} \mathbf{avgNegFreq(i)} \text{ [12]}$$

5.1.1 Sentiment Analysis with Stanford CoreNLP

As shown in Figure 5.1, the twelve days' tweets display a fluctuating level of sentiment polarity. However, when each day's data is plugged into the

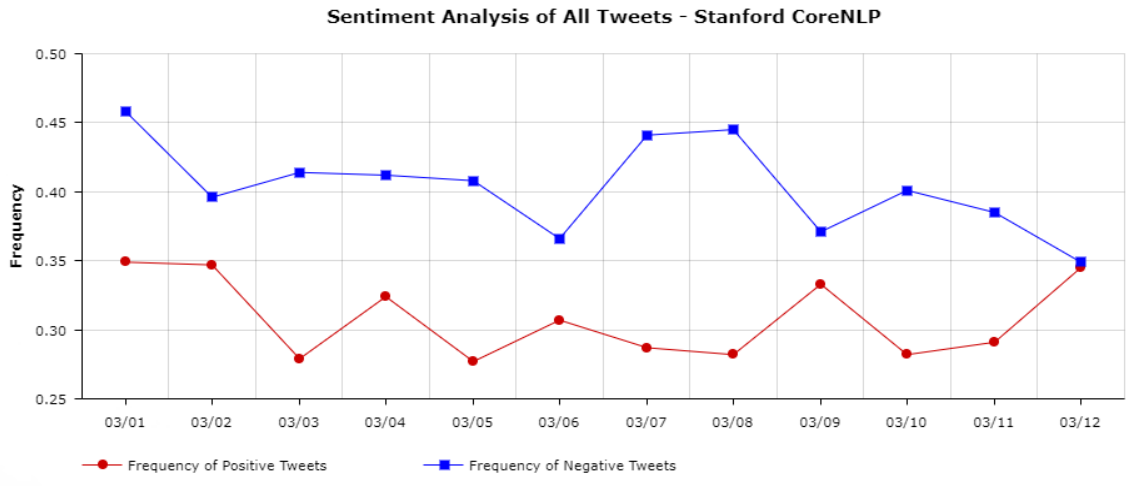

Figure 5.1: Sentiment analysis of all tweets using Stanford CoreNLP

sentiment spike detection formula, neither positive spike nor negative spike is found.

5.1.2 Sentiment Analysis with Hedonometer

As shown in Figure 5.2, the change of frequency of both positive and negative tweets is much less striking compared with the output of Stanford CoreNLP. Also noticeably, both positive (fluctuating at 8%) and negative (fluctuating at 2%) tweets only make up a small proportion of the whole dataset, meaning that about 90% of tweets are categorized by Hedonometer as having neutral emotions, which is contradictory to the actual sentiment displayed by the collected tweets.

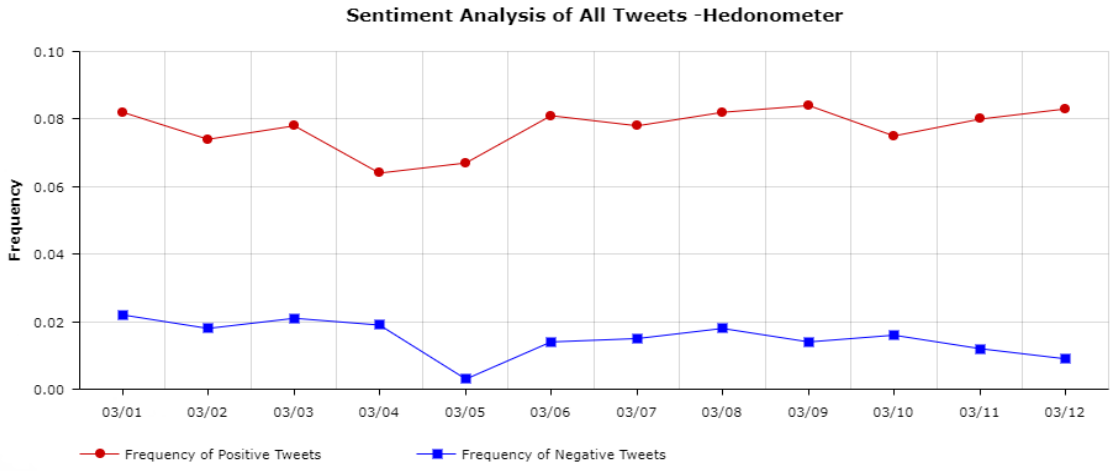

Figure 5.2: Sentiment analysis of all tweets using Hedonometer

5.1.3 Location-Specific Sentiment Analysis with Stanford CoreNLP

Since sentiment-based event detection fails when the dataset is unfiltered by location, we take an alternative approach of detecting hotspot events, which are social occurrences that only receive large social attention in specific area.

