北京郵電大學 Queen Mary University of London

Undergraduate Project Report
2018/19

The Spread of Rumors in WeChat

Name:	Gong Jiaying
School:	International School
Class:	2015215103
QM Student No.:	151009262
BUPT Student No.:	2015212931
Programme:	Telecommunications Engineering with Management

Date: 06-05-2019

Table of Contents

Abstract

With the increasing popularity of WeChat in society, rumors in WeChat Moment (WM) appear and quickly spread among the users due to its convenient and simple sharing function. However, previous works mainly focus on rumor identification in social networks such as QQ, Twitter and Sina Weibo, little research has been conducted on rumor identification and diffusion in WeChat. To study the problem of automatic rumor identification and rumor diffusion in WM, we build an important dataset of both rumors and non-rumors in WMs. A novel Automatic Crawling and Identification System Model (ACISM) consisting of three main functions: rumor detection, feature extraction and rumor identification is proposed to detect and identify rumors in WMs in this paper. In order to construct a highly accurate training dataset, we use three authoritative platforms to detect rumors as ground-truth. A supervised mechanism is designed to distinguish rumor WMs from non-rumor WMs after Chinese word segmentation and keyword extraction and three features (title, content and account) for rumor identification are proposed. Additionally, a rumor connection framework based on Susceptible Infected Model is proposed to better present and analyze rumor dissemination in WM. User locations (IP addresses), WM page views and page life time are considered to further analyze rumor diffusion in detail. Experimental results on our important dataset indicate that ACISM we proposed in this paper can successfully identify rumors with 66.7% of accuracy rate.

Key Words: Dataset Construction, Rumor Identification, Rumor Diffusion, WeChat Moments.

摘要

随着微信在社会上的日益普及，微信公众号文章以其便捷、简单的分享功能在用户中出现并迅速传播。然而，过去的研究大多关注在 QQ、Twitter、新浪微博等社交网络的谣言鉴别上，对微信的谣言识别和传播研究较少。为了研究微信公众号文章中的谣言自动鉴别和传播问题，我们构建了微信谣言和非谣言的重要数据集。本篇文章提出了一个新的微信谣言自动爬虫鉴别系统模型（ACISM），该模型由三个主要功能组成：谣言检测、特征提取和谣言识别。为了构建高准确度的训练数据集，我们使用了三个权威平台来检测谣言作为标准答案。我们设计了一种监督机制，通过中文分词和关键词提取，区分谣言和非谣言信息，提出了谣言识别的三个特征（标题、内容和作者）。除此之外，我们还提出了一种基于疾病传播模型的谣言网络传播框架，以更好地呈现和分析微信公众号文章中的谣言传播。我们考虑了用户位置（IP 地址）、微信公众号文章页面浏览量和页面存在时长等因素，对微信谣言传播进行了详细分析。 基于我们提出的数据集的实验结果发现，本文提出的 ACISM 能够成功地识别谣言，其准确度达 66.7%。

关键词：数据集构建，谣言检测，谣言传播，微信。

Chapter 1: Introduction

WeChat is a mobile social media launched by Tencent in 2011, which has played an increasingly significant role in society and the emerging market is becoming mature. According to Sensor Tower, WeChat application downloads have surpassed 28 million during the first season in 2018 and there are more than one billion active users of WeChat [1]. WeChat, similar with other social media such as Sina Weibo, QQ, or Twitter, is becoming one of the most popular social networking services in China.

"WeChat Moments" (WM) is a kind of social networking function in WeChat, allowing users to freely post and share passages they are interested in anywhere and anytime. Users only need to click the link (Figure 1a) on top right corner of the page to share the information (Figure 1b), due to this convenient and simple sharing function, WM facilitate rapid diffusion of information.

Figure1a A Posted Link

Figure1b The Share Button

Figure1c The Share Platform

However, as the platform has become more popular, and sizes of users have grown, the reliability of its system of trust has decreased. Rumors appear and quickly spread among the users, which is a common problem in most social media apps such as Twitter [8], Facebook [9] and so on. According to Merriam-Webster, rumor is a statement or report current without known authority for its truth [2]. Usually we define a rumor to an unverified assertion that starts from one or more sources and spreads over time from node to node in a network [3]. In this paper, we define a rumor as a collection of WeChat Moments posting fake news and propagating through WeChat. There exists a large amount of fake news in WeChat Moments since users can freely repost and share unverified information in WeChat. Fake news can pervert people's mind, disrupt social order or even reduce government prestige. Though WeChat Official Account Platform Anti-rumor Center [4] has cooperated with a number of

professional third-party institutions to participate in anti-rumor work, it costs much time and efforts to manually assess the truth of WMs by manpower. Besides, I argue that spread of rumors in WMs can be better prevented through analyzing the diffusion of rumors in advance. Therefore, we are motivated to focus on real traffic in WeChat, identifying rumors automatically and studying how rumors spread across its users in order to limit the spread of rumors. Towards this direction, we face three major challenges, namely.

- Most traffic data in WeChat are encrypted due to privacy and security. It is hard to detect rumors in text messages of WeChat. And there is no dataset in public for people to do research on.
- Few Chinese platforms for rumor identification exist. Hence, it is necessary for us to automatic detect more rumors.
- Only considering titles of pieces of information is one-sided since the titles are always exaggerated and astounding in order to catch people's eyes. Therefore, it is not feasible to determine whether the WM is a rumor or not by only taking account of titles of WMs.

To address these challenges, our research only focus on spread of rumors in WMs because traffic data in WMs is not encrypted and WMs is a more usual and easy way to propagate fake news in WeChat since users only have to click the forwarding button. Besides, we propose an Automatic Crawling and Identification System to identify more rumors in WMs. This model mainly use three authoritative platforms to help detect rumors and then three features (titles, contents and accounts) are extracted from each WM. Based on this model, we also design a model to differentiate rumor WM with non-rumor WM after Chinese word segmentation and keyword extraction. In addition, three new features are provided for automatic rumor identification, which can help provide a more comprehensive and objective result. Experiments on dataset of WMs we have constructed show that our proposed algorithm can successfully identify rumors in WMs. Furthermore, we show many graphs to explain the spread of rumors in WeChat, including user locations, WM page views and page life time, which provides the basic for further analysis on rumor evolution.

In summary, the main contributions of this paper include:

- We build an important dataset of both rumors and non-rumors in WMs, including Uniform Resource Locations (URLs), titles, contents, accounts and timestamps. To the best of our knowledge, this is the first dataset of rumors and non-rumors in WMs.

- We propose a novel Automatic Crawling and Identification System Model to detect and identify rumors in WMs. This model can automatically extract key information and check rumors. Three features for rumor identification are proposed in this paper.
- We analyze rumor diffusion based on Susceptible Infected Model and user location, WM page views, page life time to analyze the spread of rumors in WMs in detail, which contributes to rumor evolution in future work.
- Experiments on the important dataset we have provided show that our proposed algorithm performs well on rumor identification.

The rest of this paper is organized as follows. Related works of rumor identification and rumor diffusion are introduced in Section 2. In Section 3, we present a novel Automatic Crawling and Identification System Model for rumor identification and the process of rumor spreading. Besides, we also describe the new dataset construction in Section 3. The evaluation results and analysis are shown in Section 4. In Section 5, we conclude the paper with future work.

Chapter 2: Background

Related works include work on natural language processing (NLP) about word segmentation and keyword extraction, work on rumor identification in social networks and work on networking about the diffusion and spread of rumors in WeChat. In this section, we summarize the research most related to our work and provide these works in three main areas: event detection, rumor identification and rumor diffusion.

2.1 Event Detection

There are various works on event detection and extraction. Neural Model [5] is popular in event detection and summarization. It can filter out the irrelevant information, measure the similarity through Term Frequency–Inverse Document Frequency (TFIDF), and perform relevant classification by using a global shared representation. Neural Storyline Extraction Model [6] is also popular in extracting specific events. They train word2vec to represent each entity and use TFIDF to filter out less important keywords. [7] used some machine learning techniques including keyword extraction, keyword community detection and pre-trained classifiers to detect events. Though these works outperform many other methods on event detection, they only apply their innovative methods in one specific field. They use TFIDF to do keyword extraction and Support Vector Machine (SVM) to achieve event detection. Their work enlightens us to perform these state-of-the-art methods such as TFIDF and machine learning techniques such as SVM on rumor identification, which is a better fit for the current scenario because rumors detected by authority are not enough for us to quantitatively analyze rumor diffusion, hence we need to use SVM to identify more potential rumors in WMs.

2.2 Rumor Identification

There are some relevant works on rumor identification in social networks such as Twitter, Sina Weibo, Stock Forum so on. Some platforms like Twitter delete bot/malicious accounts. [8] used edge weighting and centrality measure weighting to identify rumors in online social network Twitter. Automatic rumors identification has been done on Sina Weibo [9]. They proposed a new method to automatically annotate data from Sina Weibo and they also propose three new features for rumor identification in social networks. In addition, machine learning techniques are used in automatically identifying rumors in Stock Forum. [10] Crawling tasks of massive Internet forum data with smart computer technologies were successfully carried out to automatically identify rumors in forum. These works only focus on rumor identification

in social networks. There is few research studying rumor identification in WeChat. We apply similar method to WMs in order to identify rumors, which is more appropriate than other social networks because automatic identification is more suitable for rumor checking in WM due to its fast diffusion and relatively fixed mode. The most related work is [11]. They proposed a new algorithm to learn the latent factors of each information and then the learned factors are used to predict the rumors. Then they use the dataset of rumors labeled by WeChat to verify that their proposed algorithm can successfully identify 61% of rumors. However, the precision rate is very low. We try to improve the precision rate of rumor identification in WeChat through our proposed methods and algorithms.

2.3 Rumor Diffusion

Some researches have focused on diffusion of WeChat Moments. [12] analyzed how WeChat Moments spread by considering lots of factors such as number of views, path length, users' locations and so on. They used the results to help them develop marketing strategies. Similar method can be used for us to analyze spread of rumors in WeChat Moments. [13] developed models to analyze lifecycle of web applications in WeChat, which can be used to analyze the spread of rumors in WeChat Moments. [14] analyzed various elements in WeChat Moments such as number of viewings, IP addresses, spread path, users' locations and so on, leading to a systematic study on online reposting behaviors of users in WeChat and the influence on computer networks. Nevertheless, these works pay much attention on diffusion of any information in WeChat instead of only rumors spread in WeChat Moments. Features of rumor spread in WeChat [15] are proposed to help control and manage the diffusion of rumor. Spreading features like time and location are studied in different diffusion levels in order to provide a more authentic result. These related works present state-of-the-art methods to help analyze spread of rumors in WeChat Moments. However, most of these works only use simple graphs to show rumor diffusion in social networks, this is not persuasive enough to analyze rumor diffusion. Instead, we try to use Susceptible Infected (SI) model to better present information (rumor) dissemination in WM.

Chapter 3: Design and Implementation

This section introduces the general pipeline of Automatic Crawling and Identification System Model as shown in Figure 2. As can be seen in the figure, the model can be divided into five main steps: Dataset Construction, Rumor Detection, Feature Extraction, Rumor Identification and Rumor Diffusion. Each step is described in detail as follows.

