Research Article

Open Access

Lu An*, Xingyue Yi, Yuxin Han, Gang Li

# An Influence Prediction Model for Microblog Entries on Public Health Emergencies

https://doi.org/10.2478/dim-2018-0013
received July 2, 2018; accepted September 5, 2018.

**Abstract:** This study aims at constructing a microblog influence prediction model and revealing how the user, time, and content features of microblog entries about public health emergencies affect the influence of microblog entries. Microblog entries about the Ebola outbreak are selected as data sets. The BM25 latent Dirichlet allocation model (LDA-BM25) is used to extract topics from the microblog entries. A microblog influence prediction model is proposed by using the random forest method. Results reveal that the proposed model can predict the influence of microblog entries about public health emergencies with a precision rate reaching 88.8%. The individual features that play a role in the influence of microblog entries, as well as their influence tendencies are also analyzed. The proposed microblog influence prediction model consists of user, time, and content features. It makes up the deficiency that content features are often ignored by other microblog influence prediction models. The roles of the three features in the influence of microblog entries are also discussed.

**Keywords:** influence prediction, microblog, public health emergency, topical features, random forest, latent Dirichlet allocation

## 1 Introduction

Public health emergencies generally pose serious threats and significant losses to public health, economic development, and social stability. Since they break out in a short time and spread rapidly, the management departments of public health emergencies often face enormous challenges. They need to take effective measures in time to prevent the spread and upgrade of events, as well as eliminate the source of hazards and subsequent influence promptly.

With the continuous development of Web 2.0 and mobile Internet technology, microblog platforms such as weibo.com in China have shown increasing vigorous vitality. They have the obvious advantages that everyone can participate in the information exchange with low cost of producing information, convenient information exchange, and strong interactive functions. Thus, various types of institutions and individuals use microblogging as new information publishing platforms to enhance their communication power and influence. It is found that the results of Internet-based surveillance systems are generally consistent with the results from traditional monitoring methods, and these systems can be viewed as extensions of traditional surveillance systems (Milinovich, Williams, Clements, & Hu, 2014).

As microblog platforms increasingly become important sources of information, when public health emergencies occur, the relevant management agencies, news media, and the general public usually post a multitude of related contents, such as the spread of diseases, symptom descriptions, treatment plans, rescue progress, and epidemic control, on microblog platforms (An, Wu, & Yu, 2017). For example, the Ebola outbreak in West Africa in 2014 led to hundreds of thousands of records in major microblog platforms in China and the United States, which were far away from the outbreak center. According to Towers et al. (2015), each piece of video news about the Ebola outbreak triggered tens of thousands of microblog entries and Web searches on average. As a kind of crowdsourcing information, the user-generated content, such as microblog entries of high influence, provides a valuable source to detect potential risks of emergence. The emergency management departments can use social media to gain insight into public opinions during public health events (Finch et al., 2016). The microblog users

***Corresponding author: Lu An:** Center for Studies of Information Resources, Wuhan University, Wuhan, China, E-mail: anlu97@163.com
**Xingyue Yi, Yuxin Han:** School of Information Management, Wuhan University, Wuhan, China
**Gang Li:** Center for Studies of Information Resources, Wuhan University, Wuhan, China

who contribute information during crises play the role of "digital volunteers" (Starbird & Palen, 2011).

The massive information storage capacity of the microblog platforms provides sufficient open source information for relevant studies on public health emergencies. So far, scholars have conducted a huge volume of research using the information related to public emergencies, such as emergency monitoring, emergency information dissemination, and user behavior during emergencies, available on microblog platforms. Among them, prediction of the influence of microblog entries is an important task.

The influence of a microblog entry refers to how much attention a microblog entry attracts. In this study, we measure the influence of a microblog entry by the sum of its forwarding, commenting, and favorite counts. High-influence microblog entries of public health emergencies may have positive social impacts, such as dissemination of public health knowledge and important prevention measures, as well as negative impacts, such as the spread of rumors and false advertising, which may cause public unease or aversion and trigger secondary events online or in reality. Locating the microblog entries that may have high influence can help emergency management departments foresee potential problems and take precautions in advance. Management departments of public health emergencies are in urgent need of methods and rules to predict high-influence microblog entries from the massive amount of microblog information, discover upcoming problems, and improve the prospectiveness of decision-making.

At present, relevant research works on microblog influence are abundant. However, research on the influence of microblog in specific fields, such as public health emergencies, is relatively insufficient. This study attempts to propose a microblog influence prediction model for public health emergencies, which is composed of user, time, and content features and which uses the random forest method (Breiman, 2001) and the Best Match 25-based latent Dirichlet allocation model (LDA-BM25) (Li, 2013). As this model is constructed specifically for public health emergencies, it highlights the features of public health emergencies, such as the microblog accounts of medical institutions in the publisher types, identifying whether the publisher belongs to the medical industry and the development stage of the public health emergency. It provides a sound resolution to predict the influence of microblog entries on public health emergencies and reveals the relationships between the features of microblog entries and their influence tendency.

The findings of the study can help relevant management departments understand the public behavior patterns, effectively identify the potential influential microblog entries, and carry out targeted emergency response measures for public health emergencies. They can help management departments generate influential microblog information to effectively guide online public opinions. Through analysis of the evolution of the high-influence microblog entries, the management departments will be able to both understand the public's concerns about the public health emergencies and discern the differences among the intensity of the public's attention toward different issues at different development stages of public health emergencies.

# 2 Related Research

## 2.1 Analysis of the key factors of microblog influence

The influence of microblog entries refers to the extent of and frequencies by which microblog entries have been read, used, approved, or criticized. It can be measured by various indicators, such as likes, comments, forwarding, replies, and favorite counts for Sina Weibo, as well as retweets and favorite counts for Twitter.

Studies on the influence of microblogs can be divided into the following: a) those on the influence of microblog users; b) the studies on the influence of microblog contents; and c) those on the influence of microblog topics. Researchers are curious about the factors that contribute to the high influence of some microblog entries and have conducted some research on these factors by various methods, such as questionnaire surveys, content analyses, and statistical analyses. These factors include the number of followers (Ye & Wu, 2010), user behavior such as users being followed or mentioned and microblog entries being forwarded (Cha, Haddadi, Benevenuto, & Gummadi, 2010), the number of friends (Weng, Lim, Jiang, & He, 2010), types of microblog contents (such as entertainment and life inspiration) (Sun & Li, 2012), users' identities, the format of microblog contents (such as text and pictures) (Zhao & Zeng, 2014), and the posting time of the microblog entry (Zhang, Lu, & Yang, 2012).