After we apply location-specification, the change of frequency of positive and negative tweets in Figure 5.3 is much more salient than Figure 5.2, which are two nearly-flat lines. Plugging the statistics shown in Figure 5.3 into the sentiment spike detection formula, a negative spike is detected on March 7th, 2020 and a positive spike is detected on March 9th, 2020.

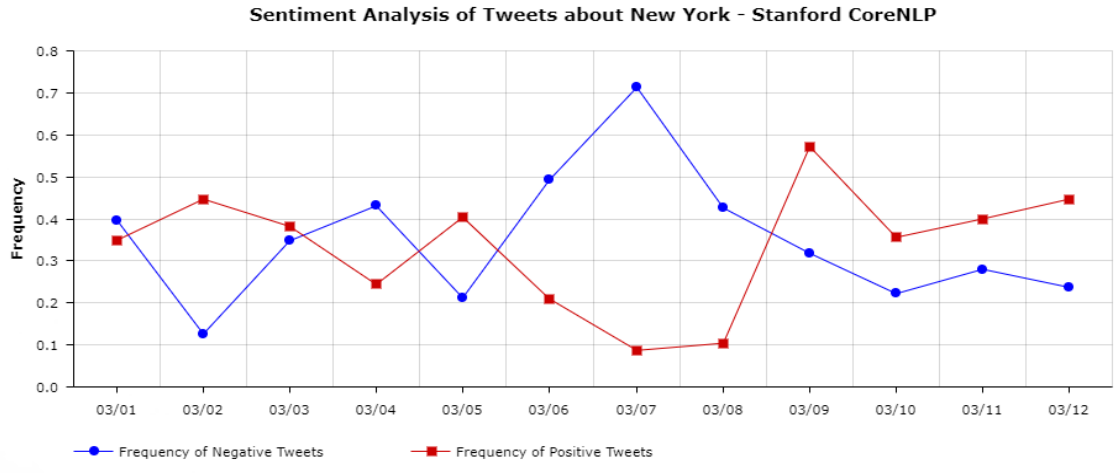

Figure 5.3: Sentiment analysis of New-York-related tweets using Stanford CoreNLP

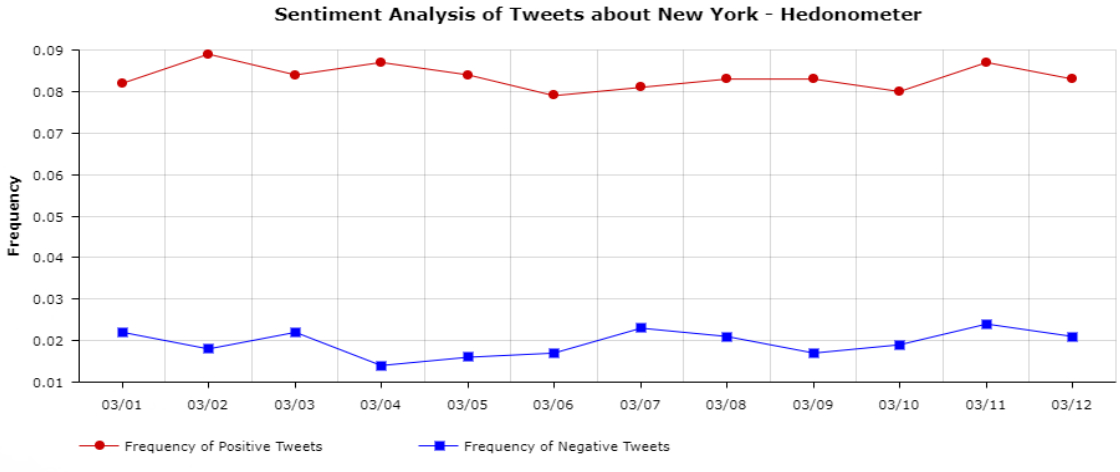

Figure 5.4: Sentiment analysis of New-York-related tweets using Hedonometer

5.1.4 Location-Specific Sentiment Analysis with Hedonometer

The figure above (Figure 5.4) displays the sentiment analysis performed on New-York-related tweets by Hedonometer. Similar to results from Section

5.1.2, Hedonometer still has the tendency to mark most tweets as "neutral".