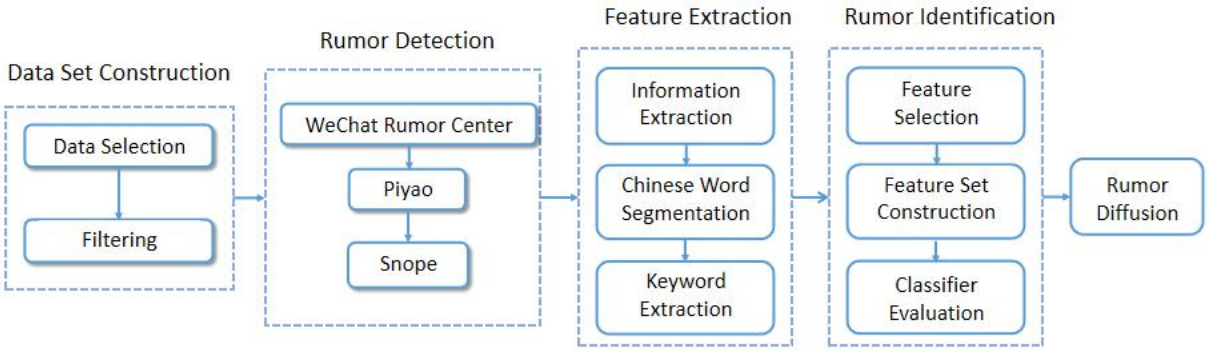


Figure 2 Automatic Crawling and Identification System Model

3.1 Dataset Construction

In this section, we aim to collect both a training dataset and testing dataset including both rumor and non-rumor WMs. Thus, a comprehensive and representative dataset is the basis for quantitative analysis of rumor identification and rumor diffusion. We make effort to carefully build such a dataset containing the following five major parts: URLs, titles, contents, accounts and timestamps. Due to privacy problems, we don't provide corresponding IP addresses in public. The format of the dataset in described as follow:

$$\langle url, t, c, a, T \rangle \tag{1}$$

Where url is the http address of the web page of WM, which is used for generating the web page of WM; t represents the title of each WM and c represents the whole content of each WM; a is the account or organization in responsible of the corresponding WM; T is the timestamp of WM, which is the local time when the WM is visited.

In order to identify rumors in WMs, we first collect traffic data about different WeChat usage with IP source addresses and timestamps from Sep.1st, 2018 to Sep.26th, 2018. After filtering out those URLs that cannot be resolved, we collect 109404 URLs about WMs (mp.weixin.qq.com). Next, in order to get a highly accurate training dataset, we collect the URLs of confirmed top popular rumor WMs in WeChat Official Account Platform Anti-rumor Center [4]. Meanwhile, we also collect URLs of the same number of top popular non-rumor

WMs in the non-rumor dataset. Then, we use the method introduced in Feature Extraction to extract the titles, contents and accounts to be part of our dataset.

3.2 Rumor Detection

First, we use time difference to detect rumors. If the URLs are fresh enough, we can't find any rumors because official platforms need time to identify rumors. In general, the average delay between the reported time of a suspicious rumor and its judge time is more than 24 hours [9]. Hence, we wait for 24 hours, and then check the WMs through three main platforms that aggregate trustworthy external sources, which are mainly from governmental or official organizations, for detecting rumors. We provide a brief introduction of each platform below.

WeChat Official Account Platform Anti-rumor Center: This center has cooperated with professional organizations in various fields, to identify rumors spread over WeChat, and to popularize the facts behind the rumors through the anti-rumor mechanism [4].

Piyao.org.cn: is a website with the functions of rumor report, rumor verification and authoritative rumor refute information. It is the most authoritative rumor refute website in China [16].

Snopes.com: is a website that documents Internet rumors, urban legends, and other stories of unknown or questionable origin. It is a well-known resource for validating and debunking rumors [3][17].

WeChat official account platform directly label WMs as rumors if they contain fake news. For example, in Figure 3a, the circled information is a label made by WeChat, indicating that this article is identified as a rumor. In Figure 3b, the words in the middle means that the WM page is deleted by the account, indicating that this WM may probably be a rumor. Figure 3c shows that there is something wrong with the web page, and nothing is found in this URL.

Figure3a Rumor Labeled by WeChat

Figure3b Deleted Example

Figure3c 404 Not Found Example

Now the dataset contains two parts: marked rumor WMs and unmarked WMs. For unmarked WMs, we can find similar information or reports in other authoritative platforms such as piyao and snope. Next, we crawl the titles of each WM and the titles in any links of each WMs since these links can also lead to rumors. Then we encode these titles into URLs and build request HTTP URLs in order to automatically check these WMs in Chinese through Piyao.org.cn. Besides, we also pass all URLs through www.snope.com to check whether any WMs in English are fake news. As for every WM in the unmarked dataset, if we find similar information in these two platforms, it will be removed from the dataset, otherwise, it will be left in unmarked dataset. After first detecting the 78088 WMs, we find that 29 of these WMs are labeled as rumors by WeChat Official Account Platform Anti-rumor Center, no WM is labeled as rumor by Piyao and no WM written in English is labeled as rumor by Snopes. Detailed results will be described in Section 4.2.

3.3 Feature Extraction

3.3.1 Information Extraction

In this subsection, we use Beautiful Soup [18], a HTML parser written in the Python programming language designed for screen scraping, to extract three kinds of information (title, contents and account) in each WM of both rumors and non-rumors classified in Rumor Detection. After we crawl the source code of the webpage of WM, we analyze the webpage source code and find out the corresponding tags of title, content and account. Then, we use Beautiful Soup to extract the information in these three kinds of tags. Take the account tag as an example, figure 4 shows part of the webpage source code in one WM.

```
<div id="js_profile_qrcode" class="profile_container" style="display:none;">
  <div class="profile_inner">
    <strong class="profile_nickname">青岛市旅游发展委员会</strong>
    <img class="profile_avatar" id="js_profile_qrcode_img" src="" alt="">
```

Figure 4 Part of webpage source code

We can easily find that the account is in the tag strong class. Hence we use *soup.find('strong', class_="profile_nickname").get_text()* to extract the account of each WM. In this method, all the information we need can be successfully extracted from each URL of WM.

3.3.2 Chinese Word Segmentation

Because of Chinese special characters, there is no space between words. We need to do segment Chinese sentences into words in order to find keyword. We segment both the title and content of each WM using Jieba Chinese Word Segmentation Utilities [19], which outperforms other Chinese Word Segmentation techniques such as Stanford Chinese Word Segmenter [20], SnowNLP [21] and so on. Compared with Stanford Chinese Word Segmenter and SnowNLP, which need to occupy a lot of memory space and CPU, Jieba runs faster with lower memory. Besides, Jieba also supports the function of adding new words to the custom dictionary by developers in order to increase accuracy rate, though it has the function of new word identification ability. Table1 shows the advantages and disadvantages of different Chinese Word Segmenters. Jieba word segmentation algorithm segments the sentence based on prefix dictionary, generating a Directed Acyclic Graph (DAG) according to segmentation positions. Then it uses dynamic programming to find the maximum probability path. For unknown words, Jieba uses the Hidden Markov Model (HMM) based on ability of Chinese

word formation and Viterbi algorithm to segment the words.

Due to these advantages, Jieba has proved to provide outstanding performance on Chinese Word Segmentation tasks. Therefore, we choose to use Jieba to do Chinese Word Segmentation before keyword extraction task.

Table 1: Comparison of different Chinese Word Segmenter

Name	Advantages	Disadvantages
Jieba Chinese Word Segmentation	Runs faster, lower memory Custom dictionary New word identification	Do not have Part-of-Speech Tagging System
Stanford Chinese Word Segmenter	Better accuracy rate	Occupy much memory space and CPU
SnowNLP	Support simplified and complex chinese characters conversion	Slow

3.3.3 Keyword Extraction

One of the most critical process of the Automatic Crawling and Identification System is keyword extraction. Extracting keywords from each WM is extremely essential to the performance of the entire system. We use Term Frequency–Inverse Document Frequency (TFIDF), a numerical statistic that is mainly reflects the importance of the feature item in the document representation [22], to calculate the feature weight in both titles and contents in each WM to do feature extraction for further machine learning techniques in rumor identification. Hans Peter Luhn (1957) put forward the first form of term weighting frequency (TF) and Karen Spärck Jones (1972) proposed Inverse Document Frequency (IDF). Salon proposed TF-IDF algorithm and repeatedly proved the effectiveness of the TF-IDF algorithm in information retrieval [10]. The main idea of TF-IDF as follows: If a word in one document appears much more frequently than it appears in other documents, it can be considered that the word has good ability of classification.

In a given WM, term frequency is the appearing frequency of a term in one WM, which is the normalization of word counts. The value of weight can be calculated as follows:

$$\mathrm{tf}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{2}$$

Where $n_{i,j}$ is the number of times that term t occurs in WeChat Moment WM_i, and $\sum_k n_{k,j}$ is the sum of all words appearing times in WM_i.

At the same time, if there are fewer WMs of term i, the larger inverse document frequency is, meaning that term i has a better ability of classification. The inverse document frequency can be defined as measurement of how much information the word provides.

$$idf_i = \log \frac{N}{df_i} \tag{3}$$

Where N indicates the total number of WMs and df_i indicates the number of WMs containing term i.

In order to deal with the phenomenon that df_i has the probability to be zero, which means that there is no WM containing the keyword, the formula can be amended as:

$$idf_i = \log \frac{N}{df_i + 1} \tag{4}$$

According to the idea of TF-IDF model above, TF-IDF can be calculated as follows:

$$TF - IDF_{i,j} = tf_{i,j} \times idf_i \tag{5}$$

Therefore, if a word in either rumor WM or non-rumor WM appears in high frequency and appears in low frequency in other WMs in the whole dataset, then the value of TF-IDF is high.

3.4 Rumor Identification

Rumor Identification can be considered as a classification problem of machine learning (supervised learning), whose task is to train a function from the training data. As for classifications of rumors in WM, the selection of classifier is of great importance. In this paper, we choose and compare Support Vector Machine (SVM) [23], a method for supervised learning that can analyze data for binary classification, Naïve Bayes [24], a simple probabilistic classifier based on Bayes theorem and Decision Tree [25], a basic classification and regression method that uses a tree-like model of decisions and their possible consequences, to classify WMs as rumors or non-rumors. SVM projects the vector to a higher

dimension space and represent the decision boundary through a subset of training data, which contains pairs of input features and desired outputs. Then, an algorithm is used to analyze the training data and refers a function, which can be used for labeling the testing data. Naïve Bayes only needs a small scale of data to estimate parameters for classification, which can be quite useful if our dataset is not large enough. Decision Tree classify instances into different categories based on input features. Therefore, the input features are very important since different features can lead to different binary results.

3.4.1 Feature Selection

As mentioned before, we consider automatic rumor identification problem as a binary classification problem. It is extremely important to choose suitable features from each WM and build an effective feature set in order to identify rumor diffusion in WeChat. We propose three features for rumor identification in this paper.

First, title is an essential feature for each WM. Authors usually invest a great deal of time and work in the title of articles because a good title can attract the attention of readers. In addition, title sometimes is the outline of the whole passage, representing the central idea of the passage. Hence, it should be heavily weighted in the input feature. However, only considering title as the feature is not considerate enough because the titles are always exaggerated and astounding in order to catch people's eyes. Therefore, we take account of content information from each WM because content is a more comprehensive overview for the WM. Furthermore, account of each WM is also playing an important role due to "brand recognition", which means that checking the source of each WM can not be neglected. There is a higher possibility that the rumor WMs written by the same accounts may also be fake news. It is hard to check the source automatically, so we manually check the accounts from each WM including both rumors and non-rumors.

3.4.2 Feature Set Construction

After Chinese word segmentation and TFIDF is done, each word has a weighted value. We separately consider title and content of each WM since they have different representative of each WM. The general evaluation criteria is that positive value represents rumor information and negative value represents non-rumor information. Hence, in the training dataset of rumors, each TFIDF value is positive whereas the value is negative if it is in the training dataset of non rumors. For title feature, after we calculate the TFIDF of each word, we add the TFIDF value together if the word appears in the corresponding title of WM. For content feature,

similar method is used to calculate the TFIDF of content in each WM. However, due to the fact that content has much more words than the title in a WM, we add different weight to title and content of each WM. In this paper, to make the feature data balance and in the same scale, and we divide the TFIDF value of content by 10.