These factors can be directly obtained from the microblog information, which are called direct indicators, such as the user authentication status, the number of microblog entries posted, mentions, topics, content, and

emotional tendencies of published microblog entries (Riquelme & González, 2016). Among the studies on the influence of microblog users, scholars have also integrated or combined multiple direct indicators by constructing certain index systems or adopting mathematical modeling methods, in addition to putting forward indirect indicators, such as a mathematical model based on the PageRank algorithm to define the integrated indicator of users' influence (Kim, Lee & Straub, 2018), the location of the users' network structure, and the dynamic interaction of the user in the network structure (Hong, He, Ge, Wang, & Wu, 2017).

The evaluation methods of microblog influence include PageRank based on the number of followers (Kwak, Lee, Park, & Moon, 2010), the topic-sensitive method of TwitterRank (Weng, Lim, Jiang, & He, 2010), Twitter User Rank (TURank) based on user behavior (Yamaguchi, Takahashi, Amagasa, Kitagawa, & Kitagawa, 2010), the method based on the order relationship of forwarding (Bakshy, Hofman, Mason, & Watts, 2011), the InfluenceRank algorithm applied by constructing a regression model (Nargundkar & Rao, 2016), the WeiboRank algorithm from the comment dimension (Liao, Wang, Han, & Zhang, 2013), the personalized PageRank algorithm that considers the user's network topology characteristics (Alp & Öğüdücü, 2018), and the Learn-to-Rank algorithm that relates the user's behavior to the sentiment of published content (Chen, Liu, & Zou, 2017). Some researchers have tried to establish evaluation indicator systems for microblog influence, such as microblog influence evaluation system for major emergencies (Luo, 2013) and the evaluation system of information dissemination influence of individual microblog users (Xiao & Qi, 2013).

It shows that existing studies on the influence of microblog information generally explore the key factors, evaluation methods, and evaluation systems. Their research findings lay a theoretical foundation for predicting the influence of microblog entries. However, studies on the key factors of the influence of microblog entries mainly focus on revealing external characteristics, such as users' features, information types, posting time, and formats of microblog contents. Studies on internal characteristics (such as topical features) of the relevant high-influence microblog entries in specific areas (such as public health emergencies) and those on the coordination between internal and external characteristics are relatively insufficient.

## 2.2 Studies on predicting microblog influence

Studies on predicting microblog influence can be carried out from two perspectives, i.e., predicting the behavior of microblog users, and predicting microblog influence by judging the salience of microblog topics.

The influence of a microblog entry refers to how much attention a microblog entry attracts. It can be reflected by the users' behaviors, such as forwarding, commenting, and clicking on the favorite button. For example, if many users forward, comment on, or click on the favorite button of a microblog entry, it means that the microblog entry has high influence. In related research, the forwarding count is usually taken as a common metric for the influence of a microblog entry, while users' favorite counts, commenting, or other behaviors are rarely investigated.

This kind of influence prediction models are mainly based on users' characteristics to predict whether users will forward a microblog entry. The attributes of the forwarding behavior consist of forwarding quantity, forwarding time, and forwarding intention according to the research purpose (Palovics, Daroczy, & Benczur, 2013). Researchers predict microblog forwarding behavior through a variety of algorithms and models, including the factor graph model (Yang, Guo, Cai, Tang, Li, Zhang, & Su, 2010), social network analysis and content analysis (Kim, Hou, Han, & Himelboim, 2016), the classification method (Hong, Dan, & Davison, 2011), the probabilistic collaborative filtering model (Zaman, Herbrich, Gael, & Stern, 2010), improved passive-aggressive algorithm (Petrovic, Osborne, & Lavrenko, 2011), the predictive model for retweeting based on conditional random fields (CRFs) (Peng, Zhu, Piao, & Yan, 2011), and so forth. It has been found that users' competition or cooperative interaction (Kong, Mao, & Liu, 2016), characteristics of the forwarding agent (Maleewong, 2016), and network structure (Gao, Ma, & Chen, 2014) all influence their forwarding behavior. However, the accuracies of these predictions are very limited. The highest accuracy of these algorithms only reached 69.3%. Zhang et al. (2012) predicted retweeting by using support vector machine (SVM) based on weighted features, and their accuracy rate reached 86%. On this basis, Li et al. (2013) reanalyzed the influence factors of forwarding behavior, using the SVM classification algorithm again to predict forwarding behavior, and reached an accuracy rate exceeding 93%. Luo et al. (2014) proposed a forwarding prediction algorithm based on random forest, and the precision rate reached 92.9%. The commonly used machine learning methods include the BP neural network, logistic regression analysis, SVM,

the hierarchical Bayesian model, and so forth. The above studies can predict microblog influence to some extent in terms of forwarding scales.

Secondly, from the perspective of predicting microblog topics' salience and trends, researchers mainly choose Sina Weibo and Twitter as data sources. Such studies take microblog information as the research object and predict the mode, scope, and cycle of information dissemination. Relevant achievements include the prediction model for future trends of microblog events based on regression analysis (Tian, 2010), a method to monitor real-time risk information dissemination trends in a microblog system (Pei, Yu, Tian, & Donnelley, 2017), the information diffusion model of user characteristics (Liu, Ding, Hao, & Huang, 2016), the multiagent information diffusion model (Lemahieu, Canneyt, Boom, & Dhoedt, 2015), prediction of trending topics based on the stochastic model (Asur, Huberman, Szabo, & Wang, 2011), prediction of popularity of news items on Twitter based on the multidimensional feature space (Bandari, Asur, & Huberman, 2012), the prediction model for long-term development trends of online public opinions (Gao, Wang, & Fu, 2011), prediction of topics' popularity in microblogs based on wavelet transformation (Chen, Fang, Guo, & Guo, 2015), and prediction of microblog popularity based on forwarding level (Zhai, Liu, Cheng, Hu, & Li, 2015).

In the aspect of microblogging topical heat prediction, some methods have been put forward, such as a susceptible–antidoted–infectious–recovered (SAIR) model based on the epidemic-information propagation model (Liu et al., 2016), a Bayesian network model constructed through a social network chain (Varshney, Kumar, & Gupta, 2017), a maximum-likelihood adaptive neural system to predict microblog topics (Gromov & Konev, 2017), and the method of identifying rumors in the public health field based on LDA (Ye, Li, Yang, Lee, & Wu, 2018). Some scholars have constructed the "conductivity" attribute of topics, introduced the physical concept to microblog information prediction, and predicted whether the topic can be achieved by calculating the second derivative of conductance (Bora, Singh, Sen, Bagchi, & Singla, 2015).