As a result, neither positive nor negative spike is detected by Hedonometer. Since we knew that there were several major events in the time frame from March 1st to March 12th, 2020 (e.g., Cuomo declaring a state of emergency on March 7th), which should have triggered discernible emotion spikes, we conclude that Hedonometer has failed on event detection. The failure of this model will be discussed in the section of error analysis.

5.2 URL Link Groupings

Although we have prior selected 13 major news websites, we notice that among the 13 websites, some rarely appear in the Covid-related tweets. Therefore, we decide to do another round of website selection that only keeps five websites with the highest popularity in our collected tweets. This step greatly reduces the time of web scraping for our study, though admittedly, also causes some omissions of news reports that are covered by less-popular websites but not the chosen five media.

Figure 5.5 illustrates the distribution of news websites that appear in Covid-related tweets posted on March 1st. Though only one day's data is listed here, we ensure to calculate the distribution of news websites for the

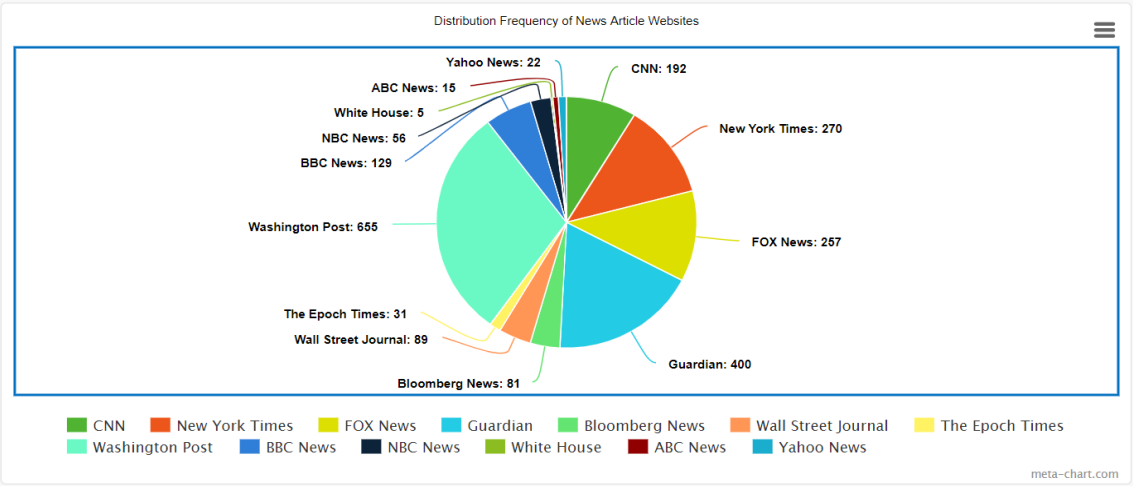

Figure 5.5: News Website Distribution for Tweets Posted on March 1st

entire 12-day frame of data to avoid sampling bias. The statistics we collect are mostly in line with the distribution of March 1st, and we determine the five most-frequently-referred news websites are Washington Post, CNN, New York Times, Guardian, and FOX News.

5.2.1 News Article Scraping

Based on the timestamps generated in Section 5.1.3, March 7th and March 9th have possible events. To begin with, we firstly take the tweet dataset that has not gone through HTML removal. We take the tweets posted on the two dates and repeat the location filtering method to only keep the tweets about New York, because these two dates of interest are generated by location-specified

event detection model. Then, we extract the URL links that lead to the five chosen news websites. Lastly, the news articles from the five major websites are scraped by Python Goose3 for topic modeling and NER in the following steps.

News Agency	**March 7**	**March 9**
Washington Post	27	2
CNN	19	3
FOX News	2	2
New York Times	14	10
Guardian	15	3

Table 5.1: Number of news reports collected from the five websites

As shown in Table 5.1, the five selected news websites in total have 77 articles mentioned in New-York-related tweets on March 7th, and 20 articles on March 9th.

