**Cardiff University**

**School of Computer Science and Technology**

**MSc computing in IT management**

Managing Misinformation [English vs Chinese Social Media]

Student Name: HONGLANG LUO student ID: C1961145

Project Supervisor: Alun Preece

Table Of Contents

[Abstract 3](#_Toc56693432)

[Introduction 4](#_Toc56693433)

[Research background and motivations 4](#_Toc56693434)

[Research objectives 4](#_Toc56693435)

[Research methodology 5](#_Toc56693436)

[Research structure 10](#_Toc56693437)

[Literature Review 11](#_Toc56693438)

[Politics on social media 11](#_Toc56693439)

[Rumors and offensive comments on social media 14](#_Toc56693440)

[Previous research on rumor identification on social media 17](#_Toc56693441)

[Limitations of the previous research 20](#_Toc56693442)

[Data collection 22](#_Toc56693443)

[Data analysis 26](#_Toc56693444)

[Results and Discussion 30](#_Toc56693445)

[Process change 30](#_Toc56693446)

[Final result and discussion 34](#_Toc56693447)

[Conclusion 48](#_Toc56693448)

[Innovation and Main findings 48](#_Toc56693449)

[Project limitations 50](#_Toc56693450)

[Future recommendations 51](#_Toc56693451)

[Project summary 52](#_Toc56693452)

[Reference 54](#_Toc56693453)

[Appendix 59](#_Toc56693454)

# Abstract

With the rapid development of social media information, emerging media platforms represented by Twitter and Facebook have become the main platforms for people to obtain information, share information, and spread information. However, the convenience and richness of information access brought by these social media platforms. The widespread dissemination of rumors in social media has severely interfered with people’s access to reliable information and may cause huge economic losses or cause some emergencies. Serious consequences. Public panic. The management of rumors and the credibility of social media has dropped sharply, which brings challenges to related research in the field of rumor detection and management. Here starts with the politicization of social media, and at the same time, mentions the spread of rumors on social media and the corresponding detection methods. The main research work includes 1. Data extraction. Extract useful fields and text data from the original data set. A python toolkit for JSON data processing and text mining. 2. Try to find rumors and misinformation from negatively sentiment Twitter. Through the test, the machine learning method is not suitable for this project, so the data set was screened many times and finally, a negative sentiment text data set was obtained. 3. In-depth study of the forms and characteristics of rumors in different countries and rumors in different languages for understanding and research. The rumors in different countries were analyzed through visual charts, and the topics involved were different, but there were also cases of mutual dissemination of false information. The project provides examples of managing rumors and misinformation on social media. It provides useful analysis ideas and methods for future research on rumor management and identification.

**Keyword:** Rumors and misinformation, python, data analysis, text mining, visualization

# Introduction

Research background and motivations

The rapid development of the Internet has achieved explosive growth and decentralized dissemination of information, and social media has gradually become an important channel for people to obtain or share information. This way that everyone can participate in information release and dissemination provides great convenience for people’s information sharing At the same time, it also brings some outstanding problems, especially the continuous breeding and rapid spread of online rumors, which poses serious challenges to the effective use and scientific management of social media. To some extent, social media has become rumors and false information. The main areas of communication. Due to the public nature of social media, rumors and false information will continue to strengthen their "public" attributes and spread in the "public" space, generating huge energy and chain reactions, which are unpredictable. And it usually has a strong public opinion, and it is easier to attract people's attention. This is conducive to the spread of rumors. At the same time, with the online application of social media, rumors spread in this new way, showing new characteristics different from the traditional environment. So this project will study what kind of rumors and misinformation exist in social media. According to rumors in different countries, there will be some different themes. Whether there are differences in rumors between different languages. These will help to better manage rumors and social media to better identify and sort out different types of rumors.

Research objectives

Therefore, this project will try to compare the characteristics and differences of rumors between different languages and countries based on data in different languages. This includes finding hot events and topics that have occurred around the world in recent months. These will cause people to vent and express a certain amount of information on social media, and in recent months, the mood of people on social media in various countries has been changing. At the same time, many political and military issues related to the country will appear on social media, and these issues will cause some rumors and misinformation to spread on social media. This error message can be visually analyzed to discover connections between multiple countries and find some specific examples. Therefore, this research will focus on three main issues. 1. After extracting a specific social media data set and performing text mining, try to find some tweets similar to rumors or misinformation in the text data. 2. Extract and analyze data sets in different languages (mainly English and Chinese), and try to find the differences in characteristics and content between them. 3. For the fields in the collected text data, try to slice Twitter according to different countries and regions or popular topics. And explore the content and sentiment differences and characteristics of these tweets.

Research methodology

|  |  |
| --- | --- |
| 1. Research approach and Research design | The selection and reasons of the main research methods used in the project will be explained. The main process of the whole research is described in detail |
| 2. Research tool | The main tools and software involved in the research |

Research approach and Research design

This project is a data analysis and mining project. The research methods adopted are a mixture of quantitative analysis and qualitative analysis. Then, quantitative analysis is a research method and process to express problems and phenomena by quantity, and then analyze, test and explain, to obtain the significance, or to find out the quantitative change rule among some factors. Qualitative analysis is to investigate whether the object of study has such or such attributes or characteristics and whether there is a relationship between them through observation, experiment and analysis (Marvasti 2018). Firstly, quantitative analysis in this project is based on digit and data. When analyzing the Twitter data in Chinese and English, it is necessary to check the original data for missing data, delete outliers and transform variable operations. Besides, some data visualization operations are needed to show the data frequency and the amount of data, and emotional scores are also required for the text. In addition, qualitative analysis in this project is language-based. While mining the text data of Twitter, it is necessary to analyze and mine the content, words and emotional preference of the text, and find out the differences between tweets of different languages.

CRISP-DM is an excellent choice for the entire research process (Wirth 2000). First, it has no specific tool limitations or domain limitations and is a standard methodology for all industries. Secondly, CRISP-DM methodology defines data mining practice as six standard stages, namely, business understanding, data understanding, data preparation, model building, model evaluation and model publishing. Therefore, for this project, the business understanding part can be translated into the discovery of research problems, the determination of data mining goals and the formulation of data mining plans. The model publishing part can be understood as data mining reports. Through data analysis and mining results, the hidden information and knowledge behind the data are revealed.

图片包含 游戏机, 文字, 地图

描述已自动生成

Figure 1: The process of CRISP-DM (Pete et al. 2004)

The main issue of the study has been to research the emotional tone and aggression between Chinese and English twitter text data. That's the first part of CRISP-DM. According to this research goal, the whole research can be divided into several important processes. Data collection process, data analysis process, data mining process, report writing process. In the data collection part, the main work is to find the data fields useful for this project from the original Twitter data file, and then extract and integrate these fields. Then data analysis, mainly involving exploratory data analysis and data visualization, through in-depth understanding of the data, to find some useful information for the next stage. The data mining part is the key to solve the problem of the project, what differences exist in tweets between different languages, find offensive tweets, and find the current hot topics and targets of these tweets. At last, need to write a data mining report, Through the analysis and mining process in the previous stage, can get some conclusions and charts, this information needs to be included in the report.

Research tool

Research tools include hardware and software as well as data sources and tools for collecting data sources. The hardware is mostly a Well-functioning windows laptop. The important research tool is the software component. The project is a data analysis mining project, so it involves some toolkits and data analysis environment. Python is a very compatible language for both data analysis and mining, includes various interactive compilers to provide a more sophisticated coding environment. In this project, Jupyter Notebook in Anaconda (Anaconda 2020) as a platform for coding and showing some results. There are also a lot of toolkits will be used. Pandas, Numpy, Matplotlib, Seaborn, pyecharts, JSON, Scikit-Learn for data mining, NLTK for text data processing, and Jieba for Chinese text processing. The Mango database where the data is stored for Chinese Twitter data. There are separate toolkits for dealing with text sentiment. And some text corpus dictionaries and test data sets. The English and Chinese data sources are from Twitter, and the English dataset is from the Cardiff Institute of Crime and Security. Chinese dataset is collected through the TwitterAPI. The code comes from the TwitterHarvest project (David M Rogers 2020).

A list will be used below to show all the research tools used and why they were used in this project.

|  |  |
| --- | --- |
| Tools | Reason |
| Well-functioning windows laptop | Necessary equipment for the project |
| Twitter (Twitter 2020a), TwitterAPI (Twitter 2020b) | Twitter provides APIs used by developers to facilitate data extraction. And Twitter data has a large number of fields for analysis. It is also the social media service platform with the most users. And there are a lot of misinformation and rumors on Twitter. |
| Anaconda (Anaconda 2020) | Anaconda contains more than 180 scientific packages and their dependencies, including conda and Python. At the same time, it has the characteristics of open source, simple installation process, and high-performance use of Python and R language, which is very suitable for data analysis and data mining projects. |
| Jupyter Notebook (Anaconda) | Jupyter Notebook helps to write code, run the code, and obtain the running result of the code. The feature is that it can easily add description, explanation and explanation of the document to the code and its running result. |
| Pandas (pydata.org 2019), Numpy (NumPy 2020), Matplotlib, Seaborn, JSON | pandas is a fast, powerful, flexible and easy-to-use open source data analysis and processing tool, very suitable for use in Notebook in the python environment. Usually NumPy is used with pandas and has powerful functions. Matplotlib and Seaborn are drawing toolkits that support python for data visualization, and JSON is used to read data sets in python |
| Scikit-Learn (Pedregosa et al. 2011) | Sklearn is a very powerful machine learning library provided by a third party in Python, which covers all aspects from data preprocessing to training models. |
| NLTK (Project 2020) | NLTK is a natural language toolkit. It collects a large number of public data sets and provides a comprehensive and easy-to-use interface on the model, covering word segmentation, Part-Of-Speech tag (POS-tag), Syntactic Parse and other NLP functions. The data analysis in this project mainly involves text data, so NLTK is very suitable for this project. |
| Jieba (Junyi 2013) | Stuttering word segmentation is a very popular Chinese open source word segmentation package, which has the characteristics of high performance, accuracy, and scalability. Currently, it mainly supports python. |
| SentiStrength (Thelwall 2017) | It is a sentiment analysis tool with a very simple algorithm. It comes with a sentiment dictionary, which can accurately obtain the sentiment scores in the network text. |
| MongoDB (MongoDB 2020) | The Chinese data obtained from the Twitter API is a file in JSON format. A non-relational database should be used. And MongoDB, which stores data in the form of files, is very suitable for this project. |
| bixin(bung 2020) and cnsenti (Thunderhit 2020), baidu\_stopwords(goto456 2020) | Due to the lack of sentiment analysis support for Chinese texts. two different sentiment analysis packages specially used to analyze Chinese text to score the sentiment of the text. At the same time, stop words are also imported from outside to remove stop words in Chinese text. |
| Pyecharts (Pyecharts 2020) | Echarts is a data visualization open sourced by Baidu. With good interactivity and sophisticated chart design, pyecharts combines python and echarts. Very suitable for data processing. |

Research structure

First, the literature review chapter will start with the rapid development of social media and investigate the reasons for the rapid spread of rumors on social media. Look at the main forms of rumors in the past few years, and how scholars in the past have detected and identified them and pointed out the deficiencies of past research.

Second, the research method chapter. The description of the data collection process, including the collection and extraction of Twitter data in two languages. The data analysis process mainly involves the processing of text data.

Third, the description of the data collection process includes the collection and extraction of Twitter data in two languages, including the use of data sets and the selection of fields. The data analysis process mainly involves the processing of text data. And the method of screening negative sentiment tweets for rumors.

Fourth, the conclusion chapter. This chapter will make a final summary and sublimation of the results of the previous chapter. And discuss some of the limitations of the project, and what can be changed in the future. And the final project summary.

# Literature Review

This chapter begins with a review of why social media has become a political tool and weaponized. Hence the rumors and offensive statements on social media, which then describe the development of technology to identify them. Through the previous explanation of misinformation in social media, that the mass dissemination of misinformation may cause the misinterpretation or arouse the negative feelings of the public, and even raise the possibility of high Internet crime and the very serious social problems (Patton et al. 2014). In view of the impact and danger of rumors, in recent years, research on social media has led to widespread attention. It is necessary to understand and study the types and ways of spreading rumors on social media, and some scientifically feasible ways to test these statements.