In summary, researchers have proposed a variety of microblog influence evaluation systems and models. However, these evaluation methods mainly focus on the influence of microblog users. As for specific issues, such as the influence of microblog entries regarding public health emergencies, less attention has been paid. In the studies on microblog influence prediction, researchers have achieved some results by continuously establishing and optimizing prediction models. Nevertheless, the research contents of existing literature are mostly topics' features such as popularity and trends. Internal factors such as the content features of topics have been hardly analyzed together with external factors such as publishers and publishing time. In fact, when the microblog entries are limited to those related to specific issues (such as public health emergencies), the role of microblog topics is equally important as the role of microblog publishers, which researchers usually consider. Revealing topical features of microblog entries with high influence will provide management departments of public health emergencies much inspiration for predicting the influence of relevant microblog entries. For instance, regardless of publishers and publishing time, microblog entries on successfully developing drugs for major infectious diseases or the spread of infectious diseases to new areas are likely to have higher influence. As for certain features of publishers, for instance, publishers belonging to authorities or from an industry that highly correlates with the topics of the microblog entries may further enhance the influence of microblog entries.

As for influence indicators, most researchers have used microblog forwarding behavior as a predictor rather than using comments, favorite counts, and other influence indicators. Therefore, one research task of this study is to identify the internal or external key factors that have impacts on influence of microblog entries regarding public health emergencies and thereafter establish a microblog influence prediction model.

# 3 Data and methods

## 3.1 Data collection and preprocessing

The data used this study come from weibo.com, which is a well-known microblog platform in China. As of March 2016, Sina Weibo has >261 million monthly active users (Sina Technology, 2016). The 2014 West African Ebola outbreak is taken as the investigated case in this study since it has been one of the most prevalent and major public health emergencies in recent years. Microblog entries on the Ebola outbreak are selected as samples. With the keyword "Ebola"(埃博拉 in Chinese), a total of 230,274 relevant microblog entries posted between February 1 and October 31 in 2014 were collected. The influence of microblog entries was measured by the sum of their forwarding, commenting, and favorite counts. After examining the influence distribution of microblog entries, we set a threshold of 20 for discriminating low-influence messages

from high-influence ones to ensure that the high-influence messages have an obviously higher sum of forwarding, commenting, and favorite counts than the low-influence ones. Microblog entries with a total count of not less than 20 were considered as high-influence microblog entries. Those with a total count of zero were considered as low-influence microblog entries. A total of 14,567 microblog entries were selected as test data, 7,101 of which were of high influence and 7,466 were of low influence.

## 3.2 Feature extraction of microblog entries

The features of microblog entries can be divided into three types, i.e., features of publishers, time, and contents. As this study aims to predict the influence of microblog entries for public health emergencies, some unique features are considered, such as whether the publisher belongs to the medical industry and the development stage of epidemics. Thus, publishers' features include publishers' type and whether the publisher belongs to the medical industry, and temporal features include period of time, day of the week, whether the publishing date is an official holiday, and the development stage of epidemics. As the topical features are seldom investigated in similar studies, content features are also considered and represented by the topics involved in microblog entries.

First, in the category of publishers' features, the publisher's type is determined by the certification information and the name of the publisher. According to the Sina Weibo's certification system, all of the certified users are divided into seven classes, e.g., official microblog accounts of new media. The users with certification of we-media or personal authentication are considered as celebrities. Those unauthorized are considered as grassroots users. As public health emergencies may involve medical knowledge, whether the publisher belongs to the medical industry is considered as a separate publisher feature.

Second, in the category of temporal features, the time period of a day is divided into six intervals (Wang, Zheng, Wang, Xiong, & Xie, 2014). A week consists of 7 days. Each day is categorized into an official holiday or a working day. The Ebola development is divided into two stages according to the World Health Organization (WHO). The dates before the time when the WHO declared the Ebola outbreak as a public health emergency are regarded as the initial period and those after that are regarded as the outbreak period (Jia, An, & Li, 2015).

Third, the content features of microblog entries are represented by the topics identified by the LDA-BM25. A total of 50 topics (Topic0–Topic49) have been found, and each topic consists of 50 terms. If a microblog entry's content contains one or more term(s) of a topic, the value of the element for the topic and the microblog entry in question is one. Otherwise, it is zero. The features, serial numbers, and the value range for each feature are shown in Table 1. The influence tendency will be explained later.

## 3.3 Topic discovery algorithm based on LDA-BM25

The LDA model (Blei, Ng, & Jordan, 2003) is a common topic model. The LDA model is a Bayesian probabilistic model of topic mining for natural language, and a three-level hierarchical model of words, topics, and documents. It can identify the potential topical information in a large-scale document collection or corpus. However, the LDA model has limitations in dealing with the short text of microblog entries, since the microblog entries have little amount of information, high-dimensional word vector space, nonstandard language, and highly context-dependent content. The generated topics are quite sparse with poor readability.

In order to protect the independence and enhance the readability of topics, the LDA-BM25 algorithm (Li, 2013) is used to identify topics from microblog entries. The principle of LDA-BM25 is to identify the topics from the documents using the LDA model. The word with the highest distribution probability in each topic is regarded as the core word of the topic. Other words in this topic are regarded as candidate words. The BM25 similarity between the core word and each candidate word is calculated respectively. Then, the candidate words are sorted by their similarity to the core word. For each topic, the core word and the top $m$ candidate words with the highest similarity are chosen to form a new topic.

## 3.4 The classification models

Commonly used classification models include the C4.5 decision tree (DT), Bayesian belief networks (BN), naive Bayes (NB), logistic regression (LR), multilayer perceptron (MLP), SVM, random forest, and so forth. LR can be used to construct classification models and regression models. An LR model classifies the continuous regression functions by a step function. However, the disadvantage is that it is difficult to process data sets of high-dimensional features. The data need to be linearly processed, and it is sensitive to missing data values. NB uses the Bayesian

theorem in statistical probability as a classifier to solve the classification problem. The model calculates the probability that the samples of the unknown category belong to each category based on the assumption that the individual features in the data set are independent of each other. However, in practical applications, the independent assumptions are difficult to apply.

Random forest is a classification and prediction algorithm proposed by Leo Breiman (2001). The essence is to use the bootstrap method to sample the data set and generate *k* training sets. Each training set is used to infer a separate classification and regression tree (CART), generated by extracting *m* attributes from all *M* attributes without pruning. Random forest is a combination of *k* decision trees. To solve practical problems, each tree selects classification attributes and votes. The output is decided by the votes from each decision tree. The final classification results are the most popular classes with the most votes. The advantage of random forest is that it is able to tolerate noise, does not need feature selection, has strong processing capacity with high-dimension attributes, and will not overfit. Its random strategy allows for greater variability between subclassifiers in the random forest, resulting in superior classification performance (Fu & Chen, 2014). Therefore, it can work well with microblog attributes with complex composition and high dimensions.