5.3 Key Word Extraction

5.3.1 Topic Modeling with Gensim on Tweets

After storing each tweet as an individual file, topic modeling is performed to generate word clustering. We choose to extract five topics every time we employ Topic Modeling, the reason is two-fold. Firstly, we manually go over covid-related reports on news websites and determine that on a day where

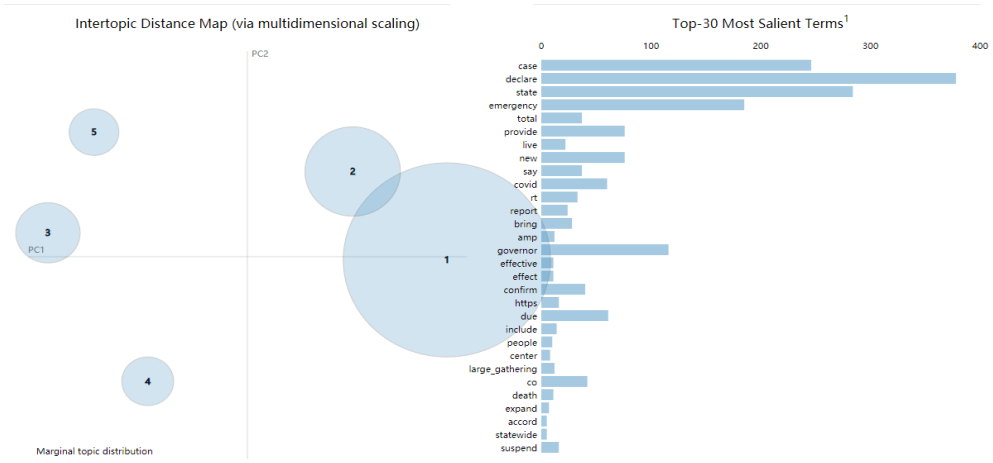

Figure 5.6: Topic Modeling of New-York-related Tweets on March 7th

there are some emergent news, the numbers of such news never exceeds 5. Secondly, we run Gensim multiple times and choose different number of topics. After comparing the results, we discover that setting the number of topics to be between 5 and 10 yield the most ideal result (graphs containing segregated and relatively large bubbles).

As shown in Figure 5.6 and Figure 5.7, the bubbles generated by Gensim are in reasonable size and scattered in different quadrants, indicating that there are common subjects among the documents. However, the extracted words are mostly verbs, giving little information about other key components of an event, such as who, why, and how. Thus, we still need more words to know the details of the events, a task that will be fulfilled by NER in the Section 5.3.3 and 5.3.4.