For the feature of account, we manually check the accounts of each WM since there is no quantitative way to represent each account feature automatically. Similar with the criteria before, positive value represents rumor and negative value represents non-rumor information. Additionally, a higher positive value means a higher possibility of rumors. Table 2 shows our scoring mechanism. Value 5 means the account is deleted. It has the highest possibility that the WMs this account released are fake news because an account will be deleted only under two circumstances. One situation is that the owner actively close the account, however, more often owners will transfer the account, which can provide other information. The other situation is that the account is deleted after two many fake news tip-offs. Value 2 means the account is not available temporarily, having the chance of providing potential rumors. Some times the account can be temporarily closed in order to rectify false information. Value 1 means irregular account. These accounts are created individually without any official authentication, having the chance to release rumor WMs. Value 0 is neutral account, which means it is hard to distinguish rumors from non-rumors only considering the name of the account. Value -1 is the company or organization's official account such as Tencent Technology. This value is negative because company's official account will check each WM many times before releasing the WM. Value -2 is national official account such as Xinhua, which can hardly release rumors because these official accounts represent the image of a country. They will check the truth of information from an official channel.

Table 2: Scoring Mechanism of Account Feature

Value	Meaning
5	The account is deleted.
2	The account is not available temporarily.
1	Irregular account.
0	Neutral account.
-1	Company or organization's official account.
-2	National official account.

3.4.3 Evaluation Metrics

Making use of an open source machine learning library Scikit-learn [25], we train the classifiers by using top 99 public popular rumor WMs and the same number of non-rumor WMs. Then, we use the 12 rumor WMs labeled in rumor detection and the same number of non-rumor WMs as testing dataset. To test and ensure the accuracy of the classifier for rumor classification, we evaluate and compare the classifiers after sample testing. For assessing the performance of the proposed method and features, we use Precision, Recall, F-measure and Accuracy as evaluation metrics. Precision is the measure of quality that is defined as the number of correctly detected rumors out of all detected rumors. Recall is the measure of quantity that is defined as the number of correctly detected rumors out of all ground truth. Accuracy is the measure of quantity that is defined as the number of all correctly predicted results out of all testing samples. F-measure is defined as the weighted harmonic average of both precision and recall, reflecting the quality of the model. The specific formula is expressed as follows:

$$\Pr ecision(P) = \frac{a}{a+b} \tag{6}$$

$$\text{Recall}(R) = \frac{a}{a+c} \tag{7}$$

$$F-measure(F) = \frac{2 \times P \times R}{P+R} \tag{8}$$

$$Accuracy(A) = \frac{a+d}{a+b+c+d} \tag{9}$$

Where a is the number of correctly detected rumors; b is the number of wrongly detected rumors; c is the number of wrongly detected non-rumors and d is the number of correctly detected non-rumors.

3.5 Rumor Diffusion

A rumor WM can be released by an organization's official account or an individual account, then the friends in the same WeChat Network can view the rumors and they are potential to repost the rumor WM. Supposing that user i and j are friends in WeChat Network. If user i shares a link of rumor WM in WeChat Circles, user j has the probability click the shared

link to view the fake news and further repost the rumor WM to his friends, resulting in rumor diffusion. This process is similar to the spread of infection so that this rumor diffusion process of rumor WM can be described by a Susceptible Infected Model (SI Model) [27]:

- Susceptible: A user is susceptible before infection, which means this user is possible to be infected by his friends. We call a user a susceptible user of a rumor WM if he is the friend of an infected user who reposts the rumor WM link.
- Infected: A user infected with rumors and he has the probability to share the rumors with his friends. We call a user an infected user of a rumor WM if he views or shares the rumor WM in his WeChat Circle.

When susceptible nodes are infected, they can not be recovered. Supposing that only one node is infected at time t_0, which is the transmission source of rumor WMs. Other nodes in the WeChat Network are in a susceptible status. The source node infects its neighboring nodes at a specific rate of infection. Then these infected neighboring nodes randomly select their neighboring nodes and propagate the rumor WMs, so that rumors start to diffuse in the WeChat Network at the same time. The dissemination mechanism of rumor WMs can be expressed as follows:

$$S(i)+I(j)\xrightarrow{\lambda} I(i)+I(j) \tag{10}$$

Where $S(i)$ represents susceptible users whereas $I(j)$ represents infected users and constant λ is spreading probability. Equation 10 means when susceptible nodes in WeChat Networks contact with infected nodes, they are possible to change to infected nodes. We assume that each node has equal chance to contact with other nodes in WeChat Network based on mean field theory. Hence, based on information dissemination dynamic theory:

$$\begin{cases} \dfrac{ds(t)}{dt} = -\lambda i(t)s(t) \\ \dfrac{di(t)}{dt} = \lambda i(t)s(t) \end{cases} \tag{11}$$

Equation 11 means that proportion of infected users $i(t)$ in WeChat Network increases with time t at the speed of $\lambda s(t)i(t)$; Proportion of susceptible users $s(t)$ in WeChat Network decreases with time t at the speed of $\lambda s(t)i(t)$. At any time in spreading of rumors, total nodes do not change:

$$i(t)+s(t)=1 \tag{12}$$

Combining equation 11 and equation 12, we can get:

$$\frac{di}{i(t)(1-i(t))}di = \lambda dt \tag{13}$$

The solution for equation 13 is:

$$i(t) = \frac{1}{1+(\frac{1}{i_0}-1)e^{-\lambda t}} \tag{14}$$

Hence, the final coverage of rumors is:

$$\lim_{t\to\infty} i(t) = 1 \tag{15}$$

Which means that no matter how small the initial rumor coverage and spreading probability is, the final rumor coverage approaches to 1.

In order to find the time t_m when $\frac{di}{dt}$ is maximum, growth of dissemination node is:

$$v(t) = \frac{di(t)}{dt} = \frac{\lambda(\frac{1}{i_0}-1)e^{-\lambda(\frac{1}{i_0}-1)}}{(1+(\frac{1}{i_0}-1)e^{-\lambda t})^2} = \frac{\lambda(\frac{1}{i_0}-1)}{e^{\lambda(\frac{1}{i_0}-1)} + 2(\frac{1}{i_0}-1) + (\frac{1}{i_0}-1)^2 e^{-\lambda(\frac{1}{i_0}-1)}} \tag{16}$$

$$\frac{dv(t)}{dt}|\, t = t_m = \frac{d(\frac{\lambda(\frac{1}{i_0}-1)}{e^{\lambda t} + 2(\frac{1}{i_0}-1) + (\frac{1}{i_0}-1)^2 e^{-\lambda t}})}{dt} = 0 \tag{17}$$

So t_m is:

$$t_m = \lambda^{-1} \bullet In(\frac{1}{i_0}-1) = \frac{1}{2} \tag{17}$$

When proportion of dissemination node among users is 50%, growth of dissemination node reach its maximum. The specific time is decided by spreading probability λ and rumor coverage rate.

Then, we will look into detail such as users' location (IP addresses), total numbers of accounts who repost the rumor WMs, viewing times and timestamps, etc, to provide more specific analysis of rumor WMs diffusion in China. Detailed analysis will be introduced in section 4.5.

Chapter 4: Results and Discussion

In this section, we compare and evaluate the performance of our proposed Automatic Crawling and Identification System Model based on a real-world WeChat Moment dataset, in a one-month period from Sep.1st, 2018 to Sep.26th, 2018.

4.1 Dataset

We collect 168,477 URLs in total. The average number of URLs in one day during the period from Sep.1st, 2018 to Sep.26th, 2018 is 6480. We use WeChat dataset collected above to carry out experiments and evaluate the performance of proposed model. Figure 5 indicates the number of URLs on different days in the dataset. The blue line represents total URLs on different days, the red line represents URLs of WeChat Moments on different days and green line represents the unique WM URLs on different days. X-axis represents the date from Sep.1st, 2018 to Sep.26th, 2018 whereas y-axis represents the number of collected URLs.

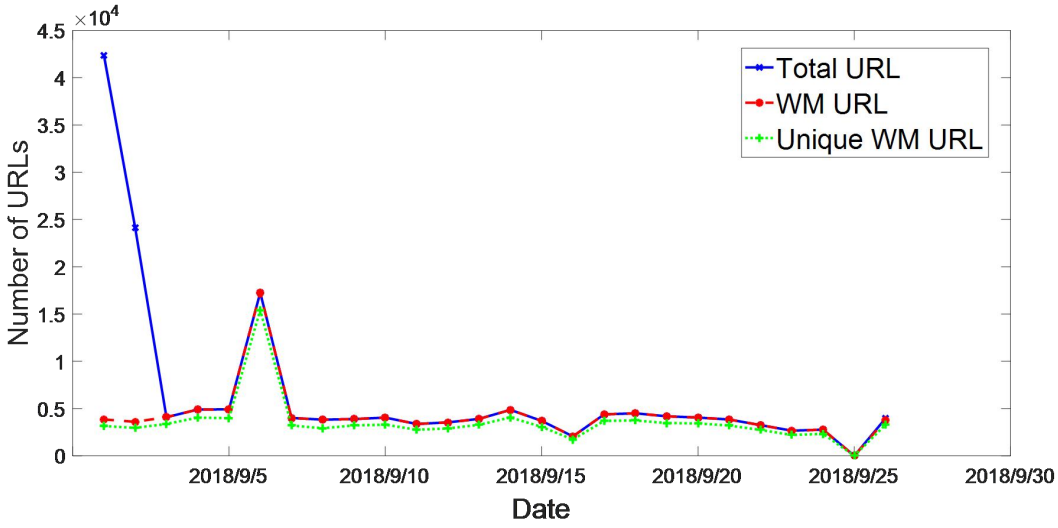


Figure 5 The number of different types of URLs

We totally collect 168,477 URLs of WeChat, including any type of URLs such as payment, shopping, wallet and so forth. Then we filter out the URLs that are not related to WeChat Moments, such as payment (pay.weixin.qq.com), office management tools (work.weixin.qq.com) and so on and the URLs that generate the same WM. Table 3 provides the total number of URLs, average, maximum and minimum value left after each step of filtering.

Table 3: Number of URLs left after filtering

	Total of URLs	Average	Maximum	Minimum
Total URL	168,477	6480	42387	0
WM URL	109,192	4200	17269	0
Unique WM URL	91,436	3517	15416	0

The data now involves 301 different IP addresses. Then we collect top ten most popular WM topics. Figure 7 indicates that people are more interested in reading WMs about tourism or traveling. X-axis represents the brief name of articles while y-axis represents the number of clicking times.