Politics on social media

There are some positive effects of social media. The book (Seib 2012) argues that the age of social media has made an indispensable contribution to political decision-making and some diplomatic developments. This book discusses the famous 2011 Arab uprisings and the diplomatic challenges posed to many governments by popular movements and media developments. This shows that social media already has a great influence, and the book also explains how policymakers can actively adapt to and use the real-time features of new media to make more rational political decisions. At the same time, social media, which in previous years became a vehicle for social and political activities, now plays an important role in influencing government decisions. Political Information Weekly (Sandoval-Almazan and Ramon Gil-Garcia 2014) at the time noted that social protest had become a powerful expression of social, policy and opposition to government. Social media has been used in this regard as a vehicle and tool for rallying people and calling out grievances. The research indicates that this kind of large-scale network activity has a four-stage development cycle. Something happens, the media reacts, the Internet goes around, and eventually people start to mobilize. The importance of social media in this process was illustrated through three different periods of social protest movements. Social media provides more opportunities for citizens of various countries to participate in and discuss political and social issues. To some extent, it makes the countries more democratic.

In addition, there are discussions about Chinese social media in this paper (King et al. 2017). Some Western social media believe that there will be some people in China's social media to defend the Chinese regime. These people are called "50C". It is a special kind of person who is used to publish clandestinely or fabricate a lot of comments on social media to mislead and make the Chinese regime more stable. But the study looked at the content of relevant posts by 50c members on Chinese social media Sina Weibo and screened them. It turns out that most of them are incentives to the state of the nation. So these 50Cs are not emotional agitators with a special purpose, nor are they paid to fabricate and publish special comments. At the same time, it can be found that social media has become some political tools and citizens are also willing to express their personal views and opinions on political issues.

However, the technology of social media is also easy to weaponize (Reynolds 2014). The book focuses on the role and characteristics of social media in cyber warfare. And how the media can influence people's beliefs and attitudes, and even mobilize or direct their actions. Social media played an important role in mobilizing and assisting military activities through the conflicts between countries at that time. This can be divided into several strategic points. First, increase the visibility of messages and use some subject tags to expand the influence. Then start to spread misinformation and rumors to target important events and start to target attacks. Meanwhile, the book holds a negative attitude towards the use of relevant technologies to control these malicious rumors, believing that social media is operated by public opinion. More evidence of this can be found in another book (Singer and Brooking 2018). Having revolutionized everything from dating to business to politics, social media is now changing the war itself. Smartphones and social apps have clearly changed the fundamentals of violent conflict. A "us vs. them" (Cikara et al. 2011) narrative prevalent on social media, virtual communications platforms have become an integral part of war strategy. All political, social and national conflicts spread rapidly through social media like Facebook and Twitter. Moreover, these social media maintain a high degree of anonymity, and some social platforms lack tools to verify the authenticity of their content and are difficult to verify and delete one by one, so some false information spreads rapidly among the general population. Finally, it caused some mental shock or shock to a large number of people.

At the same time, social media has become a part of people's lives. Social media sites such as Facebook and Twitter are particularly popular in the United States. The evidence in the Society weekly (Kruse et al. 2018) shows that about half of the U.S. population uses social media. As social media is accepted by the public, there is a critical discussion in the weekly about whether social media is the public sphere of politics. On the one hand, it believes that the organization of social media can enable more public to participate in social and political topics, also provide favorable conditions for people to participate in more online communication activities. On the other hand, it points out that most social media give priority to some content for certain economic interests, and people do not conduct rational or objective debates on social media. One of the main reasons for the emergence of all kinds of false information is to get more attention, so as to get more advertising revenue. The other reason is that once they express their political views, some users will suffer online harassment and even insulting and offensive behaviors. In addition, another report (Tucker et al. 2018) argues that there is a lot of political misinformation in social media. As the paper just showed, social media enables more people to participate in the discussion of political issues. In this way, when some politicians need stories and events to enhance their image, they will create false information or even rumors on social media or magnify false information from other sources to achieve a certain purpose. The report gives a detailed analysis of the spread and causes of political disinformation. An important one is that inflaming emotions, such as anger, makes people more likely to trust and support an incorrect false message, and more likely to spread information about their views. Then confusion has the opposite effect, prompting individuals to seek out information that is accurate. It is clear that there is some unfairness and subjective awareness in online political activities and discussions on social media.

Through the above research and paper, social media has influenced people's lives and the way they used to communicate. The penetration and influence of social media on politics are becoming increasingly prominent, which is changing the mode of political participation of people in various countries, the discourse space of international politics and the behavior ability of individual politics. So some politicians or organizations will spread some wrong information in order to achieve a certain purpose.

Rumors and offensive comments on social media

Social media rumors will lead to serious social problems. Especially, teenagers have no independent ability to identify these online rumors This report (Patton et al. 2014) points out that social media has become a medium of online violence among teenagers. This study shows that social media sites, such as Facebook and Twitter, are difficult to monitor and censor, leading to teenagers' easy exposure to illegal and violent behaviors or incidents on the Internet. The research shows the negative influence of social media on teenagers through a lot of online violence incidents. But there was also a lack of rigor, and the research did not suggest how to avoid or prevent these events. There is not enough real data, only some incidents of cyber violence are listed. And the report found that there was not enough experience and technology to detect the presence of violence.

With the emergence and rapid popularization of social media, online crisis events spread rapidly in the mainstream media. In this study (Pang et al. 2014), it is mentioned that when some people speak about some special events, the mainstream media will report some things because they are newsworthy. Initially, the event is targeted at some individuals or organizations, but if similar events occur in history or involve the interests of many people, the media will think that the event can be spread or packaged to get more attention. For example, a public figure buying a flight was banned by the airline for buying two tickets because of her body shape and was posted on social media. The negative news about the event spread quickly. As a result, the airline suffered a lot of damage to its reputation. It can be seen from this case that social media plays a catalytic role in the fermentation of these events. It will make these news spread quickly so as to obtain some news value.

Some scholars want to know more about the problem of fake news or rumors. The report (Civic et al. 2019) examines why so many people share false and misleading information online at all times. There is evidence that the anonymity of social platforms is very high, so that when interacting on social media, people will dare to question authority, want to reveal more information, and do not worry excessively about receiving harassment and retaliation for their behavior. Their research suggests that some users of social media in the UK share false information without knowing it. Because the person sharing the information will not deliberately verify the authenticity of their information. However, some users will fix some other users' error messages. It's also a way for British citizens to engage in political activism on social media. But such false news can undermine the healthy functioning of these countries.

It is inevitable for rumors and offensive statements to spread on social media (Hu et al. 2018). On the one hand, it is a kind of helpless catharsis, but disappointed by the reality, it is powerless to change, so extreme emotions will be vented on the Internet. On the other hand, there is an ulterior motive. Many extreme problems have ulterior motives. Either way, extreme ideas are spread on the Internet. Because of the disorder of the Internet, ordinary people tend to respond to extreme dissent with more extreme language. This leads to a more polarized mood and leads to more extreme thinking. It's a vicious circle. In addition, network platforms tend to push their favorite content to users, resulting in users can only see what they want to see, which further deepens the consolidation of extreme ideas. The report(Mathew et al. 2019) cites the case of the main culprit in the Pittsburgh synagogue shooting. Some anti-Semitic and racist posts have been uploaded on the Gab platform. The Gab platform was forced to shut down after the incident. At the same time, the Gab social platform was also rejected and uninstalled by many people at that time. Eventually, this platform was replaced by other social platforms.

The COVID-19 pandemic poses major challenges not just for the entire health system but globally. It has also sparked many online rumors, pranks and misinformation (Tasnim et al. 2020). Since the outbreak of the epidemic, some rumors and information without scientific basis have spread like a virus, and information about health protection has become the focus of rumor spreading. For example, American medical scientist (Thomas Cowan 2020) pointed out that 5G radio waves will reduce the immunity of people, leading to novel Coronavirus infection. This is based on the fact that Wuhan is the first city in the world where 5G is commercially available, and there is an outbreak of COVID-19, while Africa has no 5G base stations, and so far there has been no outbreak. In fact, some scholars have proved through scientific experiments (Uthman et al. 2020) that there is no connection between 5G and novel Coronavirus. Through the introduction of 5G mobile technology and novel Coronavirus, the origin of the disease and its transmission path were analyzed and expounded. Different studies of the two have shown that coronavirus is not a new pathogen of the disease. Different strains of the disease have been found over the years, but the most recent is COVID-19. However, 5G technology is still under development and has not been deployed anywhere in the world at all. There is no correlation between them.

In addition, other misinformation continues to spread rapidly around the world through traditional and social media. So there's a study (Kouzy et al. 2020) that quantitatively investigates misinformation about COVID-19 on Twitter. Studies have provided information on the scale of misinformation dissemination, such as the extent to which a large number of misinformation is spread in the medical field. And the approximate number of them. It is hoped that governments and public health organizations will provide strong evidence and adopt scientific methods to reduce the impact of misinformation. It also stressed the importance of taking appropriate preventive measures in order to curb this phenomenon endangering public safety. At the same time, misleading rumors and conspiracies on social media can cause massive social panic. Media coverage and public sentiment affect the public safety and security sectors of countries. For example, Chinese restaurants may have hygiene problems and Chinese tourists are bullied. These rumors led to serious trust problems (Depoux et al. 2020). At the same time, the COVID-19 pandemic is exacerbating racism and discrimination, especially against Chinese and Asians. Attacks on People of Asian descent in Germany continue to increase (Tagesspiegel 2020). There was even a rumor that "Novel Coronavirus is a Chinese virus because it originated in Wuhan". In an editorial published on 7 April 2020 (Nature 2020), the renowned scientific journal Nature called for an immediate end to the stigma of coronavirus and the avoidance of irresponsible behavior that links the virus and its disease to specific locations.

According to the above research and report, it can be proved that the number and variety of rumors and misinformation on social media has increased in the past few years. Some events on the Internet have not been proved to be true before they are spread by online media. Some of the rumors contain strong offensive remarks, denigrating the country or people. It directly affects people's normal life in real life and causes social unrest.

Previous research on rumor identification on social media

The 2012 study (Chen et al. 2012) tested a method for detecting aggression in social media text for teens. This is an approach known as lexical and syntactic feature architecture (LSF). This method can be broken down into three-part subsystems. The first part is used for basic preprocessing of the text. Second part is to detect the aggressiveness of a particular sentence. Third part is user attack detection. A combination of sentences and users' weighted values is used to determine whether a text contains offensive words or emotions. This language processing model will greatly assist in monitoring offensive language in social media texts. Other scholars studied (Kwon et al. 2013) the characteristics of online rumor propagation at the time. The research proposes a novel method to identify rumors, which makes use of the temporal, structural and linguistic characteristics of rumors. A model called periodic external shocks (PES) is proposed. This model mainly studies the temporal and structural features of rumors as well as the features of language. Analysis of temporal patterns is a new way to classify rumors.

In addition, Other research (Vosoughi and Roy 2015) mentioned some tools to detect and analyze rumors in social media at that time. And the source can be tracked and monitored. The RumorLens tool builds everything from detecting possible rumors, to categorizing rumors, and visualizing the dynamic whereabouts of rumors. But determining the target that the tool will collect, the topic of the rumor, needs to be set artificially. In short, this tool makes it possible to retrieve and tag social media rumors. Meanwhile, some language analysts see social media as an important medium for spreading rumors. The report (Zhao et al. 2015) is intended as a way to detect rumors by searching for specific words in comments about an event. Researcher wants to query and filter some rumors by using the method of retrieval and clustering. The research used the 2013 Boston bombing as a theme, which was heavily covered on social media at the time, with lots of misinformation or conspiracy theories. A public organization has made alarmist comments after its account was stolen. This kind of information has very serious consequences. Rumor comments on this aspect can be identified by the running of the detection algorithm although they cannot be detected perfectly. However, this method of rumor detection for key words provides a feasible method and an example for later scholars to conduct further research and promotion.

Many social media rumors are offensive to some countries. In 2017, two civil wars in Iraq and Syria led to a large number of refugees seeking asylum in other countries. This led to increasing social media attention at the time. Some expressed sympathy for the refugees on Facebook, but others attacked them, using words and insults. This study (Bretschneider and Peters 2017) is aimed at the collection and analysis of offensive speech in German, especially for the pattern of hate language, the dictionary has been collected and collated. This is a pattern-based method to detect text messages in offending paragraphs, including the victims mentioned. It can replace the machine learning word bag model of application sequence model. This method has a higher hit rate and accuracy compared with machine learning methods. However, it cannot be perfectly detected, only for a language type, and there are some special types of offensive remarks that cannot be identified. There have also studies of hate speech on social media in other languages (Watanabe et al. 2018). They divided tweets into three distinct, neutral ones, and the analogy does not involve rude or offensive comments. The second category is tweets with negative words, which are emotionally negative and have impolite words. The third category is offensive tweets, including racist and racist words. After a series of feature extraction and processing, the semantic patterns of tweets and emotional words were studied. Finally, Weka toolkit was used for classification, including dichotomy for words that detect whether tweets are offensive and classification for three different types of tweets (the three types mentioned above). At the same time, some indicators such as F1 score are used to detect the accuracy of its classification. This study provides a detailed explanation and research on the processing and process of collecting offensive statements. A new method to detect hate speech is proposed.