The random forest uses the bootstrap method, a method of sampling with replacement. Some of the samples are not used in the training process, but for testing. Assuming that the sample size is *N*, the probability of not sampling is $1-1/N$£$^{N}$. When *N* tends to infinity, about 36.8% of the samples are used for testing, which are known as out-of-bag (OOB) data. The error estimation that uses the OOB data is called OOB error estimation. The generalization error estimation of the random forest is the average value of the OOB error estimation for each tree (Breiman, 2001). Breiman has shown that the OOB error estimates have the same accuracy as the test set with the same size; thus, there is no need to set a test set.

In this study, we experiment with seven classification models and find the one with the best performance. In order to evaluate the prediction performance of the classification model, this study uses four common performance metrics of classifiers as evaluation criteria, i.e., precision, recall, F-measure, and receiver operating characteristic (ROC) curves. Among them, precision is a measure of exactness, which calculates the percentage of objects that the classifier labeled as positive are actually positive. Recall is a measure of completeness, which calculates the percentage of positive objects that the classifier labeled as positive. F-measure is the harmonic mean of precision and recall, which equals two times of precision and recall divided by the sum of precision and recall. The ROC curve is a trade-off between true positives and false positives. The area under this curve is a measure of the accuracy of the model. The closer the area is to 0.5, the less accurate the corresponding model is, and vice versa (Han, Kamber and Pei, 2012).

This study uses WEKA (Ian & Eibe, 2003) to establish the random forest model. WEKA is a free open source data mining platform based on Java, which is a combination of abundant data mining algorithms.

# 4 Result analysis and discussion

## 4.1 Prediction of microblog entries' influence

As a total of 14,567 microblog entries and 57 features were investigated, a matrix of 14,567 ´ 57 was obtained. The random forest algorithm was first used to construct the influence prediction model.

To ensure maximum accuracy, the maximum depth of the tree and the random number seed are set to the default values, i.e., zero and one, respectively. The OOB error estimate and precision are taken as primary indicators. The number of selected features (*m*) and the number of decision trees (*k*) are to be adjusted. The optimal result is to achieve the minimum OOB error estimate while ensuring the highest precision.

Assuming that the total number of features is *M*, *m* is generally close to $\sqrt{M}$ (Breiman, 2001). Experiments have shown that when *m* equals seven, the optimal value of *k* is 222. Keeping the value of *k* unchanged and adjusting *m* to its optimal value, the final value of *m* is always six. To simplify the experiment process, *k* is adjusted to its optimal value while *m* equals six. Figure 1 shows the OOB error estimation and the precision of random forest when *m*=6.

Figure 2 shows the relationship between *m* and the model performance when *k*=204. The precision arrives at its highest value when *m* equals 6. The final parameters of the random forest model are *m*=6 and *k*=204.

In order to evaluate the performance of the random forest model, the precision, recall, F-measure, and ROC curve are used as evaluation criteria. The final results are shown in Table 2.

Table 1
*List of Microblog Entries' Features*

| Features of microblog entries | | | Range of values | Influence tendency | |
|---|---|---|---|---|---|
| User features | 1 | Publisher type | Official microblog account of new media | 0.38 | High |
| | | | Microblog account of traditional media | 0.45 | High |
| | | | Official microblog account of enterprises | 0.23 | High |
| | | | Official microblog account of governments | 0.21 | High |
| | | | Microblog account of colleges and scientific institutions | 0.17 | High |
| | | | Microblog account of nongovernmental organizations | 0.44 | High |
| | | | Microblog account of medical institutions | 0 | Low |
| | | | Celebrities | 0.45 | High |
| | | | Grassroots user | −0.24 | Low |
| | 2 | Whether the publisher belongs to the medical industry | Yes | 0.24 | High |
| | | | No | −0.02 | Low |
| Temporal features | 3 | Period of time | Night (00:01 am–6:00 am) | 0.04 | High |
| | | | Early morning (6:01 am–8:30 am) | −0.06 | Low |
| | | | Morning (8:31 am–12:00 pm) | −0.04 | Low |
| | | | Noon (12:01 pm–2:00 pm) | −0.07 | Low |
| | | | Afternoon (2:01 pm–6:00 pm) | 0.04 | High |
| | | | Evening (6:01 pm–12:00 am) | 0.04 | High |
| | 4 | Day of the week | Monday | 0 | Low |
| | | | Tuesday | 0.05 | High |
| | | | Wednesday | −0.05 | Low |
| | | | Thursday | 0 | Low |
| | | | Friday | 0.11 | High |
| | | | Saturday | 0 | Low |
| | | | Sunday | −0.13 | Low |
| | 5 | Whether the publishing date is an official holiday | Yes | −0.18 | Low |
| | | | No | 0.11 | High |
| | 6 | Development stage of epidemics | Initial period | −0.08 | Low |
| | | | Outbreak period | 0.02 | High |
| Content features | 7–56 | Topics (Topic0–Topic49) | 1 | | |
| | | | 0 | | |