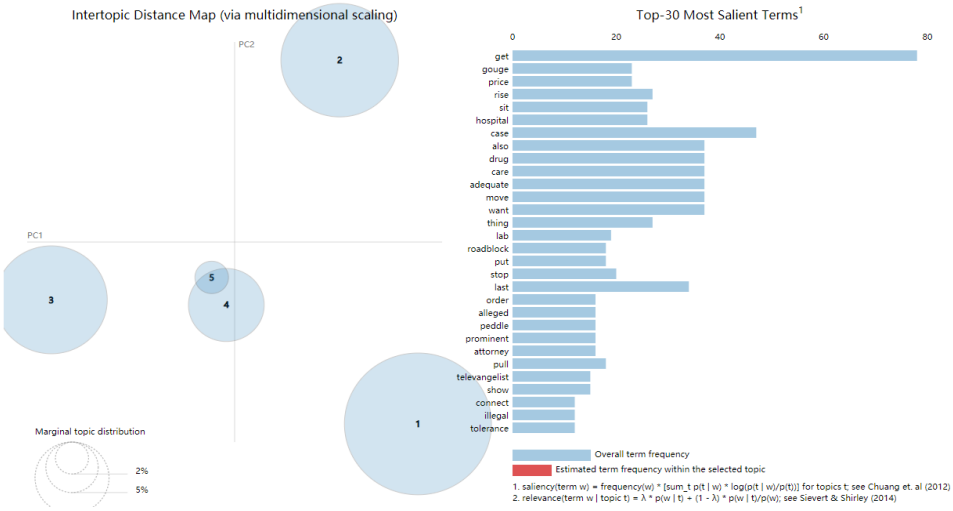

Figure 5.7: Topic Modeling of New-York-related Tweets on March 9th

5.3.2 Topic Modeling with Gensim on News Articles

Topic	Relevant Words
Topic 0	state, say, people, case, announce, new, death, die, include, westchester
Topic 1	say, test, people, ship, official, case, health, passenger, patient, state
Topic 2	say, ship, passenger, captain, supply, attend, people, go, still, come
Topic 3	airline, say, case, people, emission, country, could, health, new
Topic 4	go, test, state, say, emergency, visit, morning, cancel, people, case

Table 5.2: Topic Modeling on News Articles for March 7th

Table 5.2 displays the result of topic modeling on the 77 news articles for predicting the event(s) on March 7th. Although several topics involve similar words like “state”, “people”, and “case”, some extracted words are relatively specific which might provide great help to describe social events, such as “ship”, “Westchester” (a county in New York State), “emission”, and

"passenger".

Table 5.3 displays the result of topic modeling on the 20 news articles for predicting the event(s) on March 9th.

Topic	Relevant Words
Topic 0	say, case, state, test, government, virus, people, public, school
Topic 1	article, biden, lean, trump, campaign, business, dealing, note, plan, right
Topic 2	case, state, say, official, health, day, week, allow, supply, local
Topic 3	government, say, virus, question, public, country, people, party, case
Topic 4	test, state, case, say, school, people, city, official, virus

Table 5.3: Topic Modeling on News Articles for March 9th

Since March 9th only has 20 articles for topic modeling, which is far from enough for producing accurate topic modeling result, as shown in the table above, the generated word clustering does not give an overview of the possible events. Many words repetitively occur in different topics, such as "people", "state", and "say", suggesting that the topics have large overlap. Therefore, the topic modeling of news articles on March 9th provides limited help to event description.

5.3.3 Named-Entity Recognition on tweets

Based on results from Section 5.1.3, March 7th and 9th are marked by Stanford CoreNLP as having sentiment spike. Therefore, NER is applied to the New-York-related tweets posted on these dates. Table 5.4 displays the

dates with corresponding words having the tag “person” or “organization”.

Date	Person	Organization
March 7	Trump, Andrew Cuomo, Ned Lamont	CDC, Healthy Ministry, Trump Administration, Uber, Bridgeport Hospital
March 9	Trump, Andrew Cuomo, de Blasio	CDC, Columbia University, Association of Public Health Laboratories, FDA

Table 5.4: NER results of tweets posted on dates marked by Stanford CoreNLP

5.3.4 Named-Entity Recognition on News Articles

After merging the news articles for predicting events on March 7th into a single file, we utilize NER to extract the nouns of person’s name and organization. The nouns are sorted by frequency, so that only the nouns regularly appear in news reports are kept. The same process is performed for news articles of March 9th.

As shown above, besides “Trump” and “CDC” which are two words frequently appear in Covid-related tweets regardless of the occurrence of major events, NER has pulled out many useful information, such as Andrew Cuomo (Governor of New York), Bill de Blasio (Mayor of New York City), Alex Azar (Former U.S. Secretary of Health and Human Services), Seema

Date	Person	Organization
March 7	Trump, Andrew Cuomo, Alex Azar, Mike Pence, Seema Verma	CDC, Centers for Medicare, International Air Transport Association, Methodist Hospital
March 9	Biden, Andrew Cuomo, de Blasio, Ned Lamont	CDC, White House, Columbia University Center for Advanced Medicine

Table 5.5: NER results of news articles posted on dates marked by Stanford CoreNLP

Verma (health policy consultant and former administrator of the Centers for Medicare & Medicaid Services), etc. These people who are frequently discussed in the news are mostly political figures, experts of public health, or a combination of both.

As for organizations, NER suggests that health care and public health institutions are frequently mentioned in news reports. The three exceptions are International Air Transport Association, Columbia University, and White House. "International Air Transport Association" gives a hint that a potential event about travel control and restriction of overseas flights. "Columbia University" indicates a school policy that makes it to the headline of New York local news, which is confirmed by close reading that Columbia will suspend classes in two days. "White House" suggests that a major event might be related to an official announcement from the government.