In fact, other paper (ElSherief et al. 2018) have detailed analyses of hate speech on social media. The paper presented hate speech is different. Including hate speech directed at specific individuals is defined as directed hate speech. It can also be defined as hate speech in a broad sense if it is directed against a group of people. First of all, targeted hate speech is very personal. It can be an attack, an insult, or a form of online harassment or stalking. Second, hate speech in a broad sense is usually very broad, such as discussing religion, discriminating words, or making fun of someone's nationality or sexual orientation. The lexical and semantic features of hate speech are studied. The 14th Conference on Natural Language Processing(Siegel and Wiegand 2018). Many scholars categorize offensive comments very clearly. In one study (Risch et al. 2018), naive Bayes classifier, neural network and rule-based methods were used. And by using different combinations of useful classifiers to detect and classify the severity of offensive language and their different types. Cross-validation is performed for different classifiers. The disadvantages of single model in text classification are also pointed out. Offensive language in this research is carefully classified. Certain labels such as insults, certain religions, Races, ethnic minorities, or nationalities.

One study (Pitsilis et al. 2018) developed a deep learning algorithm to identify offensive tweets on Twitter. To address this, the team analyzed the behavior patterns of Abusive Twitter users and how they differed from other Twitter users. The long and short term memory network (LSTM) classifier is used to improve its classification ability. Based on the results of this experiment, this method is useful for classifying offensive remarks in short message or phrase classes. It also shows that users often use short sentences and words to try to hide their aggression. Some other scholars(T et al. 2020) also compared the classification results of different classifiers, providing a comparative analysis and Random kitchen Sink (RKS) based approach for Offensive language Detection. In terms of data processing, they also adopted different methods. And finally, different schemes are used for the processing of text sentences. For example FastText (Bojanowski et al. 2017), Universal Sentence Encoder (Cer et al. 2018). Finally, the results show that the validity and accuracy of different classification algorithms are different, and RKS has certain advantages in processing large data sets.

In addition, some scholars (Preece et al. 2018) have conducted detailed studies on a platform for analyzing semantics. The platform consists of several different applications. One of them is The SentiSum App. The app can be used to detect the impact of popular events on social media attention, as well as the emotional bias of some tweets. Moreover, it is proved to have a good effect in detecting the influence of some anti-social behaviors or major events through concrete examples. It has been instrumental in crime prevention and policing.

Finally, some scholars(Rogers et al. 2020) collate and analyze all the machine learning algorithms about text classification in recent years. It critically discusses the key points that machine learning algorithms need to pay attention to when categorizing text and the important steps that are required to process text-type data. And performance comparisons between classifiers. Finally, some points of view are put forward by comparison. Data sets play an important role in classification. Most of the classification results are related to data sets. The core task is lexical emotion and semantic analysis.

Limitations of the previous research

admittedly, the research on rumor recognition is a relatively new field, which has many challenges. Many scholars have done a lot of research and experiments through the detection and classification of rumors and offensive language. Comparative traditional classification methods mainly involve the features of text content (such as theme, sentence length, emotional vocabulary, and time-based change rules) and user-related information. Through these data, the category characteristics of suspected rumors are mined, and the recognized rumors and non-rumor data are used as training data for classification. There are many limitations to this approach. The characteristics of classification need to be designed manually. The process is tedious and time-consuming, and the data volume and scenario application are not very optimistic. There are also deep learning methods to automatically learn effective semantic features instead of the traditional manual method to find classification features. For example, neural network is used to analyze some user behavior patterns and combine traditional methods to extract text information to identify rumors. There is also the use of controlled language processing combined with natural language processing to analyze the difference in the dissemination of false information and real information.

The primary problem is that must wait for a rumor to spread and for a large enough amount of data to have an impact before can identify it. Moreover, due to the anonymity of social media and the fact that more and more people want to express some opinions and opinions through social media, people often want to share some novel and hot information or events. To some extent, false information is more attractive than real information. At the same time, social media will maintain a neutral attitude towards false information due to some interests. This will speed up the spread of rumors and make rumor detection more difficult.

Secondly, in recent years, rumor detection has been focused on certain data sets, and certain events are generated social text data. In addition, they all aim at the same language, and most of the researches are researches and experiments of language types based on English data. Compared with English, there are not enough materials for rumor detection of other languages.

# Data collection

According to the previous chapter mentioned a lot about detection and text classification techniques. It is not difficult to find that the original text data must be cleaned before text data analysis and mining. Because the data obtained from social media is often irregular text with special characters. The data collection process of this project is divided into two parts: The collection of English Twitter dataset and Chinese Twitter dataset. The original English Twitter dataset came from Sentinel paper (Preece et al. 2018), which was gathered using Sentinel and filter by a list of keywords about “disinformation”. The dataset contains partial data from January to July 2020. Five data sets were used in this project, which are 2020-01-08 to 2020-01-14, 2020-03-04 to 2020-03-10, and 2020-03-25 to 2020-04-01, 2020-06-15 to 2020-06-22, 2020-07-13 to 2020-07-20. Chinese original Twitter data is collected using code in TwitterHarvest project (David M Rogers 2020) by modifying some parameters and registering with the Twitter developer API. Chinese Twitter data range Due to the limitation of the Twitter developer API, the data returned can only be within a certain period. The data collection is based on the aforementioned "false information" keyword list, which has been translated into Chinese and added some of the most popular topics related to China. Next, this is just original data, so the next step is to extract the data that will be used in this project. It should be noted that the data collection process in Chinese and English is quite different due to the different languages. Firstly, the data collection in English is explained. The file type of the original data is a JSON file, which needs to be read in the Notebook. And then extract the useful fields. The fields and descriptions to be extracted are shown in table form below. The first field represents the stored data name, and the second field represents the fields in the original JSON file.

Table 1 : English Twitter field

|  |  |
| --- | --- |
| fields | description |
| #infor for this tweet |  |
| **tweet\_id = data['id\_str']** | Each tweet has a unique ID as its identifier |
| **tweet\_lang = data['lang']** | This field shows the language of the tweet |
| **tweet\_time =data['created\_at\_src']** | Represents the exact time when the tweet was sent |
| **url = data['entities']['urls'][0]['expanded\_url']** | The twitter address for this tweet |
| **classification =data['entities']['classification']** | Some of the tweets stored in this field are categorized by StanfordNLP of Google, primarily for emotional categorization of tweets |
| **text = data['extended\_tweet']['full\_text']** | Tweet text |
| **sentiment = data['entities']['sentiment']** | It shows the main emotional bias of tweets |
| **swearword = data['entities']['swearword']** | Tweets may contain some swear words |
| #infor from who tweet this tweet | The user who sent the tweet included their ID, their location of registration, and the number of followers and friends they had. |
| **user\_id = data['user']['id']** |  |
| **user\_location = data['user']['location']** |  |
| **user\_friends\_count=data['user']['friends\_count']** |  |
| **user\_favourites\_count=data['user']['favourites\_count']** |  |
| **user\_followers\_count = data['user']['followers\_count']** |  |

Extracting these fields from a JSON file by writing Python functions. And then need to do some other processing, for example there are some fields that are extracted from the contents of a dictionary that needs to be split, slices of fields. Finally, the processing is done and saved as twitter data as a CSV file. Here's an example of twitter data.

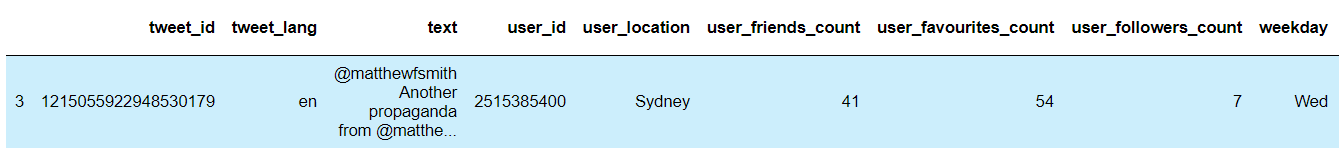


Figure 2 : twitter data\_1

社交网络的手机截图

描述已自动生成

Figure 3: twitter data\_2

Then there are several steps needed for the data of English tweet text. The processing of tweet text data includes:

Table 2 : English Twitter text processing steps

|  |
| --- |
| 1.the conversion of lowercase letters  It is necessary to convert all English characters into lowercase letters to facilitate the processing of subsequent steps |
| 2.deletion of @user and RT from each tweet  This is a handle from a tweet, usually some RT characters and @ username that must be deleted |
| 3.deletion of words with a length of less than 3  English words with less than three characters are mostly meaningless words, such as is and are |
| 4.Delete punctuation  Remove all punctuation |
| 5.removal of stop words  Because of the emotional bias analysis, the stop words have no meaning and need to be deleted. |
| 6.The part of speech reduction  It is necessary to carry out part of speech reduction to transform different expressions of the same word into the same form |
| 7.word segmentation and Part-of-speech tagging  Finally, NLTK's Twitter word segmentation toolkit was used to segment the text data and mark the part of speech of each word. |

The final processing results are shown in the data frames in the figure below.

电脑屏幕截图

描述已自动生成

Figure 4 : tweet\_text\_data

For Chinese Twitter data, Export the data stored in mongoDB into several CSV files and useful fields are extracted from CSV files and then split and integrated. Because English and Chinese are derived from different data sets, there are some differences in the fields, but the main fields to be analyzed remain unchanged. The final processing results are shown in the figure below.

社交网站的手机截图

描述已自动生成

Figure 5 : Chinese\_twitter\_data\_1

手机屏幕截图

描述已自动生成

Figure 6 : Chinese\_twitter\_data\_2

For Chinese Twitter processing steps are relatively simple.

Table 3 : Chinese Twitter text data processing steps

|  |
| --- |
| 1.Remove handles and special symbols  also need to delete the handle |
| 2.Import Chinese stop words and delete them  Chinese stop words have not been included in NLTK corpus, so a separate open source Chinese stop words library (goto456 2020) needs to be imported |
| 3.Import Chinese word segmentation toolkit and make word segmentation for tweets  The Chinese word segmentation step needs to be imported into the Jieba library (Neutrino3316 20s13). |

The final processing results are shown in the figure below

电脑屏幕截图

描述已自动生成

Figure 7 : Chinese\_tweet\_text

This is the end of Twitter data collection in Chinese and English. It is not difficult to find that the processing process of Twitter data is quite tedious, with many useless fields needing to be deleted, and a large number of emoticons, URL links, tags and other noises exist in the text data. Chinese and English text data are also very different in word segmentation. English word segmentation divides the text into one word, while Chinese word segmentation usually divides the text into two or three words.

Figure 8 shows the number of raw Twitter data collected for 5 months after processing. For Chinese data, since there is only one week in July, no detail chart display is provided.

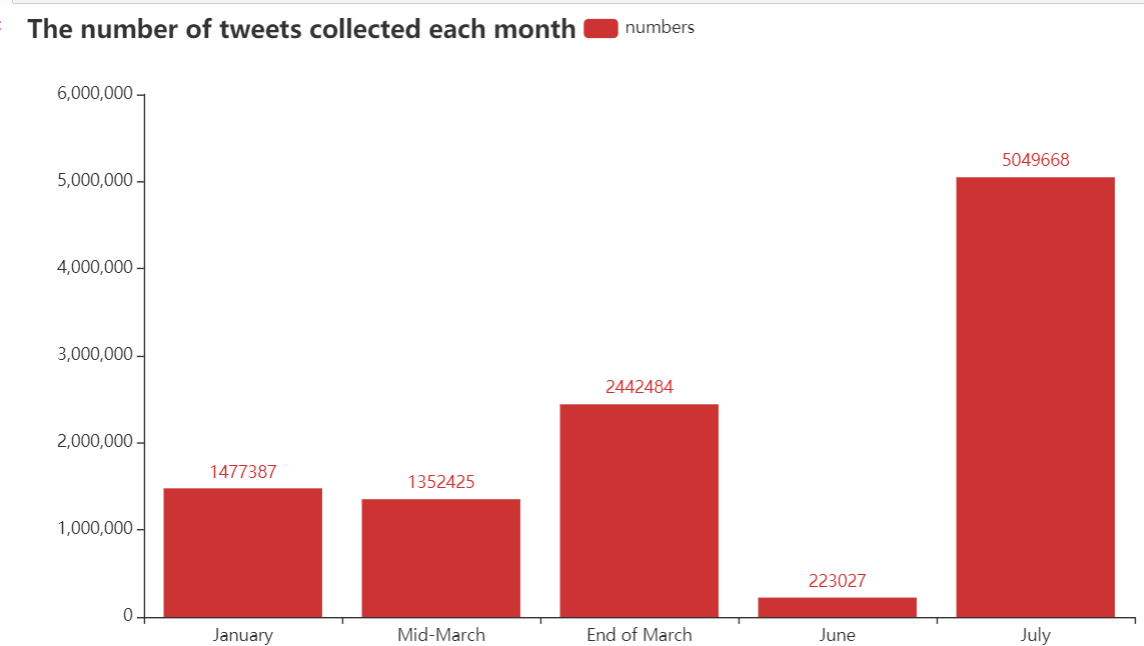


Figure 8: The number of tweets collected each month

The table below shows the original amount of data collected.