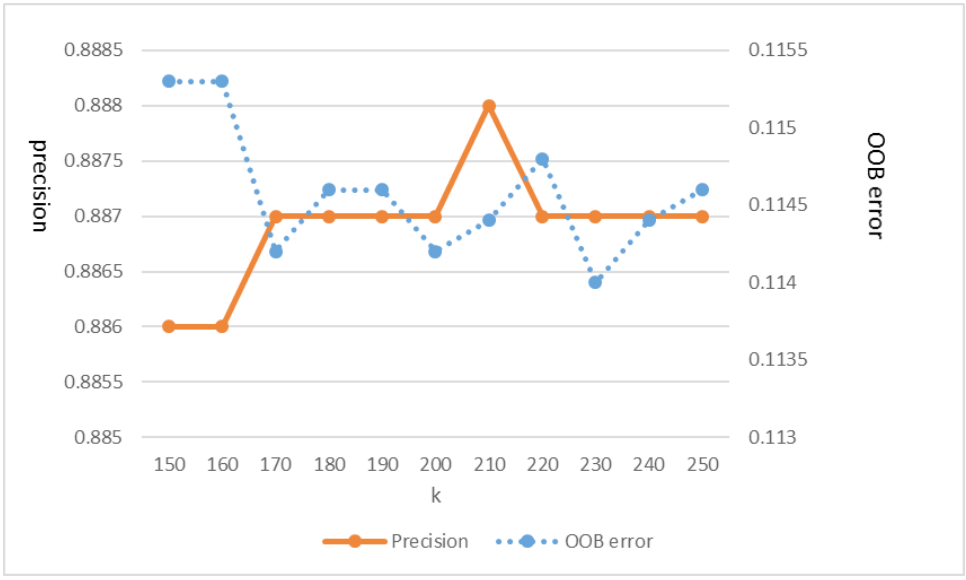


*Figure 1(a)*. {*k*|*k*=150+10*i*, 0≤*i*≤10, *i*∈N}

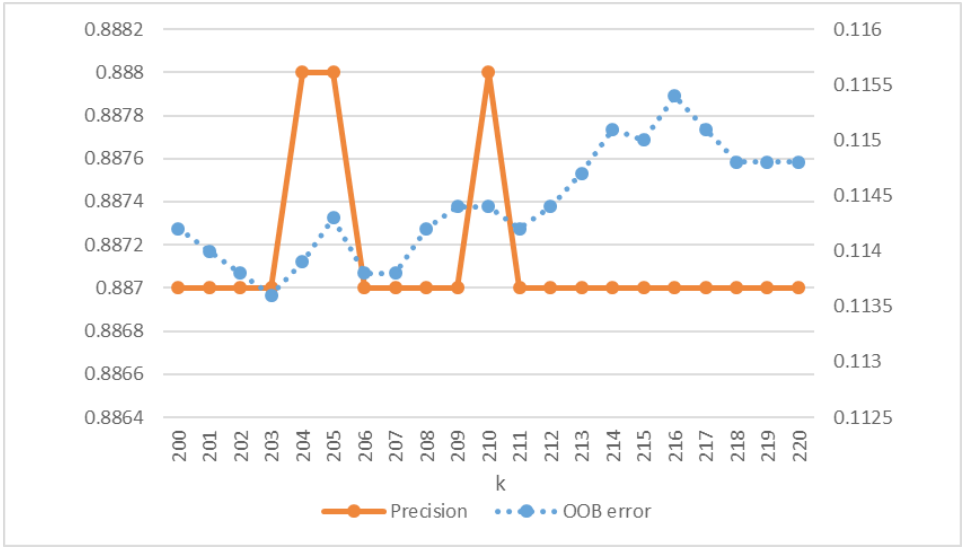


*Figure 1(b)*. {*k*|200≤*k*≤220, *k*∈N}

***Figure 1.*** **The number of decision trees *k* and the model's performance when *m*=6.**

The experimental results show that the precision of the influence prediction model reaches $\frac{6252+6681}{6252+849+785+6681}=88.8\%$, which is higher than the precision rates of the forwarding prediction model of microblogs by Petrovic et al. (2011) (69.3%) and the model by Zhang et al. (2012) (83%).

To verify whether the performance of the random forest algorithm is better than that of other algorithms, the DT, BN, NB, LR, MLP, and SVM models were selected to compare with the random forest. The results are shown in Table 3. The maximum value of each indicator is shown in bold. Table 3 shows the superiority of the random forest model for each indicator compared with other algorithms.

**Table 2**
***The Confusion Matrix of the Random Forest Model***

| | | Predicted value | |
|---|---|---|---|
| | | **High influence** | **Low influence** |
| Actual value | High influence | 6,252 | 849 |
| | Low influence | 785 | 6,681 |

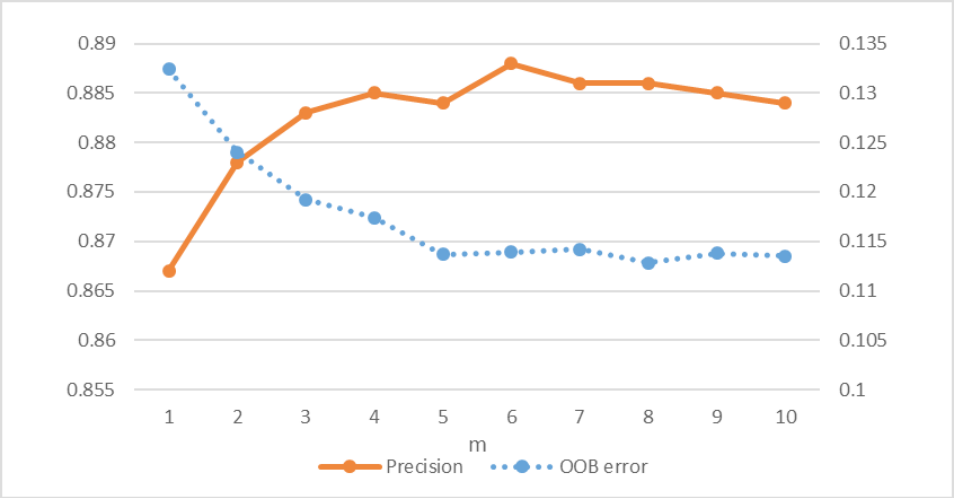


***Figure 2.*** **The number of selected features *m* and the model's performance when *k*=204.**

**Table 3**
***Performance Comparison among Multiple Algorithms***

| | RF | DT | BN | NB | LR | MLP | SVM |
|---|---|---|---|---|---|---|---|
| Precision | **0.888** | 0.858 | 0.754 | 0.749 | 0.825 | 0.857 | 0.879 |
| Recall | **0.888** | 0.858 | 0.754 | 0.749 | 0.823 | 0.857 | 0.878 |
| F-measure | **0.888** | 0.858 | 0.754 | 0.749 | 0.823 | 0.857 | 0.878 |
| ROC area | **0.952** | 0.891 | 0.836 | 0.83 | 0.905 | 0.925 | 0.877 |

RF = random forest model.

## 4.2 The relative importance of microblog entries' features

To determine the importance of individual features to the model, we calculated and compared the precision of different influence prediction models lacking a certain feature with that of the model with all of the features. Figure 3 shows the results of the comparison.

Experimental results show that the removal of any feature leads to a decline in the precision of the influence prediction model. The relative importance of each feature can be revealed by the decrease in precision of the model after removing the feature in question.

Figure 3 shows that the most important feature is the publisher's type, followed by topics. The four temporal features (e.g., period of time) also have considerable influence. The findings coincide with those of a previous paper (Cao, Huang, & Tu, 2016), which found that the development stage of an event and the release time of a microblog entry contribute much to microblogs' influence. The experimental results verify the impact of microblog entries' topical features on their influence. Considering the unique features of public health emergencies, such as whether the publisher belongs to the medical industry and the development stage of epidemics, can also improve the precision of the influence prediction model. As explained in the section "Related Research", existing studies usually examine the number of followers (Ye & Wu, 2010), the number of friends (Weng, Lim, Jiang, & He, 2010), or users' identities (Zhao & Zeng, 2014) when they consider publishers' features. The publisher's type was seldom investigated, while in this study, it was found to be the most important feature. Similarly, previous studies considered the types of microblog contents (such as entertainment and life inspiration) (Sun & Li, 2012) and the format of microblog contents (such as text and pictures) (Zhao & Zeng, 2014). Specific topics were seldom examined to predict the influence of microblog entries. Nevertheless, in this study, topics were found to be the second most important feature. Certainly, this study is different from others in that the prediction model is for microblog entries of public health emergencies. That is why the features of "the medical industry" and "development stage of the epidemic" were investigated and found to contribute to the influence.