5.4 Results

Date	Detected Event	Descriptive words
March 7	1) Governor Cuomo declares state of emergency for New York State	Cuomo, emergency, Westchester, driver, Uber, passenger, St. John's Episcopal Hospital
	2) Passengers on California cruise ship test positive	ship, passenger, positive, case, Grand Princess, trip
	3) Ned Lamont announces a confirmed case travels from NY to Connecticut	Ned Lamont, Connecticut, doctor, Bridgeport Hospital, symptom
	4) Amtrak cuts train service between NY and DC	airline, passenger, cancel, infection Washington, traveler
March 9	1) Cuomo attacks C.D.C over delays in Coronavirus testing	CDC, testing, supplies, capacity, plea, kit, emergency
	2) Columbia University will suspend classes in 2 days	Columbia University, cancel, president, Bollinger
	3) Spokesman for the Association of Public Health Laboratories denied testing shortages	Association of Public Health Laboratories, testing, capacity
	4) Mayor Bill de Blasio announces a new confirmed case in Bronx, NY	de Blasio, Bronx, resident, business, school, supplies

Table 5.6: Detected Event for March 7th and March 9th

Putting together the words extracted by the sections discussed above, this study in total extracted four events for March 7th and another four events for March 9th. Table 5.6 shows the list of events, their date, and

corresponding words pulled out from the tweets/news websites that help the researcher locate the event. As shown in the table, most detected events are local news in New York state, which is in accordance with our hypothesis that location-specified sentiment-based event detection will effectively draw out hotspot events.

For events on March 7th, event (1), (3), and (4) are local news reported on March 7th. Given that these events are all officials' responses to the spread of the virus, the detected negative spike is well explained. However, event (2) is reported on March 6th, one day prior to when the tweets are posted. A closer examination reveals that this news may appeal more to people in San Francisco, the outset city of the cruise, and for people living in New York, they might not quickly hear about this news because it's less relevant than event (1), (3), and (4). This result is in line with prior work that suggests people will react much more promptly to local news, but they oftentimes need longer time to pay attention to news happening in other regions, which is why researchers usually allow for a 24-hour-lag for detecting global news [12, 5].

For March, 9th, the four events are all reported on the same day as the tweets are posted. The range is wider than the news of March 7th, ranging from universities' announcement to public health staff's reply to New York

Governor's accusation. This result embodies the potential of Twitter dataset for open-domain event detection because when NER is used properly, the output covers people from various fields, which lead to discovering social occurrences in different industries.

On the other hand, a distinct inaccuracy of the detection result is that the model has detected a positive spike on March 9th, but all the relevant events are negative, such the conflict between Governor and C.D.C. To explain for this discrepancy between the positivity expressed in tweets and the antipathetic content of the news, we manually examine the news reports and find a plausible way to understand it: After Governor Cuomo criticizes the health organization for not testing enough people, the officials are taking measures to increase the testing capacity and ship more knits to local health centers. If the public gets informed that more people would soon have access to testing, they are likely to express positive emotions towards the seemingly-negative news.

5.5 Error Analysis

Since Hedonometer fails to detect any events for both the unfiltered dataset and the dataset preprocessed with location specification, an extensive error

analysis is performed on explaining such inefficiency of Hedonometer. As shown below, Hedonometer tends to mark most (about 90% of all) tweets as neutral, and for tweets that are not categorized into neutral, they are more likely to be marked as positive than negative, whereas Stanford CoreNLP shows the proportion of negative tweets largely exceeds that of positive tweets.

Date	**Positive**	**Neutral**	**Negative**
March 1	8.2%	89.6%	2.2%
March 2	8.9%	89.3%	1.8%
March 3	8.4%	89.4%	2.2%
March 4	8.7%	89.9%	1.4%
March 5	8.4%	90.0%	1.6%
March 6	7.9%	90.4%	1.7%
March 7	8.1%	89.7%	2.3%
March 8	8.3%	89.6%	2.1%
March 9	8.0%	90.3%	1.7%
March 10	8.7%	89.4%	1.9%
March 11	8.8%	88.8%	2.4%
March 12	8.3%	89.6%	2.1%

Table 5.7: Percentage of positive/neutral/negative New-York-related tweets on each day calculated by Hedonometer

Misclassification After manually examining the tweets that have been categorized into "neutral" by Hedonometer, the researcher notices that Hedonometer sometimes classifies tweets as neutral even if the sentiment is distinctly negative. As shown in the table below, the three examples convey negative emotions but all are marked as neutral tweet by Hedonometer. Errors

in this category have no apparent cause to understand why Hedonometer makes such assessment. Because of the large amount of tweets, deciding what proportion is misclassified requires too much human labor for close-up evaluation.