Table 4: The initial data quantity.

|  |  |
| --- | --- |
| DATASET | COUNTS |
| English Twitter data | 10,544,991 |
| Chinese Twitter data | 131,351 |

Data analysis

This research will do data analysis and visualization of the collected Twitter data. Mainly is carries on some data mining work. Based on the twitter data collected, some simple descriptive statistical analysis can be performed, for example in The English twitter data, the proportion of users in each country can be calculated according to the user's registered area, and the bar chart can be used to show the results. Or count when most users like to tweet. These are an overall understanding of the data set while processing the data, the purpose is to find useful data and information more effectively. These processes and visualizations can be found in each data\_X\_processing.ipynb file.

This project attempts to find rumors and misinformation by looking for negative and offensive Twitter. And this research will use two different methods to analyze and mine the sentiment of Twitter text. The first one is a dictionary-based approach. First, Filters text data that meets certain criteria from all Twitter texts, and then use the dictionary-based sentiment analysis tool (Thelwall 2017) to calculate the score of each tweet.

手机屏幕截图

描述已自动生成

Figure 9 : dictionary-based approach

This is done by screening out country-specific text data in advance (The countries and regions to research have been decided). Then use tools that give each tweet a clear emotional score. A negative score is negative emotion, and a positive score is positive emotion.

Table 5: Twitter text data processing

|  |
| --- |
| 1. country-specific list   This list will include keywords related to China and Hong Kong and Taiwan, USA, UK, EU, South America, Middle East and the COVID-19.( These are the topics and popular tags that are often talked about on social media now.) |
| 1. dictionary-based sentiment analysis tool   SentiStrength (Thelwall 2017) is used to discriminate positive and negative sentiment in a text. It has its own emotional dictionary.  bixin(bung 2020) and cnsenti (Thunderhit 2020). To ensure the accuracy of Chinese data, two different Chinese text sentiment analysis packages are used. |

The second approach is based on machine learning, where OLID (Zampieri et al. 2019) has data sets about offensive language. It has also been flagged for offensive words. In this way, the training data set can be processed and then the machine learning model can be built. Then use the machine learning model that has been built to directly bring the collected Twitter text and can obtain text data with offensive emotional tone. At the same time, the research prepares to use different machine learning models, find the highest F1 score or recall through multiple mediation parameter Settings, and get the best emotion classification model.

手机屏幕截图

描述已自动生成

Figure 10 : Machine learning model approach

The emotional tone of the English text data can be obtained in two different ways. Finally, prepare to experiment with two different methods, and choose one of the better methods to get the final text. But for Chinese data, it is not prepared to use machine learning methods to obtain the emotional bias of Chinese tweets. Because social media Twitter is not open for use in China, there is no suitable Twitter text data to use as training data. Therefore, the method of sentiment analysis on Chinese texts will adopt the method of sentiment dictionary matching. However, in order to ensure the accuracy of the recognition of emotion tone in Chinese text, two different emotion dictionary toolkits will be used: bixin(bung 2020) and cnsenti (Thunderhit 2020). This research will analyze the content of the text and the different emotional preferences of Chinese and English, and finally find out different themes of rumors, and then the targets of tweets in different languages and offensive tweets.

# Results and Discussion

This chapter will be divided into two parts. The first part will explain the reasons for changing the process of negative tweets. The second part is mainly the final result, most of which will be presented in the form of charts and tables.

In this chapter, the project encountered some difficulties, including the cleaning and use of additional training data sets, but in the end it abandoned machine learning methods and used other methods for data analysis. Important fields are missing, and English and Chinese data are too different. The following will explain how to solve the above problems.

Process change

The data cleaned up in the previous chapter will be sorted and saved as new CSV files. There is a total of ten CSV files (Five dataset: The data of each part is divided into a csv file of data and a csv file of Twitter text), and there are also five code files that perform the cleaning steps. The reason is that the amount of data is too large, and the function may not be executed due to insufficient memory errors, and it is too complicated and difficult to organize together. So take a separate approach to execute code files one by one (The code file is dataset\_X\_processing.ipynb, and the CSV file is data\_X.csv and text\_X.csv). Then there will be the training process of the machine learning algorithm. Finally, because the trained model is not suitable for this data set, the strategy of screening offensive Twitter is changed. And due to the differences in the structure and fields of the Chinese and English data, different methods will be used for analysis.

For English data, the entire classification process is similar to a pipeline. According to the sentiment dictionary mentioned in the previous chapter, the sentiment and offensiveness of tweets are classified for the first time. The sentiment dictionary is the first layer. The first negative sentiment tweets are selected through sentiment scores, and then the machine learning model is used for the second classification. In this process, use bag-of-words model or TF-IDF feature to select the best parameters and select the best model. Tried to use random forest, SVM, naive Bayes these three methods suitable for text classification.

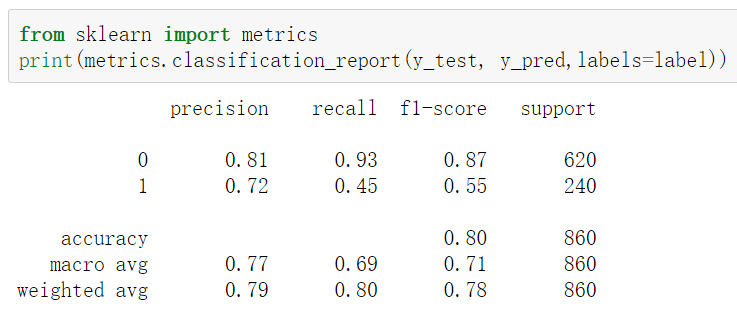


Figure 11: Classification report

After the three different algorithms adopt different parameter settings, use the training data set provided by OLID to train their model, and the result report obtained is shown in Figure 10. In this case, first of all the data itself is extremely unbalanced data, and there are many new topics and new words in the new data set. The accuracy of the classification results can also be obtained from the F1 score index. The F-1 score has high accuracy for ordinary Twitter texts of 0, but the classification accuracy for abnormal Twitter texts of 1 is very low. The result proves that the classifier can recognize normal samples well, but the recognition accuracy of abnormal samples is not high.



Figure 12:Testing model

Then put the best model into the data set in the project for testing. According to the classification results, there are still a lot of offensive comments and rumors in the data classified as normal tweets, which proves that this classification is inaccurate. of. The result is the same as the training data set. It is found that tweets classified as 0 (normal tweets) also contain a lot of offensive tweets and words. It shows that the classification of 1 abnormal text is not accurate. In other words, this machine learning training model is not applicable to current Twitter data. The process of using machine learning is saved as two files (machine learning for training data(aggressive tweet).ipynb and using model for english tweet data.ipynb)

Now, the screening method has changed instead of using machine learning methods. As shown in Figure 12 and Table the steps in the first method mentioned in the previous chapter are changed.

The first screening of countries and hot topics that need to be studied (including China, the United States, Russia, the Middle East, Europe, the United Kingdom, Taiwan, Hong Kong, South America, Africa and COVID-19), and the second screening using the above disinformation terms. Finally, it uses sentiment-based dictionary tools for sentiment analysis. The results obtained in this way will be better than those obtained through machine learning training.

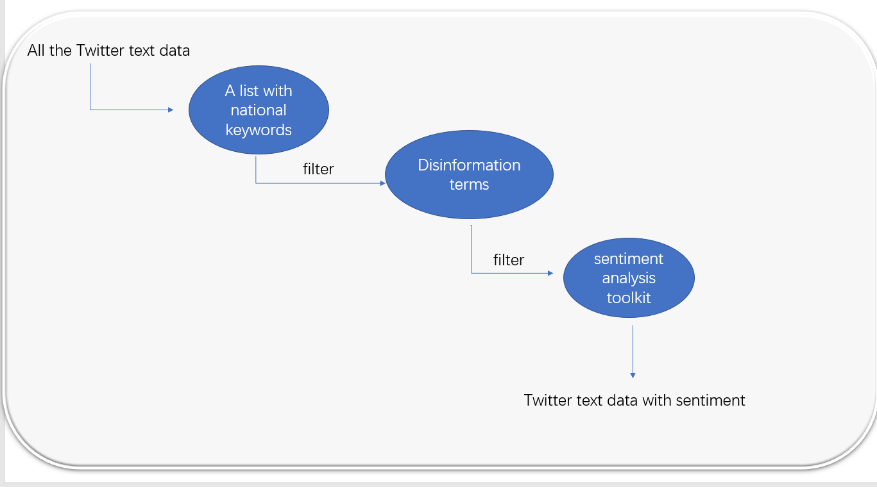


Figure 13: Final filter approach

Table 6 : final Twitter processing

|  |
| --- |
| 1. country-specific list   This list will include keywords related to China and Hong Kong and Taiwan, USA, UK, EU, South America, Middle East and the COVID-19.( These are the topics and popular tags that are often talked about on social media now.) |
| 1. Disinformation terms   disinformation = ['Fake news','Propaganda','Disinformation','Active measures','Subversion','Interference','Influence','Conspiracy','Deep state','Misinformation','Fabrication','Manipulate','Deceive','Useful idiots','Mainstream media','Populism','Untrustworthy','Hoax','Made-up','Bogus','Inaccurate','Doctored','Fact Checking','eu False','eu Fraud','eu Hoax','eu Lies','eu Rumours','eu Troll','europe False','europe Fraud','europe Hoax','europe Lies','europe Rumours','europe Troll','european False','european Fraud','european Hoax','european Lies','european Rumours','european Troll']  The Chinese disinformation terms will be shown in the appendix. |
| 1. dictionary-based sentiment analysis tool   SentiStrength and bixin and cnsenti. |

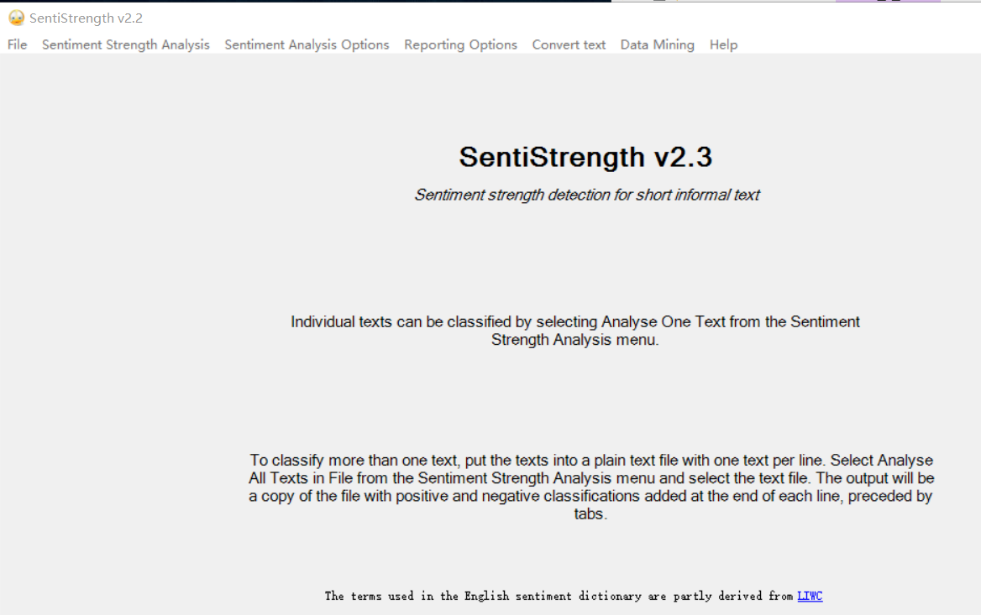


Figure 14: Sentiment analysis toolkit (SentiStrength)

Most of the data obtained in this way is related to rumors or offensive Twitter. Finally, the SentiStrength toolkit shown in Figure 13 can be used to score emotionally the text content on the social media. It is based on the scoring mechanism of the sentiment dictionary. Can detect positive and negative words. And the accuracy has reached the level of manual detection. Make each tweet text get a unique sentiment score. The higher the score, the more positive the emotion, the lower the score, the more negative the emotion. For this project, Twitter data with negative sentiments will be retained. The final process is divided into two different processes: English tweet section and Chinese tweet section (final process ----find tweets with negative sentiment (disinformation terms and dictionary matching).ipynb and tweet\_chinese\_processing and mining\_negative.ipynb)

For Chinese data. First of all, the Chinese data was thoroughly cleaned. Since the support for Chinese text in the field of text mining and python is not very high, some difficulties were encountered in processing, including the amount of Chinese Twitter data obtained was too small, and the Chinese was divided into Simplified and traditional. Therefore, only other python packages can be imported to process and analyze Chinese Twitter data. The sentiment scores are obtained through the matching of sentiment dictionaries, and the Chinese Twitter data with negative sentiment is filtered out (and to ensure that the sentiment analysis is more accurate, sentiment analysis based on two different sentiment dictionaries is used package). The lower the sentiment score, the more aggressive the tweet.