## 4.3 Analysis of the influence tendency of microblog features

To better explore the characteristics of high-influence microblog entries for public health emergencies, we propose the concept of influence tendency of microblog features. Since different values of a microblog feature may be associated with different rates of influence, the

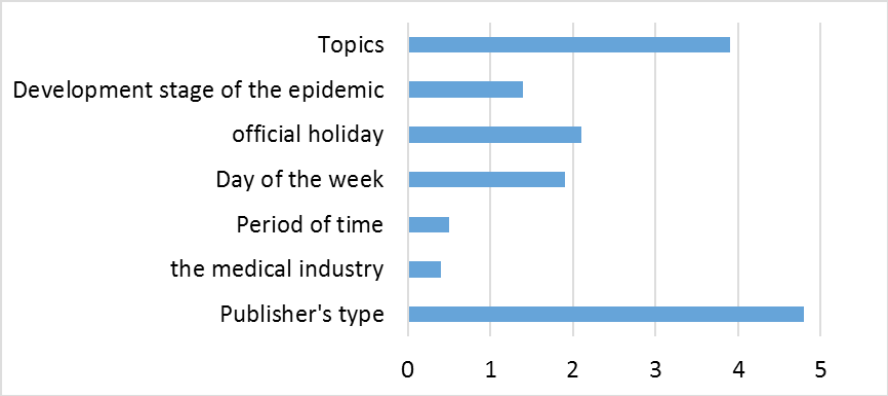


***Figure 3.*** **The decrease in precision of the influence prediction model observed after removing each of its features**

influence tendency of a microblog feature aims to reveal the relationship between the value of a microblog feature and the influence of the microblog entries in question. The calculation method is to count the numbers of high-influence and low-influence microblog entries less than a certain value of a feature and to compute the difference between the proportion of high-influence microblog entries less than this feature value and the proportion of high-influence ones among the whole microblog corpora, as shown in Eq. (1).

$$\text{Inf\_Inc}(V_i) = \frac{num(V_i|V_c=High)}{num(V_i)} - \frac{num(D|V_c=High)}{num(D)} \quad (1)$$

where $\text{Inf}\_\text{Inc}(\text{V}_\text{i})$. is the influence tendency of the attribute value *i*, $\text{num}(\text{V}_\text{i} \mid \text{V}_\text{c} = \text{High})$ is the number of microblog entries with high influence for the attribute value *i*, and $\text{num}(\text{V}_\text{i})$ is the number of microblog entries for the attribute value *i*. D refers to all the microblog entries, num(D) is the total number of microblog entries and num(D|$V_c$=High) is the total number of microblog entries with high influence in the microblog corpora. When Inf_Inc($V_i$) equals 0, the attribute value *i* has no influence tendency. When Inf_Inc($V_i$) is higher than 0, the attribute value *i* shows high influence tendency. When Inf_Inc($V_i$) is lower than 0, the attribute value *i* shows low influence tendency. The higher the value of $\text{Inf}\_\text{Inc}(V_i)$, the more obvious the influence tendency of the attribute value is.

## 4.4 Analysis of the influence tendency of user features and temporal features

Table 1 shows the influence tendency of the 28 attribute values, 13 of which have low influence tendency and 15 values have high influence tendency.

It is observed that seven types of publishers, i.e., official microblog accounts of traditional media, celebrities, nongovernmental organizations, new media, enterprises, governments, colleges, and scientific institutions have high influence tendency. Grassroots users show low influence tendency. Publishers in the medical industry show high influence tendency, while those not in this industry show neutral influence tendency.

It is found that microblog entries that are posted during afternoon, evening, and night (between 2:01 pm and 6:00 am the next day) are more likely to have high influence, and those posted during early morning, morning, and noon (between 6:01 am and 2:00 pm) are more likely to have low influence. During a week, Friday tends to show high influence, while Sunday tends to have low influence. It is interesting to find that the microblog entries posted on holidays show low influence, while those posted on working days tend to show high influence. That is to say, microblog platforms are more used on working days than on holidays. It is easy to understand that microblog entries posted during the initial period tend to have low influence, while those posted during the outbreak period tend to have high influence. With the development of epidemics, more attentions are drawn.

**Table 4**

***Topics with the Highest/Lowest Influence Tendency***

| Rank | Topic no. | Top 20 terms | Topic summary | Influence tendency |
|---|---|---|---|---|
| 1 | Topic19 | Fight, Chinese medicine, negative, confirm, staff, dispatch, Reuters, bad, sick, AFP, quarantine, fear, afraid, south, foundation, west, Gates, air ticket, southeast, other, Bill | Medical assistance and charity | 0.2 |
| 2 | Topic13 | Infected person, Chinese medicine, body, discharge, infect, estimate, American, all countries, website, process, hearsay, hopeful, level, public, institute, Brantley, volunteer, lanjinger, unfortunate | The First US Ebola case; Reports of Ebola recovery cases | 0.18 |
| 3 | Topic25 | Fight, treat, dispatch, hour, future, Youth Olympic Games, hotspot, prevent, highly, fierce, regret, people, hospital, discriminate, possible, Liu Weizhong, influence, great, national, male, have promising prospects | Precautions against the Ebola virus at Youth Olympic Games; The Ebola prevention plan with traditional Chinese medicines | 0.17 |
| ... | | | | |
| 48 | Topic36 | Client, share, dead, poison, crowd, walk, stand up, address, fox, Liberia, front page, android, watch, wake, bury, die, on the way, walk around, passenger, mobile phone | Ebola patients with symptoms of "living dead" | −0.13 |
| 49 | Topic7 | Share, abuse, terrible, reply, outbreak, incoming, truth, state-owned, story, photograph, Beta, watch, death, despair, interesting, pray, side, economist | Photographs from the Ebola zone | −0.17 |
| 50 | Topic40 | Wonderful, coat, heighten, women's clothes, long sleeve, leather, windbreaker, men's clothes, spring and autumn, father, shoe, dress, England, authentic, middle age, trousers, loose, blouse, thin, shirt | Clothing advertisements | −0.19 |

### 4.5 Analysis of the influence tendency of content features

Experiments show that among the 50 topics, 38 have high influence tendency and the remaining 12 topics have low influence tendency. Table 4 shows the top three topics with the highest influence tendency and those with the lowest influence tendency.