Tweet Content	Sentiment Value	Word List
If you are feeling low you know who to blame: New York Times Headline!	Neutral	feel, low, know, blame, new, york, time, headline
We can't rely on CDC testing.	Neutral	ca, rely, test
New York's tax base is both exceptionally wealthy—and exceptionally fragile.	Neutral	new, york, tax, base, wealthy, fragile

Table 5.8: Examples of neutral tweets marked by Hedonometer

A possible explanation for such misclassification is that Hedonometer has an inefficient parser. For instance, in the second sentence from the table above, the word "can't" is parsed as "ca" in the word list, which wipes off the negative meaning carried by the original word. Nonetheless, even though sentence 1 has most of the words correctly dissected, the sentiment value is still imprecise.

Given this analysis, we hope the challenges caused by Hedonometer are well demonstrated and become easier to be overcome in future studies. Researchers can consider utilize sentiment analysis tools that also employ a ternary classification method but have a higher accuracy when applied to social media

content than Hedonometer.

Chapter 6

Conclusion

Since a large number of social media posts emerge every day, there's the need for automated systems that processes large volumes of content from heterogeneous streams, detect and track breaking news events, and provide an informative description of events. Such resources can help the public to stay informed and make timely decisions.

This thesis presents an open-domain event detection method that takes raw data from social media as the input, goes through a series of Natural Language Processing, and produces a list of events with descriptive words as the output.

An extensive and comprehensive analysis on refining the sentiment-based method is performed to show the difficulty of event detection when the dataset contains noise from various sources. Multiple strategies to filter the raw data are experimented and reported, providing more than one way of generating

meaningful detection results. Three state-of-the-art NLP tools are run and compared, and show the potential of the Twitter dataset to provide statistical information in various fields, such as tracking the public's emotion and form a database consisting of news fragments.

This research reveals that even if the token of interest is a negative event, sentiment-based detection model is still applicable, assuming that a multi-fold preprocessing of the dataset is operated. The proposed location-specification detection model validates that even when the dataset is filtered by location, it is still capable of detecting both local and global events. The ineffectiveness of Hedonometer suggests that a more sophisticated handling of the sentiment analysis is needed, where a sentiment analysis tool that relies on the context is preferred. Finally, the researcher incorporates a tentative explanation for the failure of several models, with a hope to offer insightful retrospective and make suggestions to future deeper study. Finally, the abundant information collected from Topic Modeling and NER, such as people's and institutions' names, assists us to precisely identify a list of events. Researchers interested in event detection should consider incorporating Topic Modeling and NER in their process of descriptive keywords extraction.

For future work, the extracted events (Table 5.6) and error analysis

(Section 5.5) can serve as guidelines to further enhance the performance of the sentiment-based event detection model.

Appendix 7

Complete Results

Date	Detected Event	Descriptive words
March 7	1) Governor Cuomo declares state of emergency for New York State	Cuomo, emergency, Westchester, driver, Uber, passenger, St. John's Episcopal Hospital
	2) Passengers on California cruise ship test positive	ship, passenger, positive, case, Grand Princess, trip
	3) Ned Lamont announces a confirmed case travels from NY to Connecticut	Ned Lamont, Connecticut, doctor, Bridgeport Hospital, symptom
	4) Amtrak cuts train service between NY and DC	airline, passenger, cancel, infection Washington, traveler
March 9	1) Cuomo attacks C.D.C over delays in Coronavirus testing	CDC, testing, supplies, capacity, plea, kit, emergency
	2) Columbia University will suspend classes in 2 days	Columbia University, cancel, president, Bollinger
	3) Spokesman for the Association of Public Health Laboratories denied testing shortages	Association of Public Health Laboratories, testing, capacity
	4) Mayor Bill de Blasio announces a new confirmed case in Bronx, NY	de Blasio, Bronx, resident, business, school, supplies

Bibliography

[1] Jun Araki and Teruko Mitamura. Open-domain event detection using distant supervision. *Proceedings of the 27th International Conference on Computational Linguistics*, pages 878–891, 2018. URL `https://www.aclweb.org/anthology/C18-1075`.