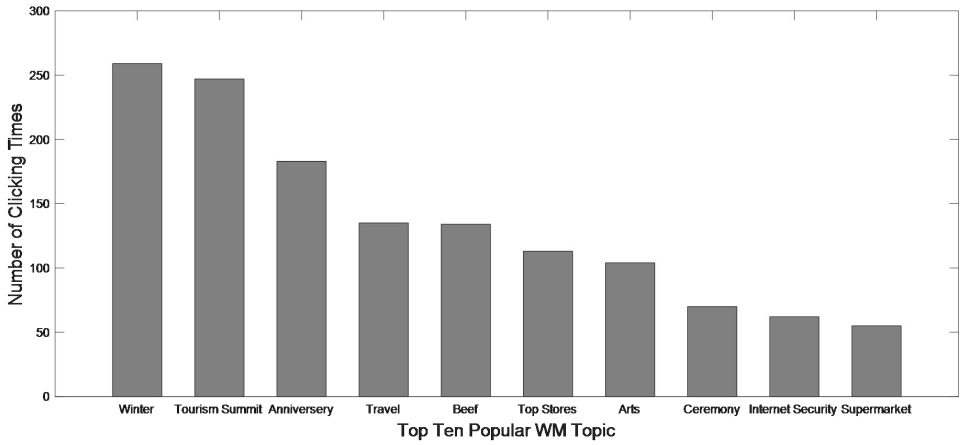


Figure 7 Top ten most popular WM topics

4.2 Rumor Detection

After first detecting the 91,436 WMs, we find that 29 of these WMs are labeled as rumors by WeChat Official Account Platform Anti-rumor Center, 20 of these WMs indicate 404 not found and 24155 among the total WMs show nothing, which means the WM page is deleted by the account. This can be a potential factor of rumors. In Figure 8, we divide the situation into four classifications: rumors, deleted, error and non-rumors. The bar plot shows the total amount of each classification. The x-axis is the 4 different classifications while y-axis as the log scale represents the number of URLs in each classification. Figure 9 describes the distribution of URLs in these four classification on different days. The x-axis represents the date whereas the y-axis as the log scale represents the number of URLs. The blue line represents the WMs deleted by the account, yellow line shows the non-rumor WMs, red line means the rumor WMs and green line represents WMs that have errors.

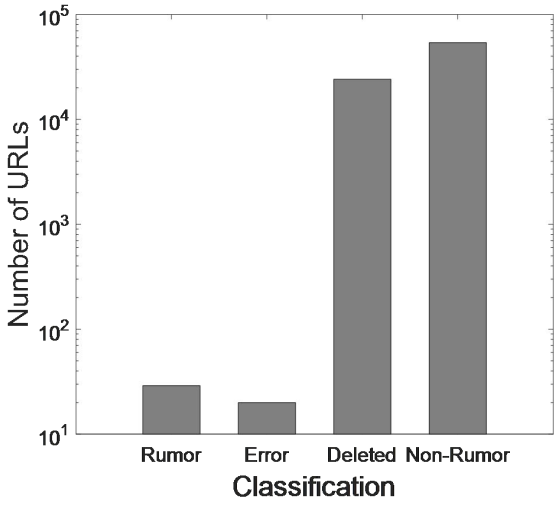


Figure 8 Bar plot of classification

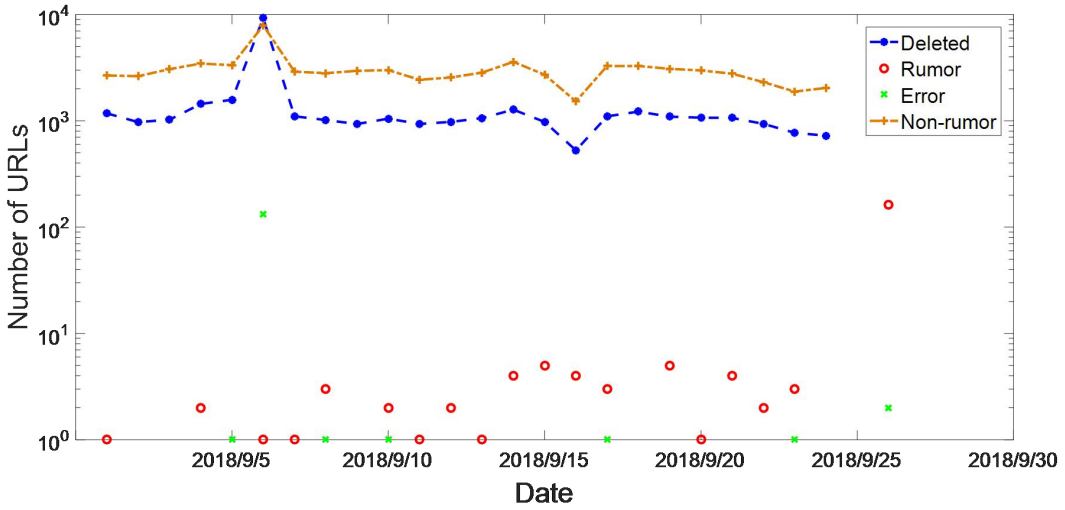


Figure 9 Distribution of different URLs

Additionally, no more rumors are detected by Piyao and Snopes. It is probably because that there is no overlap between the dataset in Piyao, Snopes and the dataset of WeChat Official Account Platform Anti-rumor Center.

Then, we construct the rumor testing dataset and non-rumor testing dataset based on the result of rumor detection. Among the total 29 rumor WMs detected by WeChat Official Account Platform Anti-rumor Center, there are 12 different rumor WMs. Besides, we also randomly collect 12 non-rumor WMs as our non-rumor testing dataset. We use these 24 WMs including 12 rumors and 12 non-rumors as our testing dataset.

For training dataset, we collect top 99 popular rumor WMs from WeChat Official Account Platform Anti-rumor Center as our rumor training dataset. Then we collect the same number of WMs from the dataset introduced in section 4.1 as non-rumor training dataset.

4.3 Feature Extraction

For both the training dataset and testing dataset, we extract the titles, contents and authors from each WM. Next, Chinese word segmentation is done separately on both titles and contents. Table 4a describes the number of different Chinese words in both training and testing dataset.

Table 4a: Number of different words

	Rumor Training Dataset	Non-Rumor Training Dataset	Rumor Testing Dataset	Non-Rumor Testing Dataset
Title	48,114	64,053	1,152	1,068
Content	947,727	1,981,089	30,060	58,476

Then, we record the keywords with TF-IDF value above zero to construct the feature set for further classification. Table 4b depicts the number of keywords whose TF-IDF value is above zero. Figure 10 depicts the comparison of percentage of keywords in different dataset. X-axis represents four different dataset and y-aixs is the percentage of keywords in the whole content. The left bar represents the title while right bar represents content of WMs.

Table 4b: Number of keywords with TF-IDF value above zero

	Rumor Training Dataset	Non-Rumor Training Dataset	Rumor Testing Dataset	Non-Rumor Testing Dataset
Title	643	721	100	92
Content	19,640	49,671	3,328	6,911

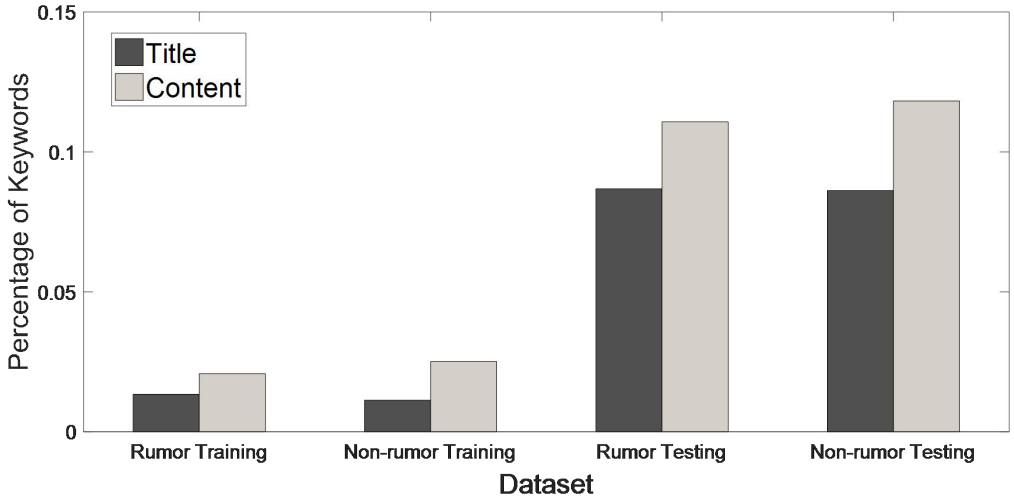


Figure 10 Percentage of keywords in each dataset

4.4 Rumor Identification

Three classification methods are deployed as the estimation functions and compared.

- Support Vector Machine
- Naive Bayes
- Decision Tree

Table 5 and Figure 10 illustrate the performance of three classifiers (Support Vector Machine, Naive Bayes and Decision Tree) chosen in Rumor Identification process. As mentioned in Section 3.4.3, we select Precision, Recall, F-measure and Accuracy as evaluation metrics. Figure 10a depicts the performance of three classifiers before adding account feature. Figure 11a shows Decision Tree performs the best with 77.8% precision, 58.3% recall, 66.7% f-measure and 70.8% accuracy. Naive Bayes performs the worst with only 46.2% precision, 50% recall, 48% f-measure and 45.8% accuracy.

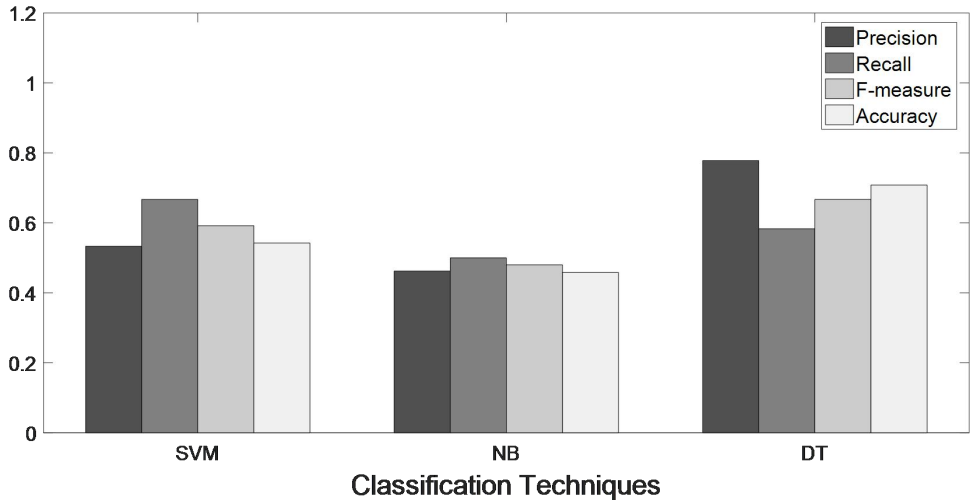


Figure 11a The performance of three classifiers

Figure 11b shows the performance of three classifiers adding all three features mentioned in Section 3.4.2. In Table 5 and Figure 9b, the overall performance of classifier based on Decision Tree outperforms the other two classifiers and its precision, recall, f-measure and accuracy are 100%, 33.3%, 50% and 66.7% respectively. Support Vector Machine performs worst and its precision, recall, f-measure and accuracy are 50%, 58.3%, 53.8% and 50% respectively. Our proposed feature and method based on Decision Tree classifier outperforms other existing methods such as algorithm proposed in [11] with an only 61% accuracy.