Final result and discussion

The final result can be expressed in the table below. The time period of offensive English Twitter data. It is the end of January, the end of March and the beginning of April and mid-June and mid-July. The time period of aggressive Chinese Twitter data is within one week of July.

Table 7: Numbers of twitter data

|  |  |
| --- | --- |
| DATASET | COUNTS |
| Offensive or negative English Twitter data | 1,287,138(798,250) |
| Offensive or negative Chinese Twitter data | 69359(69283) |

It can be seen that the number of English Twitter and Chinese Twitter is seriously unbalanced. But this is also normal, because China has not opened the social platform of Twitter for everyone to use. And there are some problems in these data that need to be dealt with. First, the user\_location field in the English data has a lot of missing. Because the research of this project is aimed at offensive Twitter data in different regions of each country, it is necessary to count the Twitter data from different countries, so it is necessary to delete these data with missing values. Therefore, the amount of data in parentheses in the above table does not have missing values. Secondly, many addresses in the Chinese Twitter data are not real addresses. Therefore, other methods of statistics and analysis need to be adopted for Chinese Twitter.

Quantitative analysis (General--English part)

First, the 798,250 pieces of data in the remaining English tweets after cleaning will be divided according to hot spots and statistical analysis will be performed. According to the figure 14 below, it can be divided into 12 parts. Use bar statistics to make the data more obvious. From the bar chart, there are more tweets in multiple areas than other areas. They are China (159,550 tweets), the United States (151,831 tweets), Russia (173,024 tweets), the Middle East (71978 tweets) and COVID (328,366 tweets). At the same time, it can also be found from the proportions in the pie chart that the above five regions occupy most of the space in the pie chart. These areas will be analyzed in more detail in later sections. These results come from the file: Research english text data.ipynb.

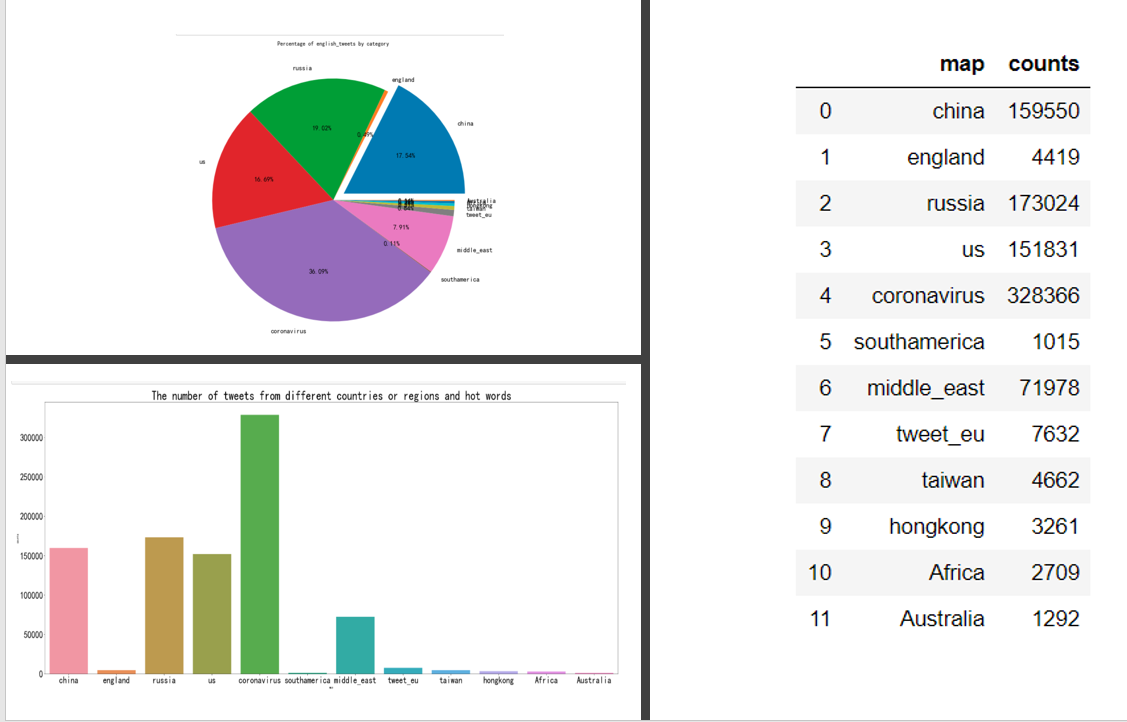


Figure 15:English\_tweet\_table\_barchart

Figure 15 is about statistics on Twitter from different countries mentioning China. According to statistics, it can be tailored in many countries. Twitter from the United States mentioned that China covers more than 70%. Followed by India and Hong Kong and the United Kingdom accounted for 11%, 5%, 4%, and other countries accounted for the rest.

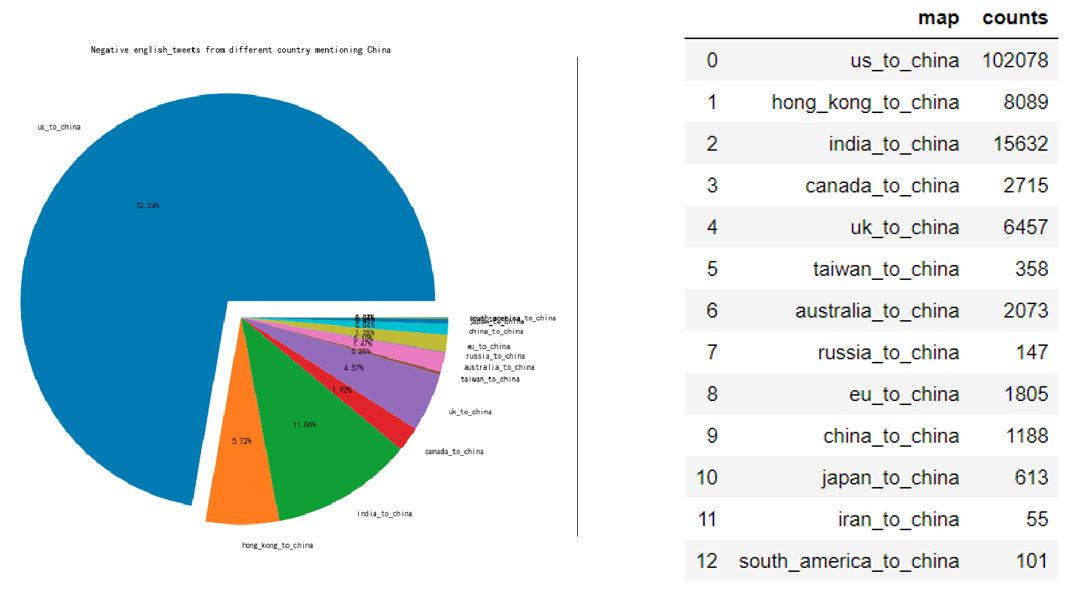


Figure 16: Negative english\_tweets from different country mentioning China

Figure 16 is a word cloud map that can show which high-frequency words exist in the Twitter mentioning China in different countries. At the same time, some tweets can be find that have appeared many times based on these high-frequency words. Several specific examples are shown on the left side of the picture (Twitter from us\_to\_china part). Finally, Russia, the United States, the Middle East, and the analysis of COVID-19 all use the same method. (the analysis process and results are stored in research english text data.ipynb).

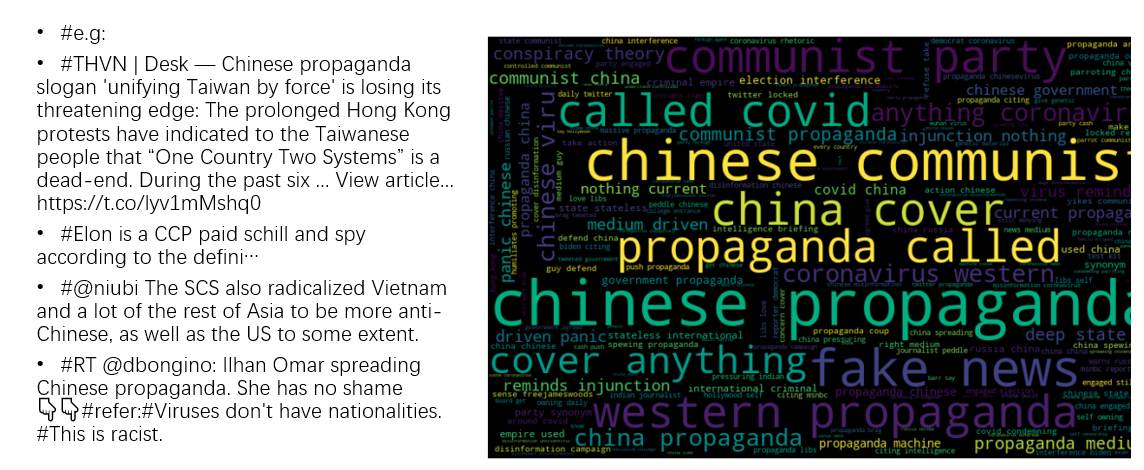


Figure 17 : mentioning China tweet\_words\_detail

Quantitative and qualitative analysis (General and detail--Chinese part)

Secondly, For Chinese Twitter, due to insufficient data and missing fields, the same research process as English Twitter data cannot be adopted. According to the figure below, it is also divided into 12 parts. From the bar graph, we can see that tweets in only three areas are more prominent than other areas. They are the United States (27248 tweets), China (27374 tweets), and Hong Kong (14704 tweets). In the pie chart, China, the United States, and Hong Kong accounted for 30%, 30% and 16% respectively. These results are from the file: research chinese text data.ipynb

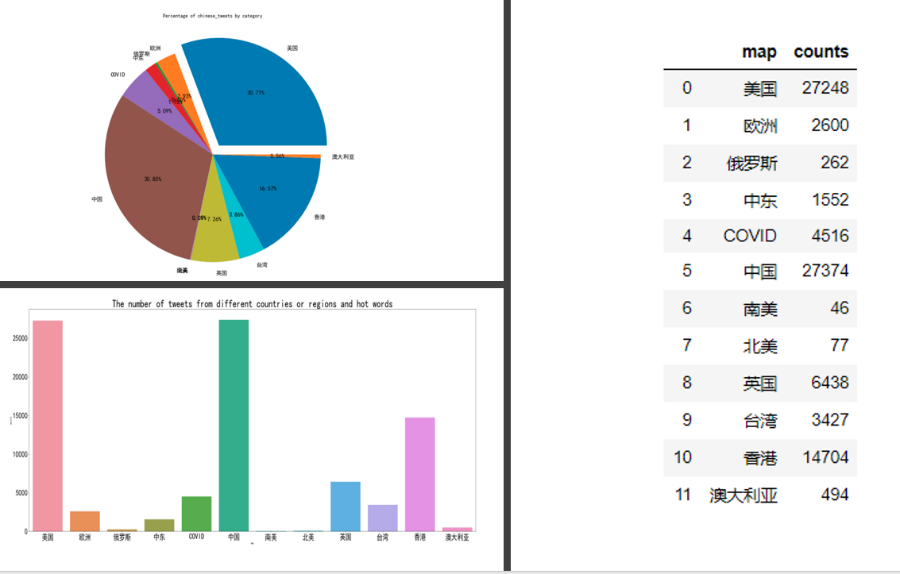


Figure 18: Chinese\_tweet\_table\_barchart

Next, due to the large number of missing or errors in the user\_location field data, and the amount of data is different, it cannot be analyzed in the same way as the English Twitter. Therefore, the following hot spots mentioned in Chinese Twitter will conduct word frequency statistics and try to analyze their content and high-frequency words.