Table 4 shows that the topics of high influence tendency are different from those of low influence tendency. Among the topics of high influence tendency, Topic19 is mainly about financial support for the frontline staff of the Gates Foundation and the public discussions triggered by frontline workers attacked by local people or those infected with viruses. Topic13 involves the real-time report of the Ebola outbreak and the condition of the infected people, including the first case of Ebola in the United States and the rehabilitation of the remaining patients. Topic25 is related to controversial events, including the role of Chinese medicine in combating the Ebola virus and the rumor of treating an Ebola carrier in Shanghai. As for the low influence tendency topics, Topic36 refers to some people who were afraid of infection and refused medical care in the affected areas. Topic7 is about the photographs taken in Liberia. Topic40 involves the clothing advertisements.

It is seen that topics with high influence tendency can be divided into three categories: a) reports on landmark events, such as "the first US Ebola case"; b) positive responses, such as "medical assistance and charity"; and c) controversial issues, such as "the Ebola prevention plan with traditional Chinese medicines". As for the topics with low influence tendency, such as "Guangdong ruled out the observed cases of Ebola", they usually have simple contents and target on limited users. Thus, they can hardly draw much attention from the public and show low influence tendency.

## 5 Conclusion

This study has measured the influence of microblog entries from the perspective of users' behaviors, such as forwarding, commenting, and clicking on the favorite button. The sum of forwarding, commenting, and favorite counts is taken as the influence indicator of microblog entries. An influence prediction model of microblog entries for public health emergencies has been proposed, in which

user, time, and content features are considered. As this study aims to predict the influence of microblog entries for public health emergencies, some unique features of public health emergencies are considered, such as whether the publisher belongs to the medical industry, the development stage of epidemics, and the topics identified by the LDA-BM25 algorithm. The concept of influence tendency of microblog features is proposed and analyzed for each attribute value. To test the performance of the model, the microblog entries about the Ebola outbreak on Sina Weibo were collected. The experimental results show that the influence prediction model can predict the influence of microblog entries regarding public health emergencies effectively, with the precision of the model reaching 88.8%. The relative importance of microblog entries' features for their influence and the influence tendency of microblog features are also explored.

It is found that making accurate prediction of microblog entries' influence requires a comprehensive consideration of all possible factors. The study shows that the internal characteristics of microblog entries (e.g., topics) play an important role in predicting the influence of microblog entries, and incorporating unique features of public health emergencies (e.g., whether the publisher belongs to the medical industry, the development stage of epidemics, etc.) into the influence prediction model of microblog entries can improve its prediction performance. Thus, the influence prediction model of microblog entries is supposed to be built with the consideration of the event's features.

The concept of influence tendency of microblog entries' features is proposed and analyzed in detail. The attribute values that are likely to lead to high influence of microblog entries are identified. For example, authenticated publishers and those of the medical industry are more likely to generate microblog entries of high influence. The microblog entries published on working days and during the time between afternoon and night time (from 2:01 pm to 6:00 am) have the highest influence. The microblog entries involving topics of landmark events, reports of positive responses, controversial issues, and rumors are more likely to gain high influence.

The constructed model can be further developed to an influence prediction system. When a new microblog entry is generated, the system can automatically determine its influence level in real time. The emergency management departments can evaluate its potential risk and respond in advance to avoid future problems. The management departments may also use the prediction system to estimate the influence of the microblog entries that they are going to post and find the best way to serve the public in terms of emergency response.

The limitation of this study is that an exact comparison between the proposed model and similar ones was not conducted. Although similar studies usually predict the forwarding behavior rather than the influence of microblog entries, which is represented by the sum of forwarding, commenting, and favorite counts in this study, we will further examine whether the performance of the model can be improved if the features in similar studies and new ones are also considered.

**Acknowledgments:** This study was supported by the Major Project of the Ministry of Education of China (grant no. 17JZD034), the National Natural Science Foundation of China (grant no. 71603189, 71790612) and China Postdoctoral Science Foundation (grant no. 2018M632287).

## References

Alp, Z. Z., & Oguducu, S. G. (2018). Identifying topical influencers on twitter based on user behavior and network topology. *Knowledge-Based Systems, 141*, 211-221.

An, L., Yi, X., Yu, C. and Li, G. (2017, October). Predicting the influence of microblog entries regarding public health emergencies. In *ISSI 2017 -Proceedings of the 16th International Conference on Scientometrics & Informetrics, Wuhan, China*.

An, L., & Wu, L. (2017). An integrated analysis of topical and emotional evolution of microblog public opinions on public emergencies. *Library and Information Service, 61*(15), 120-129.

Asur, S., Huberman, B. A., Szabo, G., & Wang, C. (2011, July). Trends in social media: persistence and decay. Paper presented at *ICWSM*, Barcelona, Spain.

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011, February). Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining* (pp. 65-74). ACM.

Bandari, R., Asur, S., & Huberman, B. A. (2012, June). The Pulse of News in Social Media: Forecasting Popularity. *In Tomkins, A.(Chair), ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media,* Dublin, Ireland.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(1), 993-1022.

Bora, S., Singh, H., Sen, A., Bagchi, A., & Singla, P. (2015). On the role of conductance, geography and topology in predicting hashtag virality. *Social Network Analysis and Mining, 5*(1), 57-71.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Cao, W., Huang, Y., & Tu, G. (2016). The research of transmission characteristics of the micro-blog topic based on time network influence model. *Library and Information Service, 60*(1), 91-97.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P. K. (2010). Measuring user influence in twitter: The million follower fallacy. International Conference on Weblogs and Social Media (*ICWSM)*, May 23-26, 2010, Washington, DC.

Chen, J., Liu, Y., & Zou, M. (2017). User emotion for modeling retweeting behaviors. *Neural Networks, 96*, 11-21.

Chen, L. (2014). The evaluation model research on information dissemination influence of micro-blog individual. *Data Analysis and Knowledge Discovery,30*(2), 79-85.

Chen, Y. Z., Fang, M. Y., Guo, W. Z., & Guo, K. (2015). Topic popularity prediction of microblog based on wavelet transformation and ARIMA. *Pattern Recognition and Artificial Intelligence, 28*(7), 586-594.

Finch, K. C., Snook, K. R., Duke, C. H., Fu, K. W., Tse, Z. T. H., Adhikari, A., & Fung, I. C. H. (2016). Public health implications of social media use during natural disasters, environmental disasters, and other environmental concerns. *Natural Hazards*, 83(1), 729-760.