[2] M. Cataldi, Luigi Di Caro, and C. Schifanella. Emerging topic detection on twitter based on temporal and social terms evaluation. *Tenth International Workshop on Multimedia Data Mining table of contents*, 2010. URL `http://nyc.lti.cs.cmu.edu/classes/11-741/s17/Papers/cataldi-mdmkdd10.pdf`.

[3] Peter Sheridan Dodds, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PLoS ONE*, 6:e26752, 2011. URL `https://doi.org/10.`

1371/journal.pone.0026752.

[4] Lev Facher. President trump just declared the coronavirus pandemic a national emergency. here's what that means. *STAT*, 2020. URL https://www.statnews.com/2020/03/13/national-emergency-coronavirus.

[5] Xiao Feng, Shuwu Zhang, Wei Liang, and Jie Liu. Efficient location-based event detection in social text streams. *Intelligence Science and Big Data Engineering*, 9243:213–222, 2015. URL https://doi.org/10.1007/978-3-319-23862-3_21.

[6] Roberto Franzosi. Natural language processing suite. 2021. URL https://github.com/NLP-suite.

[7] Xavier Grangier. Goose 3 article extractor. 2011. URL https://github.com/goose3/goose3.

[8] Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, page 55–60. Association for Computational Linguistics, 2014. URL https://www.aclweb.org/anthology/P14-5010.

[9] Emily A. Marshall. Defining population problems: Using topic models for cross-national comparison of disciplinary development. *Poetics*, 41.6: 701–724, 2013.

[10] Andew Kachites McCallum. Mallet: A machine learning for language toolkit. 2002. URL `http://mallet.cs.umass.edu`.

[11] Iraklis Moutidis and Hywel T. P. Williams. Good and bad events: Combining network-based event detection with sentiment analysis. *Social Network Analysis and Mining*, 10:64, 2020. URL `https://doi.org/10.1007/s13278-020-00681-4`.

[12] Georgios Paltoglou. Sentiment-based event detection in twitter. *Journal of the Association for Information Science and Technology*, 67:1576–1587, 2015. URL `https://doi.org/10.1002/asi.23465`.

[13] Greg Rafferty. Lda on the texts of harry potter: Topic modeling with latent dirichlet allocation. *Toward Data Science*, 2016.

[14] Andrew J. Reagan, Brian Tivnan, Jake Ryland Williams, Christopher M. Danforth, and Peter Sheridan Dodds. Benchmarking sentiment analysis methods for large-scale texts: A case for using continuum-scored words

and word shift graphs. *arXiv*, 2016. URL https://arxiv.org/abs/1512.00531.

[15] Isaac Sebenius and James K. Sebenius. How many needless covid-19 deaths were caused by delays in responding? most of them. *STAT*, 2020. URL https://www.statnews.com/2020/06/19/faster-response-prevented-most-us-covid-19-deaths/.

[16] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, page 1631–1642, 2013. URL https://www.aclweb.org/anthology/D13-1170.

[17] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. Sentiment in twitter events. *Journal of the American Society for Information Science Technology*, 62:406–418, 2011. URL https://doi.org/10.1002/asi.21462.

[18] Sayan Unankard, Xue Li, and Mohamed A. Sharaf. Emerging event detection in social networks with location sensitivity. *World*

Wide Web, 18:1393–1417, 2015. URL https://doi.org/10.1007/s11280-014-0291-3.

[19] Konstantinos N. Vavliakis, Andreas L. Symeonidis, and Pericles A. Mitkas. Event identification in web social media through named entity recognition and topic modeling. *Data Knowledge Engineering*, 88:1–24, 2013. doi: https://doi.org/10.1016/j.datak.2013.08.006. URL https://www.sciencedirect.com/science/article/pii/S0169023X13000827.

[20] Han Zhang and Jennifer Pan. Casm: A deep-learning approach for identifying collective action events with text and image data from social media. *Sociological Methodology*, 49:1–57, 2019. URL https://doi.org/10.1177/0081175019860244.