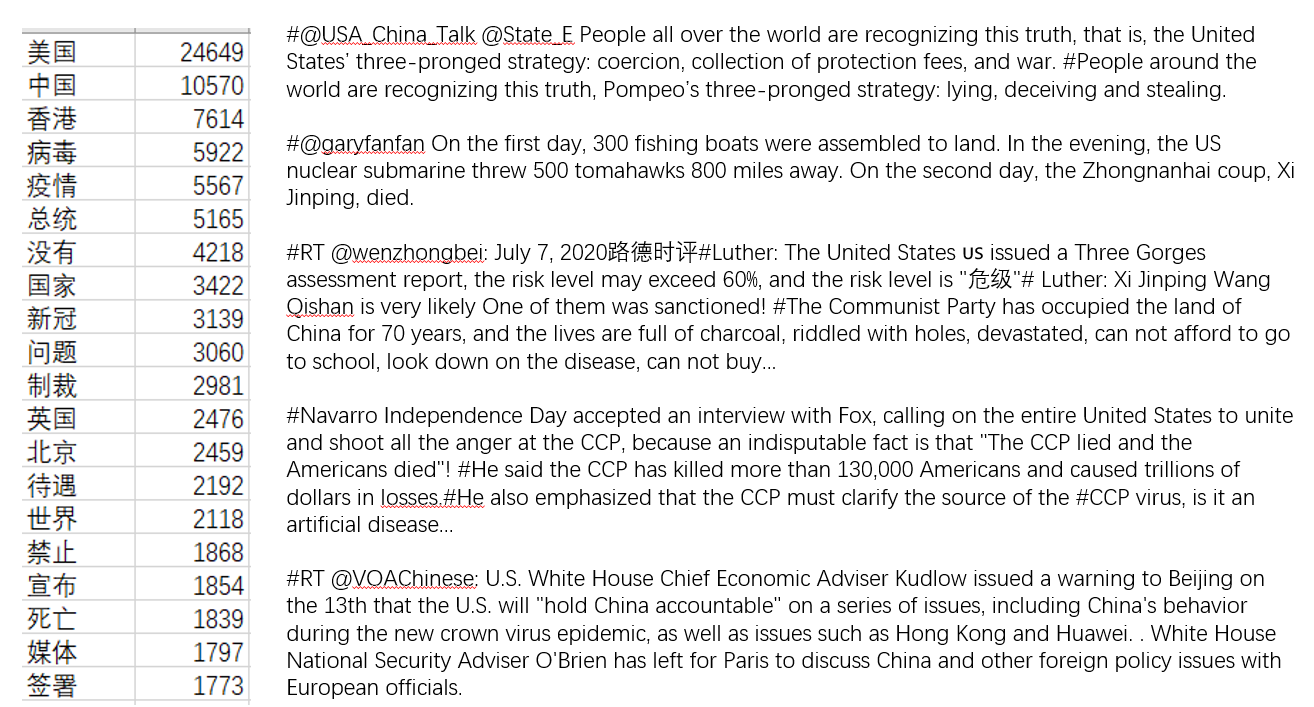


Figure 19: metion\_US\_High\_frequency\_words\_and\_tweets

The results of the specific analysis are shown in Figure 18. On the left is the statistics of high-frequency words in Twitter mentioning the United States. There are words and word counts. Among them, the United States, China, Hong Kong, as well as viruses, epidemics, and presidential words appear the most frequently. It can be concluded that most of the tweets mentioned in the US are related to the epidemic in China and the Hong Kong issue.

Through the ranking of word frequency statistics, the first few words are used to search this part of the Twitter text. Find some Twitter texts that appear multiple times. The right side of Figure 18 is their English translation text. Twitter's content mentioned popular topics between China and the United States, mainly involving political issues, including slandering the Chinese government, spreading rumors, and accusing China of not curbing the spread of COVID.

Figure 19 shows the term and Twitter text of the frequency statistics of Twitter mentioning China. These same tweets also have a high comment rate. Most of the content is related to the United States. Including vilification of the Chinese government, while the United States also praised the Chinese government's performance in the epidemic.



Figure 20: metion\_China\_High\_frequency\_words\_and\_tweets

Through a detailed analysis of Chinese Twitter, some interesting conclusions can be drawn. The most mentioned countries in Chinese Twitter are the United States and China. But in terms of quantity, the Twitter mentioning the two countries are basically the same. The tweets that mentioned the United States found more rumors, and the content was probably about politics and the launch of war. But most of the tweets that mention China are counterattacks and explanations for slander from the United States. This is the difference between them. As the number of tweets in other regions is small, no comparative analysis will be made. But it also counts word frequency and Twitter texts that appear multiple times. (In the code file research chinese tweet data.ipynb)

Quantitative and qualitative analysis (detail--English part)

Through the above general analysis, the study recorded the overall situation of English Twitter data, and also found some hot spots that require detailed analysis (China, the United States, Russia, the Middle East and the recent epidemic of COVID months). Next, these parts of the data need some special grouping. Then follow the following four steps to conduct a detailed study. #1. The data is grouped by day. #2. Find lowest point and peaks in several hot spots. #3. Emotional trends are based on the passage of time after grouping. #4. Collect some tweets to explain the reason for the peak or lowest point. The graphics will be saved as HTML files. A chart is a dynamic chart that can perform certain operations. The number of tweets is classified according to sentiment scores (-1 to -4). The lower the score, the more negative the sentiment (The score is obtained by SentiStrength).

**China part**

图表, 条形图

描述已自动生成

Figure 21：mention China’s twitter summary and sentiment trends

Table 8:Detail analysis for China

|  |
| --- |
| Information from the Bar chart.  #1. It can be seen that there are more tweets mentioned in China at the end of March and July, especially when it is low in March, there are more tweets for several consecutive days.  #2. The overall sentiment score is not very negative, but in mid-July there were more Twitter with a -2 score.  #3. The number of tweets peaked on March 28. There are 15,622 tweets in total. Among them, -1 scores accounted for the majority of 12,638 tweets, -2 accounted for a small portion of 2216, and the number of tweets with -3 and -4 scores was not much.  #4. Twitter with a sentiment score of -4 reached its peak here on 2020-03-10. The -3 score peaked on 2020-04-01, the -2 score peaked on 2020-07-16, and the -1 score peaked on 2020-03-26. |
| Information from the Line chart.  #1. The average sentiment scores of Twitter mentioned in China reached different minimum lowest point in mid-January, early March and mid-July. The average emotional score on March 10 reached -1.7, the score in mid-January was -1.6, and the score in mid-July was -1.5  #2. The reason for reaching the lower peak is due to the existence of a large number of -3 or -4 sentiment scores. Caused the score to be too low. (Note that since there are not many tweets with a -4 score and two days are missing, resulting in a shift in the value of -4 score in the figure) |
| Check the specific tweets to see the reason for these peaks and lowest point. (Frequent tweet)  **Check out the tweets on March 28 to see the reason for the peak**  #Communist China’s propaganda arm just brag-tweeted that our government just agreed to give our genetic material to their lea…  #China and Russia have seized on the coronavirus outbreak to wage disinformation campaigns that seek to undermine the U.S. and its handling of the crisis, rather than addressing public criticism of their own struggles with the pandemic  #No, China did not “cover up” anything about the coronavirus. That’s Western propaganda. It’s called COVID-19 because i…  **Pass -4 score and specific date March 10th, find the reason for the lowest average sentiment score.**  #Extremely alarming to learn that the Chinese government is stoking conspiracy theories that the coronavirus originated in the United States.  **View the day with the most tweets with a -2 score**  #The current propaganda media driven panic about the Chinese virus reminds me of FDR’s injunction “we have nothing to fear but fear itself”-media is scandalously destructive and dishonest |

**USA part**

图表

描述已自动生成

Figure 22: mention USA’s twitter summary and sentiment trends

Table 9: Detail analysis for USA

|  |
| --- |
| Information from the Bar chart.  #1. The distribution of Twitter with a score of -1 is more even than the distribution of Twitter with a score of -1 that mentions China.  #2.-2, -3 scores are more tweets. There were 8301 tweets with -2 points on June 15th, and more than 4,500 tweets with -3 points on January 8, 9th.  #3. The number of tweets peaked on January 8, with 10568. And the -3 score accounted for 50%, with 5,713. |
| Information from the Line chart.  #1. There are several lowest point, the maximum down peak is reached on January 9th, and the score on January 9th is -2.3. the score on January 8th is -2.2. The average emotional scores on March 4th, 25th and 26th are all around -1.9, and then It was -1.88 on June 15. |
| Check the specific tweets to see the reason for these peaks and lowest point. (Frequent tweet)  **Based on the -3 score and the specified date January 9th, the reason for the lowest average emotional score can be found.**  # Disgusting. As US ambassador to Russia, I became accustomed to addressing disinformation Putin propagated about Americans. I never expected that elected Members of Congress would engage in the same, making grotesque false statements about fellow Americans. Stop this nonsense  **According to the -3 score and the specified date January 8th, can get the tweets with the most -3 score.**  # MSNBC pushed Iranian propaganda tonight by reporting that Iran killed 30 Americans tonight even though the Pentagon said that that was a lie and that no Americans were killed.  **According to the specified date of the -2 score and June 15th, the reason for the most -2 tweets can be found.**  #The lib media are clawing at their faces, enraged that smart Americans are on to their scams. After decades of misinformation, & months of lies about the Wuhan virus, smart people have written them off & are totally ignoring their “advice.”Excellent.  #Association of American Physicians and Surgeons Sues FDA for “Irrational” Interference of Access to Life-Saving Hydroxychloroqu…（ No longer on twitter ） |

**Russia part**

图表

描述已自动生成

Figure 23: mention Russia’s twitter summary and sentiment trends

Table 10: Detail analysis for Russia

|  |
| --- |
| Information from the Bar chart.  #1. The tweets with a score of -1 accounted for the vast majority. There were more tweets with a score of -2, -3 on January 9, July 13 and July 16.  #2. Mention the peak of the number of tweets in Russia, there are close to 22,000 (date on July 16th), tweets with a score of -1 account for the majority, but there are also a small number of tweets with a score of -2, -3 special.  #3. During the low period of March and mid-June, only a small number of Twitter mentions Russia. |
| Information from the Line chart.  #1. There are two lowest point. The average Twitter score on January 9 was -2.5, and the score on July 13 was -1.75. |
| Check the specific tweets to see the reason for these peaks and lowest point. (Frequent tweet)  **Check the details of the peak of the number of Twitter on July 16**  #Russia propaganda project “In the Now” is making content designed for Trump-haters, the left, anti-American sentiment to engage and repost emotionally without really thinking about it. Some notable progressives have fallen for this this week.  #George Stephanopoulos asked Speaker Pelosi today, "How worried are you about Russian interference in 2020? Are we doing…  #@DiamondandSilk Let's start calling this impeachment scam against Donald\_J\_Trump by what it really is: A HATE CRIME! This scam impeachment, the Russia hoax, Mueller hoax, Ukraine hoax, all anti POTUS propaganda are hate based. Please retweet - this hashtag speaks the TRUTH!  **Check out the details of Twitter on January 9th with the lowest average sentiment score**  # Disgusting. As US ambassador to Russia, I became accustomed to addressing disinformation Putin propagated about Americans. I never expected that elected Members of Congress would engage in the same, making grotesque false statements about fellow Americans. Stop this nonsense |

**Middle\_East part**

图表, 折线图

描述已自动生成

Figure 24: mention Middle\_East’s twitter summary and sentiment trends

Table 11: Detail analysis for Middle\_East

|  |
| --- |
| Information from the Bar chart.  #1. A large number of tweets are concentrated in January. The other months are relatively small.  #2. It is also mentioned that the peak number of Twitter in the Middle East is higher, with more than 27,000. And -2.-3 emotional Twitter accounted for 50%.  #3. After the peak, there are also a large number of -2, -3 sentiment scores, but gradually decreases with time. |
| Information from the Line chart.  # The lowest peak was reached on January 8, and the average emotional score was -1.8. Except for January, there are also peaks in other months. Because the number of Twitter is not large and there is a lack of data support, no other analysis is performed here. |
| Check the specific tweets to see the reason for these peaks and lowest point. (Frequent tweet)  **View Twitter with the highest and average sentiment scores on the same day (January 8)**  #MSNBC pushed Iranian propaganda tonight by reporting that Iran killed 30 Americans tonight even though the Pentagon said that that was a lie and that no Americans were killed.  #FAKE NEWS! MSNBC Repeats Iranian Propaganda That Missile Strikes Killed 30 U.S. Soldiers |