Table 5: Performance of three classifiers

Classifier	Precision(%)	Recall(%)	F-measure(%)	Accuracy(%)
SVM	50	58.3	53.8	50
NB	60	50	54.5	58.3
DT	100	33.3	50	66.7

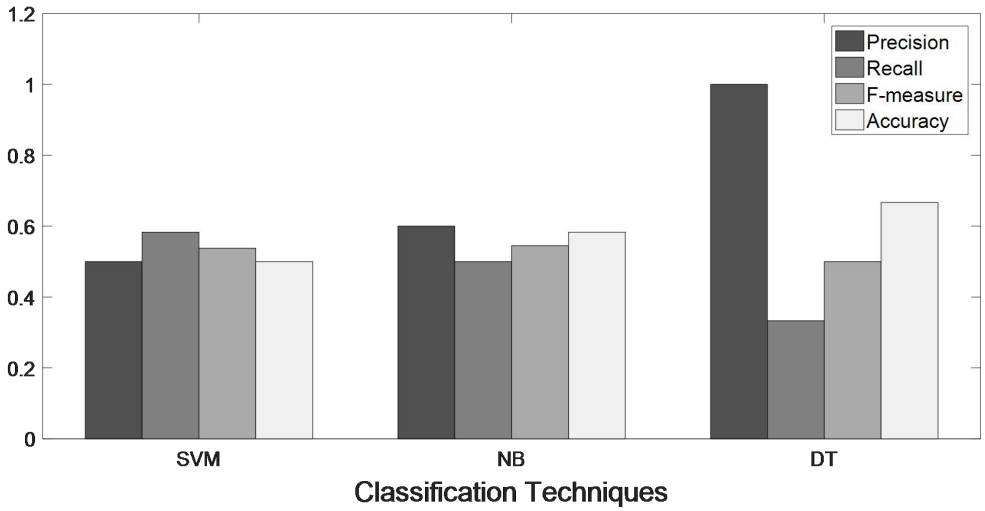


Figure 11b The performance of three classifiers adding account feature

4.5 Rumor Diffusion

4.5.1 SI Model

In this subsection, we first analyze the rumor diffusion based on SI model mentioned in section 3.5. Figure 12 shows rumor diffusion based on SI model. X-axis represents time and y-axis represents percentage of susceptible and infected users who distribute rumor WMs.

We find that in general, susceptible users tend to become infected users over time. For very long time, all susceptible users will change to infected users in WeChat network. In other words, percentage of infected users will become one if time towards infinite. Nevertheless, this situation is impossible in real world due to complexity of networks.

Therefore, we analyze rumor diffusion in detail. We take account of many real-world factors that will influence rumor diffusion in WeChat network. We analyze the characteristics of the identified rumors in the dataset, including users' location (IP addresses), total numbers of accounts who repost the rumor WMs, viewing times and timestamps.

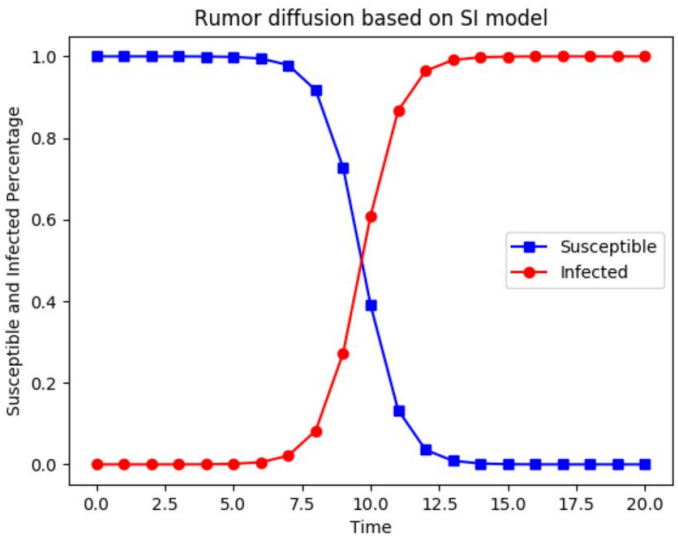


Figure 12 Rumor diffusion based on SI model

4.5.2 User Location

User Location is the IP address, which can be closely related to WeChat users' real living conditions. Due to privacy, we only provide rough locations (cities or provinces) of users who are susceptible or infected by rumor WMs. We collect the IP address of each click of rumor, error and deleted WMs.

Figure 13a depicts the main locations in China, namely the source of rumor WMs. We find that Guangdong Province (113.66.108.*), Tianjin (220.194.91.*), Chongqing (122.73.9.*) are more tended to distribute rumor WMs. The red dots in the map are the main locations distributing rumor WMs.

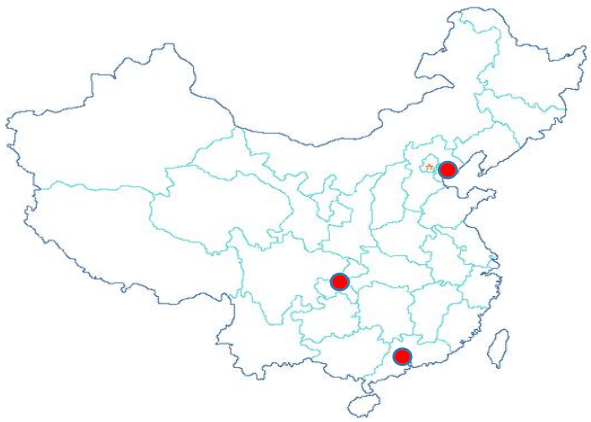


Figure 13a Rumor IP Distribution

Figure 13b depicts the main locations in China, namely the source of error WMs. We found that Tianjin (220.194.91.*), Guangdong (113.105.131.*), Beijing (61.50.96.*), Nanjing city in

Jiangsu Province (58.240.53.*), Weihai (112.247.207.*) and Jinan (124.128.76.*) in Shandong Province, Shanghai (58.247.46.*), Heilongjiang Province (60.11.231.*), Zhejiang Province (211.138.118.*) and Shenyang city in Liaoning Province (218.60.8.*) are more tended to distribute error WMs. The red dots in the map are the main locations distributing error WMs and the red lines show the propagating directions. Most red dots and lines are distributed in the eastern coastal cities in Figure 13b, so we can conclude that the coastal cities or provinces are more likely to distribute error WMs than inland cities or provinces. It is probably because that information propagates much faster in more developed cities in eastern coastal cities in China.

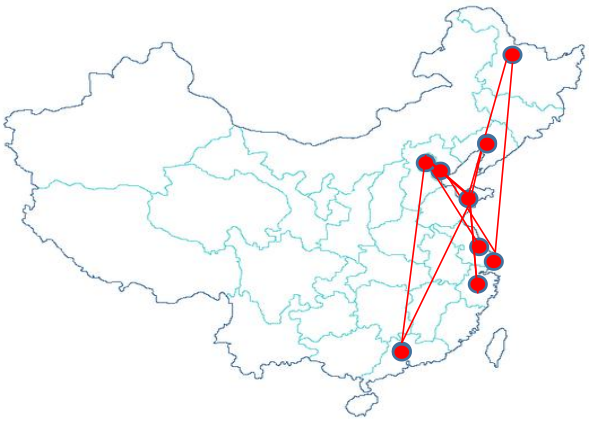


Figure 13b Error IP Distribution

Figure 13c depicts the main locations in China, namely the source of deleted WMs. We found that Shenzhen city (113.105.131.*), Guangzhou city (113.66.108.*) and Shantou ciy (116.26.84.*) in Guangdong Province, Weihai city (119.179.133.*) and Heze city (221.1.205.*) in Shandong Province, Chongqing (122.73.10.*), Henan Province (182.118.20.*), Tianjin (220.194.91.*), Handan city (221.193.177.*) in Hebei Province, Chengdu city (221.10.28.*) in Sichuan Province, Fushun city (221.203.87.*) in Liaoning Province, Nanjing city (58.240.53.*) in Jiangsu Province, Shanghai (58.246.141.*) and Heilongjiang Province (60.11.231.*) are more tended to distribute deleted WMs. The red dots in the map are the main locations distributing deleted WMs and the red lines show the propagating directions. We analyze the diffusion of deleted WMs due to the fact that the rumor WMs we have in the dataset now are not enough to get any meaningful results and these deleted WMs are potential rumor WMs. Based on Figure 13c, we find that most deleted WMs are distributed in the east of China due to its fast development of technology and information diffusion.

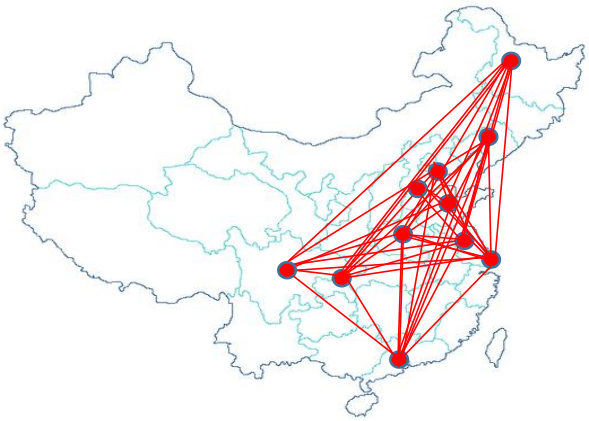


Figure 13c Deleted IP Distribution

4.5.3 WM Page Views

In this subsection, we examine the relationship between the accounts who repost the rumor WMs and the number of rumor WM page views, both of which are strong factors of rumor diffusion. Note that the number of rumor WM page views is no smaller than the account number since multiple users can view the rumor WM after one account release the fake news. Figure 14 indicates the relationship between account number and number of rumor WM page views. The x-axis represents the account who reposts the rumor WMs and the y-axis is the number of rumor WM page views. The number of page views is almost proportional to the account number, indicating that more accounts who repost the rumor WMs can lead to more infected users who view the rumor WMs. Hence, this relationship also proves that account is an important feature in analysis of rumor identification and diffusion.

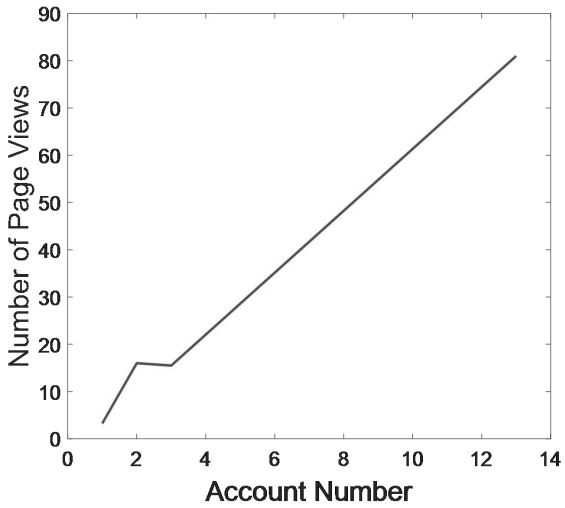


Figure 14 Relationship between account number and number of page views

4.5.4 Page Life Time

We define the page life time of a rumor WM page as the time period from the rumor WM first released in the WeChat network to the time point when no more users view the rumor WM page for at least one day. Since we have collected the dataset with timestamps, we calculated the active time period of each rumor WM based on the timestamps.

Figure 15a and figure 15b describe the life time of rumor WMs and non-rumr WMs separately. In figure 15a, x-axis represents time whereas y-axis represents number of page views. Four top popular rumor WMs are chosen to analyze the relationship between life time and page views. They are in the field of study, medicine, Chinese health and health preservation respectively. In figure 15b, the life time of non-rumor WMs, x-axis is the date and y-axis is the number of page views. Similarly, four top popular non-rumor WMs are chosen to compare with life time of rumor WMs. They are in the field of travelling, shopping, food and anniversary respectively. We find that rumor WMs are always in the field of health such as medicine, health preservation and so on, mainly due to the fact that elderly in society who want to live for a long time are more likely to follow fake news in blind confidence.

Besides, we observe that compared with diffusion of non-rumor WMs, rumor WMs spread more quickly. Within one hour, rumor WMs can spread in the infected users' WeChat circle. However, non-rumor WMs need more time to spread among users in WeChat network, which implies that rumor WMs may contain some keywords that can promote propagation. This also accords with our ACISM, which use keywords to predict rumor in WMs. Additionally, non-rumor WMs can exist for at least a few days in WeChat network. We notice that the number of page views increase rapidly in the first few days after the WM releases, while the increasing rate becomes smaller thereafter. Then the WMs become inactive after ten days of release. All non-rumor WMs are active within 10 ten days. Nonetheless, compared with life time of non-rumor WMs, the life time of rumor WMs is much shorter with only few hours' existence. All the rumors we detected can only exist no more than one day because rumor WMs can be detected by some official organizations or individuals. Then the rumor WMs are restricted to propagate in WeChat network in terms of IP restrictions, rumor tags, WM deletion, etc.