**COVID part**

图表, 条形图

描述已自动生成

Figure 25: mention COVID’s twitter summary and sentiment trends

Table 12: Detail analysis for COVID

|  |
| --- |
| Information from the Bar chart.  #1. There are more tweets mentioning COVID in general. Slowly increase from March 4th.  #2. The peak was on March 28, with more than 22,000 records.  #3. Starting from March 4th, Twitter with scores of -2 and -3 will exist until July 20th.  #4. On June 15, there were more than 18,000 tweets with a -2 score. |
| Information from the Line chart.  #1. There are several lower peaks, and most of the emotional scores are not very low, ranging from -1.5 to -1.6.  #2. But the lowest peak is very low, reaching the lowest peak on June 15th, and the average emotional score is -1.9. |
| Check the specific tweets to see the reason for these peaks and lowest point. (Frequent tweet)  **View the specifics of Twitter on the most tweeted day (March 28)**  #One of many reasons the criminal Deep State unleashed the #Covid19 biological terror attack in an election year-introduce less secure voting methods. Fellow citizens, you are currently being viciously attacked by a [hidden enemy] in plain sight.  #No, China did not “cover up” anything about the coronavirus. That’s Western propaganda. It’s called COVID-19 because i…  #I’m watching Trump’s coronavirus propaganda press conference, so you don’t have to do that. Trump looks weaker and orange than I have ever seen. His voice was deep and harsh. He is tired and speaks slowly. Betsy DeVos is there. Dr. Fauci looked from the side of the room. #MasksNow  #Twitter deletes Rudy Giuliani tweet featuring #coronavirus misinformation and false attack on Trump’s latest nemesis, Gov. Whit…  **Check the specific situation of more than 18,000 tweets with a -2 score on June 15**  #The Far Left Fake News Media, which had no Covid problem with the Rioters & Looters destroying Democrat run cities, is trying to Covid Shame us on our big Rallies. Won’t work!  #ICMR is fighting a war on two fronts. One against COVID-19 and another against misinformation campaigns to malign government and it's effort through ICMR |

**Trend map**

图表, 折线图

描述已自动生成

Figure 26:Trend map for several countries

This is a line chart of sentiment trends in four regions. Some comparisons can be made through this chart. And get some important information. 1. During this time, Twitter sentiment scores that mention the United States are usually low. 2. Mention that China's Twitter sentiment score is not as bad as other regions. 3. The lowest point of sentiment trends in several areas. The date of mentioning Russia on Twitter is January 9.

|  |
| --- |
| **Discussion** |
| **China part**  In addition to the analysis of peak data in the table, other information can also be found. There were very few tweets mentioning China in January, but suddenly increased rapidly on March 25. Then began to decrease slowly. But it started to increase rapidly in mid-July. It can be inferred that there are relatively few tweets mentioning China in most of the time, but tweets mentioning China may begin to increase during some international topics.  The content of Twitter mainly involves some political issues in the United States and China. There are also some rumors about COVID, For example, nicknames for COVID virus, discriminatory nicknames. (rumors mention that COVID is related to some political issues). |
| **USA part**  In the Twitter text mentioning the United States, it is not difficult to find that the total number of Twitter is basically the same as the number of Twitter mentioning China. However, the data is scattered in various time periods, and its sentiment is more negative than other countries. Its peaks also increased rapidly in certain time periods, which is consistent with China's previous analysis.  Tweets involving the United States mainly involve political issues in the United States and Russia and China. The international incident concerning the killing of Americans by Iran. There are also rumors and major incidents about COVID in the United States. |
| **Russia part**  The distribution of data in Russia is similar to that in China. All of them started to increase rapidly in a certain period of time. And after a few days it began to decrease rapidly. The difference is that the peaks and lowest point of China and Russia are not in the same time period.  Russian Twitter is mainly related to the United States, mainly on the topic of "Russian interference in the US election." |
| **Middle\_East part**  The data peaked in January and lasted only a few days before it began to decline. And the peak data volume is much higher than other regions. But for most of the time, the number of tweets was very small, and the Middle East was not the focus of international public opinion for most of the time. This may be caused by international events.  This international incident mainly revolved around "Iran killed 30 American soldiers." Spread some misinformation or explain the truth of this incident. |
| **COVID part**  Because the Twitter data mentioning COVID does not distinguish between countries and regions. So the amount of data is more than that in other regions. And the topic of COVID began to circulate on Twitter in March. This is related to the different states of the COVID epidemic in different countries. Moreover, the general sentiment of Twitter related to this topic is relatively negative, and there is no decreasing trend.  The main topic is rumors spreading around. Explain COVID as a man-made event. For example, bioterrorism attacks. Or apply the impact of COVID to politics. |

This chapter conducts quantitative and qualitative research on English Twitter data and Chinese Twitter data. However, the English Twitter data is more detailed, including a large number of charts and detailed sentiment trend analysis. Due to the lack of data in the Chinese Twitter data, only word frequency statistics and content research were conducted. It can be found that there are false information or rumors in both Chinese Twitter and English Twitter, and the main content is usually political conspiracy theories or misinformation.

# Conclusion

After processing and analyzing the results in detail, the paper comes to the conclusion section. This chapter will first summarize the findings in the research analysis, as well as some research limitations and future recommendations. Finally, a project summary will be mentioned.

Innovation and Main findings

This paper has addressed three issues. At the same time, different from other scholars, new innovations have been made in the method of screening rumors, the Chinese language is collected and analyzed.

Innovation in the project

Through literature review, most scholars use machine learning or deep learning to discover rumors to identify rumors and misinformation. The project tried machine learning methods, but from the results, the project data set is not suitable for machine learning models. SentiStrength has been widely used in sentiment research and rumor detection in the past few years. Before using this tool, this project conducted two screenings, one for countries and regions, one for disinformation terms, and finally for consolidation. Carried out a layer-by-layer filtering method. This is an innovation of method and process. It provides a new idea and available method for subsequent research and investigation.

The second is the research on China's Twitter data. First of all, Twitter is not yet open in China. But there is still a lot of Chinese Twitter data on Twitter. There are few reports on the research direction of non-English Twitter data. There is less research on Chinese data. The project collected Chinese Twitter data and then conducted some analysis. It provides usable prototypes and examples for collecting or analyzing Chinese Twitter data in the future.

1. Rumors found in social media data sets.

The project tried to use error messages to filter the data, while using a specific offensive speech data set to obtain the model, and introduced project data for testing. The end result is not ideal. The model's ability to recognize rumors is still limited. Machine learning methods have become a bottleneck, and the recognition of rumors has a certain timeliness and unpredictability. Because over time, the types and topics of rumors and keywords will change to a certain extent, which will make the identification task more difficult. But through the three-layer filtering method. Finally, from the negative sentiment data of Twitter, some tweets similar to rumors and misinformation were found, and some tweets can be confirmed as rumors. For example, about COVID is a man-made or politically attempted attack, these can be clearly confirmed as rumors, but some Twitter cannot confirm its authenticity. For example, Iran killed 30 American soldiers, or Russia interfered in the American election. These tweets could not find relevant information to prove their authenticity.

1. Research on rumors in non-English languages.

This paper does some research on Chinese Twitter data. Including the understanding of hot spots and word frequency statistics. Due to the particularity of Chinese, the data collected is limited. However, some interesting points will still be discovered through analysis. China's Twitter mainly talks about sensitive topics related to China, including political issues in the United States, Hong Kong, Taiwan and China. According to the specific content of its Twitter is some false information or rumors. For example, the outbreak of war. These tweets are from inland China, Hong Kong and Taiwan or other countries/regions. They publish these misinformation and spread rumors. However, due to the lack of user\_location, the real source cannot be found. But the point is that most Chinese Twitter messages can be confirmed as rumors. Because these tweets contain a lot of misinformation, they are very inconsistent with the actual situation. For example, if China starts a war, the Chinese leader will have an accident and the United States will die soon.

1. The differences and important characteristics of rumors between countries, as well as hot topics on social media.

In the analysis of Twitter data in English, the charts and tables shown in the previous chapter are adopted. The data has been visually analyzed and content researched. The data is divided into 12 parts according to countries and hot topics. China, the United States, Russia, the Middle East, COVID. These five parts have been used as hotspots for key research. Including the statistics of the number of tweets in different time periods, and the average sentiment score based on time changes. Twitter in the five places has a peak in a certain period of time. The reason is that some major incidents occurred during that time period. For example, political issues in China and the United States, as well as the outbreak of COVID in China, Russia's interference in the U.S. election, and even the spread of COVID across the world have been used as a political weapon. Its average emotional score also reaches its lowest peak along with these major events. It is not difficult to see that today's social media has been highly politicized and has even been reduced to a place where political tools and rumors are freely spread. This is also the reason why some false information and rumors can spread quickly, because people's focus and attention are on these unusual things.

Project limitations

The study has some unavoidable limitations. In this project, there are inevitably artificial restrictions in the process of data collection, data extraction, data analysis and mining. It is divided into three parts, method limitations, data limitations, and result limitations.

Methodological limitations

When processing data, two different methods are used to obtain its emotions. One is a dictionary-based method, and the other is a machine learning method. The machine learning method was abandoned in the project, and then the first method was modified. However, some human errors cannot be avoided. Because the screening of the first-tier countries and hotspots is artificially selected, it may not select all regions or select redundant regions. SentiStrength emotional dictionary does not include recent new words, so there will be errors. For the processing of Chinese data, due to the lack of support for Chinese data, some existing sentiment dictionary packages can only be used to analyze its sentiment. This is some deficiencies in the research methods.

Data limitations

When collecting data, the original English Twitter data came from the previous few months, so only a part of the data was selected for analysis. This project is just to study the offensive comments on the country on Twitter, which requires some data screening. Therefore, the final data obtained is incomplete and subjective. At the same time, the data obtained is for negative emotions, and the positive emotion part is missing. Secondly, Chinese Twitter data is obtained through Twitter API and collected through keywords. Since Twitter is not open to use in China, it is difficult to collect a large amount of Chinese Twitter data. Although using Twitter API can return a certain amount of Chinese Twitter data.

Result limitations

In analysis section, due to missing fields in the Twitter data, only part of the data can be deleted. This will cause subsequent visualizations to be less precise. At the same time, there is no way to conduct comparative analysis due to the large gap in data volume between English and Chinese analysis. And the analysis of Chinese data is not as detailed and in-depth as English data. For the detailed analysis of English data, only five hotspots were analyzed. Other regions are missing. Although to some extent it shows the situation of most of the Twitter data in recent months, it is also incomplete. Moreover, the analysis process is the independent analysis of the region, and the comparative analysis between different regions is insufficient. In the project, many tweets similar to rumors and misinformation were found through negative tweet data. Some of them can be confirmed as rumors through simple human judgment. But there are some tweets whose authenticity cannot be judged, and there is no suitable method to confirm whether these tweets are classified as rumors.

Future recommendations

Based on the main findings and limitations of the project, this project is not perfect. There are many areas that can be improved in the future.

The detection of rumors is in addition to the machine learning methods used in the project. More methods should be used to further improve the accuracy of the model for identifying rumors. Such as RNN deep learning algorithms, and time series models. Try to collect more data in terms of data volume. This includes the increase in horizontal and vertical numbers. Not only a few months of data, but one year or two years of data can be collected, and it is not limited to tweet a social media, you can collect data from other social media or platforms. The screening of negative sentiment tweets can be done in other ways, such as the RNN mentioned above or other models. Due to the limitation of data volume and the lack of fields, the research on Chinese data is not perfect in this project. In the future, we can try to collect Chinese social media data and study the differences between social media in different countries. For the analysis of text data, you can make a dashboard in the form of charts and tables or write a complete data report to show the research results.

Project summary

1. Project experience

As a student of MSc computing and IT management, I am not very suitable for this project. But I love python and data science and want to make breakthroughs and challenges. So I tried this project and learned a lot of useful experiences and techniques. The technology includes advanced data analysis of python, data extraction and storage, and text mining, and how to use text data in machine learning. I tried to analyze the text data on Twitter social media through experience, and learned about the hot topics and emotional trends in social media in recent months, the false information and rumors in Twitter, and the false information and rumors in different countries The difference.