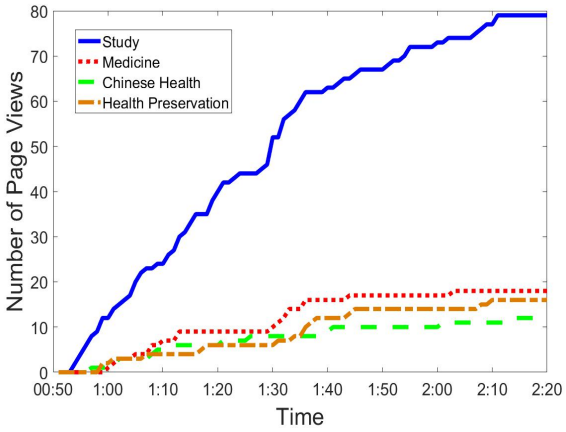


Figure 15a The life time of rumor WMs

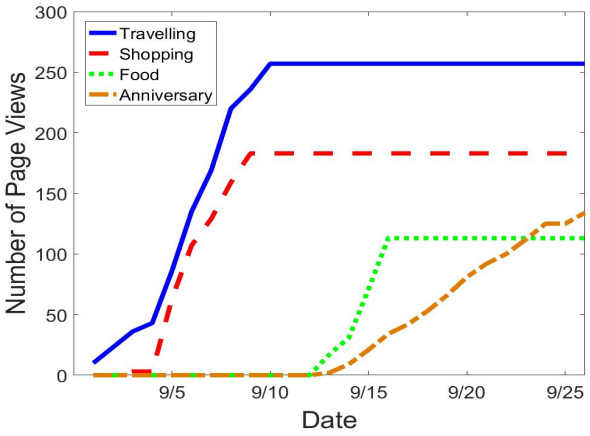


Figure 15b The life time of non-rumor WMs

Chapter 5: Conclusion and Further Work

With the popularity of WeChat, which allows users to freely post and share unverified information in WMs they are interested in anywhere and anytime, rumor WMs appear and quickly spread among the users in WeChat network. Rumor Wms are harmful to our society. Besides, due to the lack of automatic rumor WM detection works, it is necessary for us to develop a tool which can automatic identify rumor WMs propagating in WeChat network.

Detecting and identifying rumors automatically is an important and challenging task in WeChat network because of privacy problems and few existing rumor detection platforms. In this paper, we first build an important dataset of both rumors and non-rumors with 168,477 WMs in total from Sep.1st, 2018 to Sep.26th, 2018, including Uniform Resource Locations (URLs), titles, contents, accounts and timestamps. Besides, we propose a novel Automatic Crawling and Identification System Model to detect and identify rumors in WMs. After passing all URLs through our model, we find that 29 of these WMs are labeled as rumors by WeChat Official Account Platform Anti-rumor Center, 20 of these WMs indicate 404 not found, 24155 among the total WMs are deleted by the account and no more rumors are detected by Piyao and Snopes. We use machine learning methods to help us identify more rumors. The experimental results have shown that our algorithm is more accurate and effective than some existing algorithms. Three features (title, content and account) for rumor identification we proposed can successfully identify rumor WMs with an accuracy of 66.7%, which outperforms other existing approaches such as algorithm proposed in [11] with an only 61% accuracy, demonstrating the ability of our proposed ACISM and features to successfully and accurately identify rumor WMs in WeChat network.

In addition, we conclude the detected rumors into five categories.

I. Crowned with authority. They usually say that the article is released by authority.

II. News with sensational titles. Authors usually use sensational or terrifying titles such as some shocking words like amazing, miraculous to attract readers' attention.

III. Urging to read and distribute. Some articles have words such as emergency notice.

IV. Utilizing sad stories to catch people's eyes. Some articles prefer utilizing the elderly and children to make sensational stories, then they induce readers to forward these fake news in the name of love.

V. Exaggerating harm or efficacy. This method is usually used in articles of health. They blindly exaggerate the harm of some substances or they declare that they can cure some serious illness of small cost in a simple way in short time, resulting in rapid distribution.

Our model further analyze rumor diffusion based on Susceptible Infected Model and user location, WM page views, page life time to analyze the spread of rumors in WMs in detail. We find that only analyzing rumor diffusion based on SI model is not accurate enough because real world is not an ideal situation, lots of other factors such as IP address, page views and page life time will influence rumor diffusion in WeChat network. We find that Guangdong Province, Tianjin and Chongqing are more tended to distribute rumor WMs. Besides, the coastal cities or provinces are more likely to distribute error WMs than inland cities or provinces and most deleted WMs are distributed in the east of China. It is probably because that information propagates much faster in more developed cities in eastern coastal cities of China. We also find that the number of page views is almost proportional to the account number, indicating that more accounts who repost the rumor WMs can lead to more infected users who view the rumor WMs. Additionally, rumor WMs spread more quickly than non-rumor WMs and the life time of rumor WMs is much shorter than non-rumor WMs, with only few hours' existence. These findings can not only be used for limiting the spread of rumors and preventing rumor diffusion in advance, but also they are valuable for analyzing rumor evolution.

In future work, we will collect more data to help identify rumors in WeChat because only 198 WMs in the training dataset are not enough to provide accurate and precise result. In other words, we will collect enough ground truth to help us collect more different features of rumor WMs in the future. Additionally, we will investigate other new features when doing TFIDF in order to improve the accuracy and precision of rumor identification. We will add additional features such as sentiment analysis based on some Python libraries to infer the sentiment of a given piece of text. Furthermore, we will try to refer to other platforms such as http://www.fibodata.com to collect more information about user behaviors. Then we will provide a more precise analysis of rumor spread based on these user behaviors. Finally, we will study how to apply the ACISM and features in other popular social networks such as QQ, Sina Weibo, Zhihu and so on.

References

[1] Sensor Tower - Mobile App Store Marketing Intelligence: https://sensortower-china.com/?locale=zh-CN.

[2] Definition or rumor in Merriam Webster Dictionary: https://www.merriam-webster.com/dictionary/rumor

[3] Soroush Vosoughi (2015). Automatic Detection and Verification of Rumors on Twitter. Submitted to the Program in Media Arts and Sciences, in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY June 2015. MIT

[4] Official Website of WeChat Official Account Platform Anti-rumor Center: https://mp.weixin.qq.com/cgibin/opshowpage?action=dispelinfo&lang=zh_CN&begin=1&count=9&token=.

[5] Zhongqing Wang and Yue Zhang (2017). A Neural Model for Joint Event Detection and Summarization. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, {IJCAI} 2017, Melbourne, Australia, August 19-25, 2017 (pp.4158-4164). Soochow University.

[6] Deyu Zhou, Linsen Guo and Yulan He (2018). Neural Storyline Extraction Model for Storyline Generation from News Articles. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, {NAACL-HLT} 2018, New Orleans, Louisiana, USA, June 1-6, 2018 (Vol.1, pp. 1727-1736). Southeast University.

[7] Bang Liu, Di Niu, Kunfeng Lai, Linglong Kong and Yu Xu (2017). Growing Story Forest Online from Massive Breaking News. In Proceedings of the 2017 {ACM} on Conference on Information and Knowledge Management, {CIKM} 2017, Singapore, November 06 - 10, 2017 (pp.777-785). University of Alberta.

[8] Fatia Kusuma Dewi, Satrio Baskoro Yudhoatmojo, Indra Budi (2017). Identification of opinion leader on rumor spreading in online social network Twitter using edge weighting and centrality measure weighting. In Twelfth International Conference on Digital Information Management, {ICDIM} 2017, Fukuoka, Japan, September 12-14, 2017 (pp.313-318). Universitas Indonesia.

[9] Gang Liang, Jin Yang, Chun Xu (2016). Automatic rumors identification on Sina Weibo.

In 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, {ICNC-FSKD} 2016, Changsha, China, August 13-15, 2016 (pp.1523-1531). Sichuan University

[10] Hua Zhang, Jun Wang, Yan Chen, Jinghua Tan and Qing Li (2017). Research on Automatic Identification of Rumors in Stock Forum Based on Machine Learning. In 21st Pacific Asia Conference on Information Systems, {PACIS} 2017, Langkawi, Malaysia, July 16-20, 2017 (pp46). Southwestern University of Finance and Economics.

[11] Yuanxing Zhang, Kaigui Bian, Lin Chen, Shaoling Dong, Lingyang Song and Xiaoming Li (2018). Early Detection of Rumors in Heterogeneous Mobile Social Network. In Third {IEEE} International Conference on Data Science in Cyberspace, {DSC} 2018, Guangzhou, China, June 18-21, 2018 (pp.294-301). Peking University.

[12] Zhuqi Li, Lin Chen, Yichong Bai, Kaigui Bian and Pan Zhou (2016). On diffusion-restricted social network: {A} measurement study of WeChat moments. In 2016 {IEEE} International Conference on Communications, {ICC} 2016, Kuala Lumpur, Malaysia, May 22-27, 2016 (pp.1-6). Peking University.

[13] Chengliang Gao, Yuanxing Zhang, Kaigui Bian, Shaoling Dong and Lingyang Song (2018). On Lifecycle of Interactive Web Apps in WeChat. In 2018 {IEEE} International Conference on Communications, {ICC} 2018, Kansas City, MO, USA, May 20-24, 2018 (pp.1-6). Pecking University.

[14] Yuanxing Zhang, Zhuqi Li, Chengliang Gao, Kaigui Bian, Lingyang Song, Shaoling Dong and Xiaoming Li (2018). Mobile Social Big Data: WeChat Moments Dataset, Network Applications, and Opportunities. In IEEE Network (Vol.32, pp.146-153). Peking University.

[15] Wangchun Jiang, Bin Chen, Lingnan He, Yichong Bai and Xiaogang Qiu (2016). Features of Rumor Spreading on WeChat Moments. In Web Technologies and Applications - APWeb 2016 Workshops, WDMA, GAP, and SDMA, Suzhou, China, September 23-25, 2016, Proceedings (pp.217-227). National University of Defense Technology.

[16] Official Website of Piyao in China: http://www.piyao.org.cn/.

[17] Official Website of Snopes: www.snopes.com.

[18] Definition for Beautiful Soup: https://en.wikipedia.org/wiki/Beautiful_Soup.

[19] Website of Jieba Packet: https://pypi.org/project/jieba/.

[20] Pi-Chuan Chang, Michel Galley, and Christopher D Manning. 2008. Optimizing Chinese word segmentation for machine translation performance. In Proceedings of the third workshop on statistical machine translation. Association for Computational Linguistics, 224−232. Stanford University.

[21] Website of SnowNLP Packet: https://pypi.org/project/snownlp/.

[22] Definition of TFIDF: https://en.wikipedia.org/wiki/Tf%E2%80%93idf.

[23] Definition of SVM: https://en.wikipedia.org/wiki/Support-vector_machine.

[24] Definition of Naïve Bayes: https://en.wikipedia.org/wiki/Naive_Bayes_classifier.

[25] Definition of Decision Tree: https://en.wikipedia.org/wiki/Decision_tree.

[26] Official website of Scikit-learn: http://scikit-learn.org/stable/.

[27] Kundan Kandhway and Joy Kuri (2014). Accelerating information diffusion in social networks under the Susceptible-Infected-Susceptible epidemic model. 2014 International Conference on Advances in Computing, Communications and Informatics, {ICACCI} 2014, Delhi, India, September 24-27, 2014 (pp.1515-1519). Indian Institute of Science.