2. Project team.

This project is a small part of a series of projects on "Managing Misinformation on Social Media". Thanks to the project supervisor, Dr Alun preece, for organizing a weekly project sharing meeting where everyone will talk about their weekly findings and the progress of the project, as well as some of the problems that need to be resolved. This gave me the opportunity to learn and understand different projects. Then thank the staff of the Crime and Security Institute of Cardiff University. They provided all the Twitter data needed in the project. And provide and support the Twitter data collection code for other languages. Some technical difficulties encountered in the project, I also got their suggestions. Finally, I would like to thank all the members of the project team. Everyone shared their projects in the meeting. I got the opportunity to communicate and learn in the process. And it brought more different views and inspirations to my project, and at the same time exercised my project management ability and presentation skills.

# Reference

Anaconda 2020. Data science technology for groundbreaking research.a competitive edge.a better world.human sensemaking. Available at: https://www.anaconda.com/.

Bojanowski, P. et al. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5, pp. 135–146.

Bretschneider, U. and Peters, R. 2017. Detecting Offensive Statements towards Foreigners in Social Media. *Proceedings of the 50th Hawaii International Conference on System Sciences (2017)* , pp. 2213–2222. doi: 10.24251/hicss.2017.268.

bung 2020. bixin. Available at: https://github.com/bung87/bixin [Accessed: 12 November 2020].

Cer, D. et al. 2018. Universal sentence encoder for English. *EMNLP 2018 - Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Proceedings* , pp. 169–174. doi: 10.18653/v1/d18-2029.

Chen, Y. et al. 2012. Detecting offensive language in social media to protect adolescent online safety. *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012* , pp. 71–80. doi: 10.1109/SocialCom-PASSAT.2012.55.

Cikara, M. et al. 2011. Us versus them: Social identity shapes neural responses to intergroup competition and harm. *Psychological science* 22(3), pp. 306–313.

Civic, O. et al. 2019. media : misinformation , SURVEY REPORT.

David M Rogers 2020. twitterharvest. Available at: https://gitlab.cs.cf.ac.uk/scmdmr/twitterharvest [Accessed: 17 July 2020].

Depoux, A. et al. 2020. The pandemic of social media panic travels faster than the COVID-19 outbreak. *Journal of Travel Medicine* 27(3), pp. 1–2. doi: 10.1093/jtm/taaa031.

ElSherief, M. et al. 2018. Hate lingo: A target-based linguistic analysis of hate speech in social media. *12th International AAAI Conference on Web and Social Media, ICWSM 2018* (ICWSM), pp. 42–51.

goto456 2020. chinese stopwords. Available at: https://github.com/goto456/stopwords [Accessed: 17 July 2020].

Hu, Y. et al. 2018. Rumor spreading model with the different attitudes towards rumors. *Physica A: Statistical Mechanics and its Applications* 502, pp. 331–344. Available at: https://doi.org/10.1016/j.physa.2018.02.096.

Junyi, S. 2013. jieba. Available at: https://github.com/fxsjy/jieba [Accessed: 12 November 2020].

King, G. et al. 2017. How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review* 111(3), pp. 484–501. doi: 10.1017/S0003055417000144.

Kouzy, R. et al. 2020. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. *Cureus* 12(3). doi: 10.7759/cureus.7255.

Kruse, L.M. et al. 2018. Social media as a public sphere? Politics on social media. *Sociological Quarterly* 59(1), pp. 62–84. Available at: https://doi.org/10.1080/00380253.2017.1383143.

Kwon, S. et al. 2013. Prominent features of rumor propagation in online social media. *Proceedings - IEEE International Conference on Data Mining, ICDM* , pp. 1103–1108. doi: 10.1109/ICDM.2013.61.

Marvasti, A. 2018. Research methods. *The Cambridge Handbook of Social Problems* 1(3), pp. 23–37. doi: 10.1017/9781108656184.003.

Mathew, B. et al. 2019. Spread of Hate Speech in Online Social Media. *WebSci 2019 - Proceedings of the 11th ACM Conference on Web Science* , pp. 173–182. doi: 10.1145/3292522.3326034.

MongoDB 2020. Choose which type of deployment is best for you. Available at: https://www.mongodb.com/try [Accessed: 12 November 2020].

Nature 2020. Stop the coronavirus stigma now. *Nature* . Available at: https://www.nature.com/articles/d41586-020-01009-0.

Neutrino3316 2013. jieba. Available at: https://github.com/fxsjy/jieba [Accessed: 17 July 2020].

NumPy 2020. Numpy. Available at: https://numpy.org/ [Accessed: 12 November 2020].

Pang, A. et al. 2014. Negotiating crisis in the social media environment: Evolution of crises online, gaining credibility offline. *Corporate Communications* 19(1), pp. 96–118. doi: 10.1108/CCIJ-09-2012-0064.

Patton, D.U. et al. 2014. Social media as a vector for youth violence: A review of the literature. *Computers in Human Behavior* 35, pp. 548–553. Available at: http://dx.doi.org/10.1016/j.chb.2014.02.043.

Pedregosa, F. et al. 2011. Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research* 12, pp. 2825–2830.

Pete, C. et al. 2004. Crisp-Dm 1.0. *CRISP-DM Consortium* , pp. 1–76. Available at: ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf%5Cnpapers3://publication/uuid/8209D32B-855C-47AD-9C5D-9BBBF10698AB.

Pitsilis, G.K. et al. 2018. Detecting Offensive Language in Tweets Using Deep Learning., pp. 1–17. Available at: http://arxiv.org/abs/1801.04433%0Ahttp://dx.doi.org/10.1007/s10489-018-1242-y.

Preece, A. et al. 2018. Sentinel: A Codesigned Platform for Semantic Enrichment of Social Media Streams. *IEEE Transactions on Computational Social Systems* 5(1), pp. 118–131. doi: 10.1109/TCSS.2017.2763684.

Project, N. 2020. Natural Language Toolkit¶. Available at: https://www.nltk.org/ [Accessed: 12 November 2020].

pydata.org 2019. Python Data Analysis Library. Available at: https://pandas.pydata.org/ [Accessed: 17 April 2019].

Pyecharts 2020. A Python Echarts Plotting Library. Available at: https://pyecharts.org/#/ [Accessed: 12 November 2020].

Reynolds, A. 2014. *Social Media As a Tool of Hybrid Warfare*. Available at: http://portal.sinteza.singidunum.ac.rs/paper/81.

Risch, J. et al. 2018. Fine-grained classification of offensive language. *Proceedings of GermEval (co-located with KONVENS)* , pp. 38–44.

Rogers, D. et al. 2020. Real-time Text Classification of User-Generated Content on Social Media : Systematic Review. *IEEE Transactions on Computational Social Systems* , pp. 1–16.

Sandoval-Almazan, R. and Ramon Gil-Garcia, J. 2014. Towards cyberactivism 2.0? Understanding the use of social media and other information technologies for political activism and social movements. *Government Information Quarterly* 31(3), pp. 365–378. Available at: http://dx.doi.org/10.1016/j.giq.2013.10.016.

Seib, P. 2012. *Real-time diplomacy: Politics and power in the social media era*. doi: 10.1057/9781137010902.

Siegel, M. and Wiegand, M. 2018. ULMFiT at GermEval-2018-A Deep Neural Language Model for the.pdf. Available at: https://arxiv.org/abs/1801.06146.

Singer, P.W. and Brooking, E.T. 2018. *LikeWar: The weaponization of social media*. Eamon Dolan Books.

T, V.M. et al. 2020. Offensive Language Detection: A Comparative Analysis. Available at: http://arxiv.org/abs/2001.03131.

Tagesspiegel 2020. Entfesselter Rassismus in der Coronakrise. Available at: https://www.tagesspiegel.de/themen/reportage/entfesselter-rassismus-in-der-coronakrise-er-sagte-man-muesse-mich-mit-sagrotan-einspruehen/25750740.html.

Tasnim, S. et al. 2020. Impact of rumors or misinformation on coronavirus disease ( COVID-19 ) in social media Coronavirus disease ( COVID-19 ) and flow of ( mis ) information in social media. *School of Public Health, Texas A&M University* , p. 8.

Thelwall, M. 2017. The Heart and Soul of the Web? Sentiment Strength Detection in the Social Web with SentiStrength. 5, pp. 119–134. doi: 10.1007/978-3-319-43639-5\_7.

Thomas Cowan 2020. DR. THOMAS COWAN CONNECTS THE DOTS BETWEEN 5G AND THE CORONAVIRUS. Available at: https://www.bitchute.com/video/CmgSpSizVVQZ/.

Thunderhit 2020. cnsenti. Available at: https://github.com/thunderhit/cnsenti [Accessed: 12 November 2020].

Tucker, J. et al. 2018. Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. *SSRN Electronic Journal* (March), pp. 1–95. doi: 10.2139/ssrn.3144139.

Twitter 2020a. Twitter. Available at: https://twitter.com/home [Accessed: 12 November 2020].

Twitter 2020b. Twitter API. Available at: https://developer.twitter.com/en/docs/twitter-api [Accessed: 12 November 2020].

Uthman, M. et al. 2020. 5G Radiation and COVID-19 : The Non-Existent Connection. (April), pp. 2–7.

Vosoughi, S. and Roy, D. 2015. A Human-Machine Collaborative System for Identifying Rumors on Twitter. *Proceedings - 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015* , pp. 47–50. doi: 10.1109/ICDMW.2015.221.

Watanabe, H. et al. 2018. Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection. *IEEE Access* 6, pp. 13825–13835. doi: 10.1109/ACCESS.2018.2806394.

Wirth, R. 2000. CRISP-DM : Towards a Standard Process Model for Data Mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining* (24959), pp. 29–39. doi: 10.1.1.198.5133.

Zampieri, M. et al. 2019. Predicting the Type and Target of Offensive Posts in Social Media. In: *Proceedings of NAACL*.

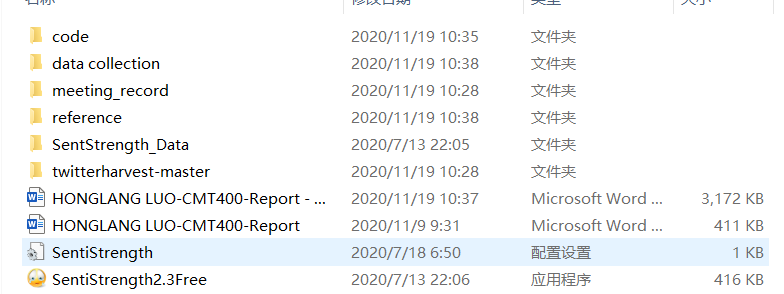
Zhao, Z. et al. 2015. Enquiring Minds : Early Detection of Rumors in Social Media from Enquiry Posts Categories and Subject Descriptors Detection Problems in Social Media. *Www* , pp. 1395–1405.

# Appendix

|  |
| --- |
| Chinese disinformation terms=[种族歧视,战争,白痴,影响,阴谋,宣传,虚假信息,虚假,事实检查,欺骗,捏造,假新闻,干扰,恶作剧,颠覆,篡改,扰乱,不准确,不信任,歧视,批评,谋杀,制裁,残害,残暴,严重,可怕,骚乱,骚动,悲剧,分裂,屠杀,企图,犯罪,仇恨,恶毒,骗局,破坏,胁迫,冲突,特朗普,美国,新冠,中国共产党,欧洲国家,英国,疫情,灭绝,消灭,叛国,煽动,反对,中国病毒,欧盟,习近平,台湾问题,香港问题,独立,台湾独立,香港独立]  The Chinese words in this list are modified according to English disinformation terms, and some high-frequency words or offensive words are added according to English Twitter data. |

This is the document structure of the project.

1. code folder-all code files, including data extraction, data processing, data analysis, text mining, and visualization code.
2. data\_collcation folder - contains mongoDB database software, OLID machine learning data set, and disinformation terms in English and Chinese Twitter
3. meeting\_record folder - It includes all meeting minutes, PPT used in the meeting, and project plan.
4. Reference - Here are all the cited articles, web pages, materials used for writing the literature review, and all the pictures of the project in the project.
5. SentStrength\_Data - Here is SentiStrength's emotional dictionary storage directory.
6. Twitterharvest-master - Here is the code for collecting Chinese Twitter data, instructions for use, and storage of Twitter API account password.
7. The draft of the project report, and the final version.
8. SentiStrength's configuration files, and applications.



This is the document structure of all datasets in the project

1. The first five folders are the English Twitter data set used in the project. Inside is a file in JSON format.
2. clear\_data folder - All data extraction, data processing, data analysis, as well as visualization charts, text mining files. The main formats are CSV, HTML, and PNG.
3. Not used folder - English Twitter data for other time periods downloaded in the teams space.
4. tweet\_china folder - All Chinese Twitter data collected through the Twitter public API has passed the CSV file exported by mongoDB.