Acknowledgement

I would like to express my great gratitude to my supervisor, who is a responsible, considerate and warm-hearted scholar. He has provided me with valuable advice in every step of my final year project. I also would like to thank an Assistant Professor from BUPT, he provided me with the important dataset (millions of URLs in WeChat) of WeChat Moments. No further analysis could be done without this valuable dataset.

This work has been supported by both Beijing University of Posts and Telecommunications and Queen Mary University of London.

Appendix

北京邮电大学 本科毕业设计（论文）任务书

Project Specification Form

Part 1 – Supervisor

论文题目 Project Title	The spread of rumors in WeChat		
题目分类 Scope	Networks and Wireless	Research	Software
主要内容 Project description	Wechat is one of the most prominent ways of communications in China and beyond. As false news and ”rumors” have become a more and more polemic issue, with multiple researhc efforts looking at Facebook or Twitter for the proliferation and dynamics of these ”rumors”. This project will look at real traffic from wechat, identify ”rumors”and study how do they spread across its users.		
关键词 Keywords	traffic analysis		
主要任务 Main tasks	**1** Preliminary analysis of the wechat data and state fo the art on rumors in social networks		
	2 Identification of fake news or rumours (e.g., using https://www.snopes.com/)		
	3 Characterization of the sites accessed through wechat		
	4 Characterization of the sites accessed through wechat		
主要成果 Measurable outcomes	**1** A dataset with websites identified as a ”rumor”		
	2 A characterization of the sites in Outcome 1		
	3 A study of the spread of the sites in Outcome 1		

北京邮电大学 本科毕业设计（论文）任务书

Project Specification Form

Part 2 - Student

学院 School	International School	**专业 Programme**	**Telecommunications Engineering with Management**		
姓 Family name	Gong	**名 First Name**	Jiaying		
BUPT 学号 BUPT number	2015212931	**QM 学号 QM number**	151009262	**班级 Class**	2015215103
论文题目 Project Title	The spread of rumors in WeChat				
论文概述 Project outline **Write about 500-800 words** **Please refer to Project Student Handbook section 3.2**	WeChat application is a mobile social media software launched by Tencent in 2011, which has played an increasingly significant role in society and the emerging market is becoming mature. According to Sensor Tower, Wechat application downloads have surpassed 28 million during the first season in 2018 and there are more than one billion active users of Wechat.[1] However, as the platform has become more popular, and group sizes themselves have grown, the reliability of its system of trust has decreased.[2] Fake news appear and quickly spread among the users. It's time for us to focus on real traffic from Wechat, identifying rumors and studying how do they spread across its users. For privacy and security, all the traffic data in Wechat are encrypted. Therefore, our research only focus on fake news in WeChat Subscription. Data about URLs of WeChat Subscription will be collected with the help of teachers due to privacy and security. The main challenge of our research is identification of fake news since most data are encryted and there are few existing Chinese platforms for rumor identification. There are some relevant works on rumor identification in social networks like Tweeter, Weibo.[3] However, there's no research studying rumor identification in Wechat. We will use similar methodologies such as using time difference to identify rumors and then collecting the rumor features to help build our own dataset and classifier. Similar works has been down in other applications. For example, some platforms like Twitter delete bot/malicious accounts.[4] We can therefore repeatedly check the news and assume it is malicious if the news is deleted by Wechat. Besides, we will also check the URLs through different rumor-identifying platforms to build our dataset with websites identified as rumors. After identification of fake news, we will use timestamps and IP souces to study the spread of rumors.[5] The predicted outcomes of our research will be a new dataset with websites identified as rumors, the model of spread of the rumors and the characterization of the sites accessed through Wechat. The main programming language in our research is Python, since it is much convenient and simple to crawl relevant information in Python. We will crawl the relevant information through Wechat Subscription URLs and encode the relevant information in order to automatically go through				

	rumor-identifying platforms to identify rumors. Besides, EmEditor will be used to do some basic operations such as filtering relevant URLs about Wechat Subscriptions on a large scale data. Furthermore, Matlab will be used to generate plots of the basic characterisation of the results such as distributions of rumors, box plot of the number of views for each article topic to better understand the URLs and the final results. [1] https://sensortower-china.com/?locale=zh-CN [2] https://www.wired.com/2017/04/how-wechat-spreads-rumors-reaffirms-bias-and-helped-elect-trump/ [3] Automatic Rumors Identification on Sina Weibo, 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) [4] Identification of Opinion Leader on Rumor Spreading in Online Social Network Twitter Using Edge Weighting and Centrality Measure Weighting, The Twelfth International Conference on Digital Information Management (ICDIM 2017) September 12- 14, 2017, Kyushu University, Fukuoka, Japan. [5] Features of Rumor Spreading on WeChat Moments
道德规范 **Ethics**	Please confirm that you have discussed ethical issues with your Supervisor using the ethics checklist on QMPlus. [YES/NO]YES
	Summary of ethical issues: (put N/A if not applicable) Data about Wechat URLs was provided by in an already anonymised and secure form.

中期目标 **Mid-term target.** **It must be tangible outcomes, E.g. software, hardware or simulation.** **It will be assessed at the mid-term oral.**	Identifying fake news in Wechat articles. Building a dataset with websites identified as fake news.

Work Plan (Gantt Chart)

Fill in the sub-tasks and insert a letter X in the cells to show the extent of each task

	Nov	Dec	Jan	Feb	Mar	Apr	May
Task 1 break down							
Understanding different meanings of traffic patterns in Wechat and can provide examples.	X						
Collecting traffic data about different WeChat usage.	X						
Filtering the useful Wechat traffic such as URLs of articles in Wechat.	X						
Crawling information(news) from all filtered URLs.	X						
Task 2 break down							
Considering different approaches to finding rumors.	X						
Making URLs go through platforms to identify rumors.		X					
Generating simple plots of the basic characterisation of the results.		X					
Building a dataset with URLs identified as rumors.		X					
Task 3 break down							
Collecting features of fake news.			X				
Building a classifier to predict fake news.			X				
Building a new dataset with predicted fake news.			X				
Generating plots of the new results.			X				
Task 4 break down							
Proposing the characterization of the sites accessed through wechat.				X			
Providing the spread of rumors across Wechat users.				X			
Generating plots of the spread of rumors.				X			
Completing final draft report and slides.					X		

北京邮电大学 本科毕业设计（论文）初期进度报告
Project Early-term Progress Report

学院 School	International School	**专业 Programme**	**Telecommunications Engineering with Management**		
姓 Family name	Gong	**名 First Name**	Jiaying		
BUPT 学号 BUPT number	2015212931	**QM 学号 QM number**	151009262	**班级 Class**	2015215103
论文题目 Project Title	The spread of rumors in WeChat				

已完成工作 Finished work:

What material was read or researched?

First, I have read several papers related to rumors in Wechat or the spread of news in Wechat. Here is the summary and my ideas of these papers.

On Diffusion-restricted Social Network: A Measurement Study of WeChat Moments.IEEE ICC 2016

I think this paper may be useful for analyzing the spread of rumors. It analyzed how Wechat Moments (Wechat articles) spread by considering lots of factors such as number of views, path length, users' locations and so on. They used the results to help them develop marketing strategies. I think we can use their method or idea of analyzing spread of wechat articles to help us analyze the spread of rumors.

On Lifecycle of Interactive Web Apps in WeChat.IEEE ICC 2018

This paper is similar to above. They developed models to analyze lifecycle of web apps in Wechat, which can be used to analyze the spread of articles.

Mobile Social Big Data: WeChat Moments Dataset, Network Applications, and Opportunities. IEEE Network, 2018

The most useful part in this paper is their data set. They use and cooperate with http://www.fibodata.com to help them analyze lots of things about Wechat Moments. I found this website is very amazing. It can show the number of viewings, ip addresses, spread path, users' locations and anything we want to see. I just see the demo of this website. I think if we want to use this website, it may probably charge.

"Early Detection of Rumors in Heterogeneous Mobile Social Network". IEEE DSC 2018

This paper is quite difficult to understand. It's about rumor detection. The whole idea is that they proposed a new algorithm to learn the latent factors of each information and then the learned factors are used to predict the rumors. Then they use the dataset of rumors labeled by Wechat to verify that their proposed algorithm can identify 61% of rumors. Their consideration is very complexed. They take lots of factors into consideration. For instance, they think that the articles I repost to you is different from the same article you repost to me.

What work was done?

Collecting and analyzing the data set.

In order to identify rumors in Wechat articles, we first collect traffic data about different

Wechat usage with ip source addresses and time stamps in a time period. After filtering out those Uniform Resource Locators(URLs) that can not work well, we collect 109404 URLs about Wechat articles. All these 109404 URLs are generated from Sep.1st, 2018 to Sep.26th, 2018. After filtering out the URLs that are not related to Wechat articles, we still have 108765 URLs left.

Rumor Identification

The raw data we have are URLs containing lots of information in Wechat such as payment (pay.weixin.qq.com), oce management tools (work.weixin.qq.com), articles (mp.weixin.qq.com) and so on. Due to encryption problems and speed of spread, we choose to filter URLs about Wechat articles. We use the software EmEditor, which can be used to analyze data on a large scale, to filter URLs. We first input mp.weixin.qq.com and filter all articles in Wechat. We then crawled the information(news) of all filtered URLs in a few weeks after generation of these URLs. If the URLs are fresh enough, we can't find any rumors because social platforms need time to do rumor identification. Therefore, we wait for a few weeks, and then we crawl the information in Wechat articles. In this way, we tend to use time difference to identify rumors. We choose to record all Wechat article titles because all key information is fully generalized in article titles and these information can be used later to easily go through other rumor identifying platforms to identify rumors. After crawling all titles in Wechat articles, we discover that some articles are labeled as rumors by Wechat. Since all the articles in these URLs are in Chinese, we find no more rumors after we pass the URLs through www.snopes.com. Next, we encode these titles into URLs and we make these URLs automatically go through a platform(http://www.piyao.org.cn) to check rumors in Chinese. Then we collect the results whether the Wechat articles are rumors or not. After we pass all URLs through www.piyao.org.cn, there is no more rumors detected. It is probably that the there is no overlap of between the data set in piyao platform and the data set of Wechat articles.
Then we do some statistics plots about the results.

What problems were faced?

1. Excel can’t deal with millions of data to filter relevant information.
2. Can’t identify rumor articles in Chinese automatically through the platform www.piyao.org.cn
3. Can’t find rumors labeled officially by Wechat.
4. When generating plots, since the difference between labeled rumors and others is so large that the number seems to be zero of rumors.

What solutions were found?

1. Using EmEditor, a software which can help filter relevant data on a large scale.
2. Generate the key word through encoding, so that specific URLs containing the relevant keywords are generated to go through some websites automatically.
3. Wait for a few weeks and use time difference to identify rumors.
4. Draw the plot in the log scale so that number of rumors can be clearly seen.

是否符合进度？On schedule as per GANTT chart?

[YES/NO]YES

下一步 Next steps:

After reading several papers about articles spread in Wechat, we found a website http://www.fibodata.com, which can be used to analyze lots of information about Wechat Moments such as number of viewings, ip addresses, spread path, users' locations and so on. We will next look into this website and find anything useful for us to analyze the spread of rumors.

Then we will try to find any other existing websites besides www.piyao.org.cn to help us identify rumor articles in Wechat since the platform of piyao can't do well in this field.

Then we will collect features of these articles labeled as rumors by Wechat and try to build a classifier to predict whether an article is a rumor or not based on our own data set.

Then we will generating some new plots to better explain the results.

Finally we will look into characteristics about rumor articles such as the ip addresses, times-tamps and so on to help analyze the spread of rumors and generate some relevant plots.

北京邮电大学 本科毕业设计（论文）中期进度报告
Project Mid-term Progress Report

学院 School	International School	**专业 Programme**	**Telecommunications Engineering with Management**		
姓 Family name	Gong	**名 First Name**	Jiaying		
BUPT 学号 BUPT number	2015212931	**QM 学号 QM number**	151009262	**班级 Class**	2015215103
论文题目 Project Title	The spread of rumors in Wechat				

是否完成任务书中所定的中期目标？Targets met (as set in the Specification)?
[YES/NO]YES

已完成工作 Finished work:

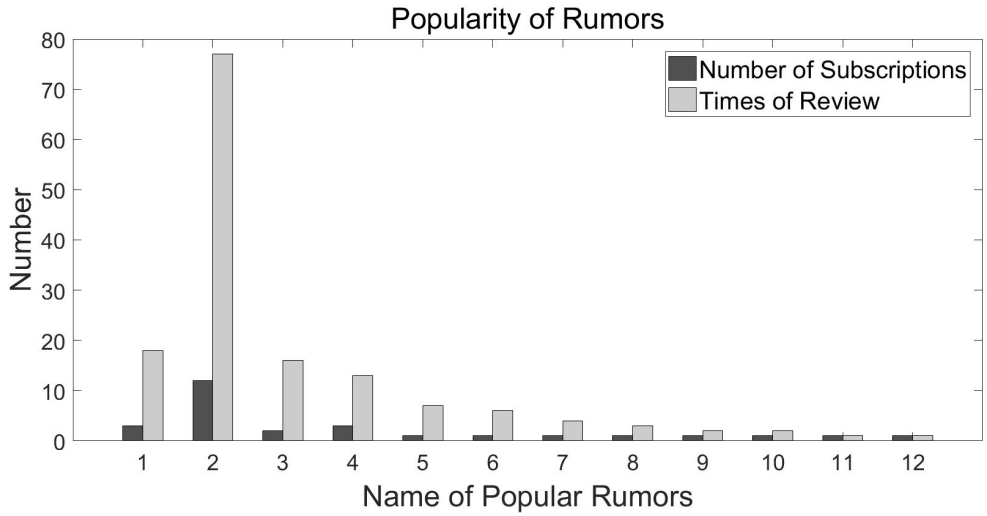
The main challenge of our research is identification of fake news since most data are encrypted and there are few existing Chinese platforms for rumor identification. All the data of message text are encrypted in Wechat besides Wechat articles, therefore, we choose to analyze rumors in Wechat articles. Besides, little works have been done to identify rumors in Wechat. There are some relevant works on rumor identification in social networks like Twitter, Weibo. Therefore, we imitate some similar works done in Tweeter or Weibo such as using time difference to identify fake news. After we first record the information labeled as rumors by Wechat officially, we get 29 pieces of fake news.

Rumor Identification

We look through all these URLs and find out the links in these articles. We go through all these links and extract information of these links appeared in each URL. The result shows that there is no more rumors and most of these articles are deleted by the author. After we pass all URLs through www.piyao.org.cn, which is used for identifying fake news in Chinese, there is no more rumors detected. It is certain that the there is no overlap of between the data set in piyao platform and the data set of Wechat articles. Besides, we also pass all URLs through www.snope.com , which similarly to piyao.com, is a platform used for fake news detection, but it can only check fake news in English, to check whether any articles in English are fake news. The result shows that no English articles in Wechat are rumors. Next, we check whether any links in the articles can lead to fake news.

After first crawling the 78088 Wechat articles, we find that 29 of these articles are labeled fake news by a third-party organization within Wechat, 20 of these articles indicate 404 not found and 24155 among the total articles show nothing, which means the article is deleted by the author. This can be a potential factor of rumors. We divide the situation into four classifications: rumors, deleted, 404 not found and others.

Analysis of Rumors



In Figure 6, the total data is 29 rumors. The x-axis from 1 to 12 represents twelve different rumors among the 29 rumors we have already detected. The left bar represents the number of different article authors who have copied the fake news and shared the rumors. The right bar represents the number of reviews. Among the 29 rumors, the most popular rumor is the second one in Figure 6, which is an introduction of the first full mark composition for the college entrance examination. This fake news is distributed by twelve different Wechat Subscriptions using almost the same title. This article is reviewed by 77 different users.

Rumor Feature Extraction

After analyzing titles and contents of these rumors, we can classify rumors into five categories based on their features.

I. Crowned with authority. They usually say that the article is released by authority. For example, the titles usually include that central authority speaks that...or top secrets are declared by authority.

II. News with sensational titles. Authors usually use sensational or terrifying titles such as some shocking words like amazing, miraculous to attract readers' attention.

III. Urging to read and distribute. Some articles have words such as emergency notice, read immediately and it will be deleted right away and so forth.

IV. Utilizing sad stories to catch people's eyes. Some articles prefer utilizing the elderly and children to make sensational stories, then they induce readers to forward these fake news in the name of love. For instance, be careful if you have kids, reading the article with tears and so on.

V. Exaggerating harm or efficacy. This method is usually used in articles of health. They blindly exaggerate the harm of some substances or they declare that they can cure some serious illness of small cost in a simple way in short time, resulting in rapid distribution. The common titles contain words like carcinogenic, lethal, poisonous, effective and so on.

尚需完成的任务 Work to do:

Since we have already collected the features of rumors, next we need to transform these qualitative data into quantitative data in order to build classifiers to predict more rumors if needed.

Next, we have to conduct a large-scale analysis, which means we need more URLs to identify rumors because only 29 rumors are not enough to provide a reasonable result. If more URLs are provided, we will get a more accurate prediction of rumors.

If we have more fake news identified, maybe we can use “brand recognition” to check the source of articles. There is a higher possibility that the articles written by the same authors

may be fake news.
Analysis of spread of rumors is also needed to be done. Because when we collecting the information about Wechat Moment, we also collected the time stamps and ip addresses as well, we can analyze the spread of rumors based on these information. Besides, some websites such as http://www.fibodata.com can also be used to analyze lots of information about Wechat Moments such as number of viewings, ip addresses, spread path, users' locations and so on.
Finally, we will look into characteristics about rumor articles and provide some new plots to better evaluate the result of rumor identification and rumor spread.

存在问题 Problems:

1. The data set we have constructed about rumors in Wechat Moment is too small. In other words, we don't have enough ground truth to help us collect features of rumors.
2. After the rumor feature extraction, it is hard to build the classifiers to predict fake news since the features are qualitative and subjective to some extent. It is hard to describe the features quantitatively.
3. Spread of rumors is also a big challenge since we only have ip addresses and time stamps.

拟采取的办法 Solutions:

1. We will ask others to help us collect more raw data (URLs of Wechat Moments, its relevant ip addresses, time stamps and so on). Then, we can use the same method as before(time difference, other rumor identification platforms) to help us identify more rumors, resulting in a large data set of fake news.
2. After crawling the information of the whole passage, we will try to collect the highest frequency of appearance of rumor features and record quantitatively. Then, we will try to use some basic machine learning methods such as Support Vector Machine, Naïve Bayes and so on to predict rumors. We will improve the accuracy based on a large-scale of data.
3. We will try to refer to other platforms such as http://www.fibodata.com to collect more information about user behaviors. Then we will provide a more precise analysis of rumor spread based on these user behaviors.

论文结构 Structure of the final report:

Abstract
Introduction
Related Work
  Identification of Rumors
  Spread of Rumors in Wechat
Methodology
  Data Set Construction
  Rumor Identification
  Rumor Feature Extraction
  Spread of Rumors
Result and Evaluation
  Analysis of Rumors Identification
  Analysis of Rumor Spread
Conclusion and Future Work
Reference

北京邮电大学 本科毕业设计（论文）教师指导记录表
Project Supervision Log

学院 School	International School	**专业 Programme**	**Telecommunications Engineering with Management**		
姓 Family name	Gong	**名 First Name**	Jiaying		
BUPT 学号 BUPT number	2015212931	**QM 学号 QM number**	151009262	**班级 Class**	2015215103
论文题目 Project Title	The spread of rumors in Wechat				

Please record supervision log using the format below:

Date: dd-mm-yyyy
Supervision type: face-to-face meeting/online meeting/email/other (please specify)
Summary:

Date: 13-07-2018
Supervision type: online meeting
Summary: introduced the project

Date: 26-07-2018
Supervision type: email
Summary: received written feedback on the problems I have met

Date: 03-08-2018
Supervision type: online meeting
Summary: discussed the project progress

Date: 20-08-2018
Supervision type: online meeting
Summary: discussed the project progress

Date: 24-08-2018
Supervision type: email
Summary: received written feedback on the problems I have met

Date: 07-09-2018
Supervision type: online meeting
Summary: discussed the project progress and challenges

Date: 08-09-2018
Supervision type: email
Summary: received written feedback on the problems I have met

Date: 18-09-2018

Supervision type: email
Summary: checked the traffic of data whether usable or not

Date: 27-09-2018
Supervision type: face-to-face meeting
Summary: discussed the project progress and challenges

Date: 29-10-2018
Supervision type: email
Summary: received written feedback on the problems I have met

Date: 15-11-2018
Supervision type: email
Summary: received written feedback on the problems I have met

Date: 17-11-2018
Supervision type: email
Summary: discussed the project specification

Date: 19-11-2018
Supervision type: email
Summary: received written feedback on the project specification

Date: 29-11-2018
Supervision type: email
Summary: created a sharing document on overleaf to better display the words and plots

Date: 11-12-2018
Supervision type: online meeting
Summary: discussed the project progress and problems I met

Date: 17-12-2018
Supervision type: email
Summary: reported feedback on several related papers

Date: 25-01-2019
Supervision type: online meeting
Summary: discussed the project progress and problems I met

Date: 31-01-2019
Supervision type: other(Skype chatting)
Summary: discussed the project progress and problems I met

Date: 15-02-2019
Supervision type: other(Skype chatting)
Summary: discussed the project progress and problems I met

Date: 21-02-2019
Supervision type: email

Summary: discussed the progress of the project

Date: 29-03-2019
Supervision type: email
Summary: updated project progress and data problems

Date: 04-04-2019
Supervision type: online meeting
Summary: discussed the project progress and solutions of problems

Date: 08-04-2019
Supervision type: email
Summary: received written feedback on the draft report of introduction and background section

Date: 13-04-2019
Supervision type: email
Summary: received written feedback on the draft report of part of implementation section

Date: 15-04-2019
Supervision type: other(Skype)
Summary: discussed the project progress by chatting

Date: 18-04-2019
Supervision type: email
Summary: received written feedback on the submitted draft report

Date: 21-04-2019
Supervision type: email
Summary: discussed the problems of IP addresses in draft report

Date: 24-04-2019
Supervision type: online meeting
Summary: mock viva

Risk Assessment

This project focuses on analyzing the spread of rumors in WeChat.

Table 6: Risk Table

Description of Risk	Description of Impact	Likelihood rating	Impact rating	Preventative actions
Not enough raw data	The result is less accurate	1	1	Always follow up with the new data
IP addresses are exposed	Personal Privacy	0	1	Not show IP addresses in public

Environmental Impact Assessment

There is almost no environmental impact in this project. No cost of manufacture, no waste of disposal and recycling and almost no energy use in service. Because all we need in this project is only a laptop which can connect to the website. All the energy we use mainly is electricity. For savings in energy because of using electronic devices, there are lots of savings on travelling. Mentor and I always have video conferencing instead of travelling to have face-to-face talks. Additionally, there are also many savings on plants since we do not need any paper to conduct this project.

In conclusion, besides electricity usage, there is almost no environmental impact in this project.