Fu, Y., & Chen, Y. (2014). Relationship analysis of microblogging user with link prediction. *Computer Science, 41*(2), 201-205.

Gao, H., Wang, S. S., & Fu, Y. (2011). Prediction model for long-term development trend of web sentiment. *Journal of University of Electronic Science and Technology of China, 40*(3), 440-445.

Gao, S., Ma, J., & Chen, Z. (2014). Popularity prediction in microblogging network. In *Asia-Pacific Web Conference* (pp. 379-390). Springer, Cham.

Gromov, V. A., & Konev, A. S. (2017). Precocious identification of popular topics on Twitter with the employment of predictive clustering. *Neural Computing & Applications, 28*(11), 3317-3322.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Singapore: Morgan Kaufmann.

Hong, L., Dan, O., & Davison, B. D. (2011). Predicting popular messages in twitter. In *Proceedings of the 20th International Conference Companion on World Wide Web* (pp. 57-58). ACM

Hong, R., He, C., Ge, Y., Wang, M., & Wu, X. (2017). User vitality ranking and prediction in social networking services: A dynamic network perspective. *IEEE Transactions on Knowledge and Data Engineering, 29*(6), 1343-1356.

Ikeda, K., & Kurihara, S. (2017). An examination of a novel information diffusion model for social media. In K. Endo, S. Kurihara, T. Kamihigashi, & F. Toriumi (Eds.), *Reconstruction of the Public Sphere in the Socially Mediated Age* 93-117 Singapore: Springer.

Jia, Y., An, L., & Li, G. (2015). On the online information dissemination pattern of city emergencies. *Journal of Intelligence, 34*(4), 91-96.

Kim, E., Hou, J., Han, J. Y., & Himelboim, I. (2016). Predicting retweeting behavior on breast cancer social networks: Network and content characteristics. *Journal of Health Communication, 21*(4), 479-486.

Kim, Y. K., Lee, D., Lee, J., Lee, J. H., & Straub, D. W. (2018). Influential users in social network services: The contingent value of connecting user status and brokerage. *The Data Base for Advances in Information Systems, 49*(1), 13–32.

Kong, Q., Mao, W., & Liu, C. (2016, August). Popularity prediction based on interactions of online contents. *In 2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS),* pp. 1-5. IEEE*, Beijing, China.*

Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web* (pp. 591-600). ACM, *North Carolina, USA.*

Lemahieu, R., Van Canneyt, S., De Boom, C., & Dhoedt, B. (2015, November). Optimizing the popularity of Twitter messages through user categories. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)* (pp. 1396-1401). IEEE*, Atlantic City, NJ, USA.*

Li, Y. (2013). *Research on the Related Issues of Short Text Topical Analysis*. Beijing: Beijing University of Posts and Communications.

Li, Y. L., Yu, H. T., & Liu, L. X. (2013). Predict algorithm of micro-blog retweet scale based on SVM. *Jisuanji Yingyong Yanjiu, 30*(9), 2594-2597.

Liao, Q., Wang, W., Han, Y., & Zhang, Q. (2013, December). Analyzing the influential people in Sina Weibo dataset. In *Global Communications Conference (GLOBECOM)*, 2013 *IEEE* (pp. 3066-3071). IEEE, *Atlanta, GA, USA.*

Liu, Y., Ding, Y., Hao, K., & Huang, B. (2016). User characteristics based information diffusion model for analysis of hot social events. *In 2016 12th World Congress on Intelligent Control and Automation (WCICA),* pp. 2131-2136. IEEE.

Liu, Y., Wang, B., Wu, B., Shang, S., Zhang, Y., & Shi, C. (2016). Characterizing super-spreading in microblog: An epidemic-based information propagation model. *Physica A, 463*, 202-218.

Luo, X. (2013). A communication-influence-power evaluation index system for major emergency events on microblogs. *Journal of Communication, 3*, 76-82.

Luo, Z., Chen, T., & Cai, W. D. (2014). Microblogging retweet prediction algorithm based on random forest. *Computer Science, 41*(4), 62-64.

Maleewong, K. (2016). An analysis of influential users for predicting the popularity of news tweets. In *Pacific Rim International Conference on Artificial Intelligence*, 22 th-23th August, Phuket, Thailand (pp. 306-318). Springer, Cham.

Milinovich, G. J., Williams, G. M., Clements, A. C. A., & Hu, W. (2014). Internet-based surveillance systems for monitoring emerging infectious diseases. *The Lancet. Infectious Diseases, 14*(2), 160-168.

Nargundkar, A., & Rao, Y. S. (2016, April). InfluenceRank: A machine learning approach to measure influence of Twitter users. In *2016 International Conference on Recent Trends in Information Technology (ICRTIT)* (pp. 1-6). IEEE, *Chennai, India*.

Palovics, R., Daroczy, B., & Benczur, A. A. (2013). Temporal prediction of retweet count. *2013 IEEE 4th International Conference on* Cognitive Infocommunications (CogInfoCom), 9th-13thJune, Budapest, Hungary (pp. 267-270).

Pei, J., Yu, G., Tian, X., & Donnelley, M. R. (2017). A new method for early detection of mass concern about public health issues. *Journal of Risk Research, 20*(4), 516-532.

Peng, H. K., Zhu, J., Piao, D., Yan, R., & Zhang, Y. (2011). Retweet modeling using conditional random fields. *2011 11th IEEE International Conference on Data Mining Workshops,7th-10th June,* Beijing, China (pp. 336-343).

Petrovic, S., Osborne, M., & Lavrenko, V. (2011). Rt to win! Predicting message propagation in twitter. *In Aral, S.(Chair), ICWSM 2011 - Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, pp.*586-589.

Riquelme, F., & Gonzalez-Cantergiani, P. (2016). Measuring user influence on Twitter: A survey. *Information Processing & Management, 52*(5), 949-975.

Sina Technology. (2016), 2016 Sina Weibo first quarter earnings: Monthly active users increase to 261 million, Retrieved from http://www.askci.com/news/chanye/20160513/1451556198.shtml

Starbird, K., & Palen, L. (2011, May). Voluntweeters: Self-organizing by digital volunteers in times of crisis. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1071-1080). ACM, *Vancouver, British Columbia*.

Sun, H., & Li, L. (2012). The Characteristics of High-frequency Forwarding Weibo and the Analysis of User's Turning Engine – -Based on the Content Analysis of Sina Weibo's "Day Forwarding Leaderboard". *Modern Communication*, 34(6), 137-138.

Towers, S., Afzal, S., Bernal, G., Bliss, N., Brown, S., Espinoza, B., Castillo-Chavez, C. (2015). Mass media and the contagion of fear: The case of Ebola in America. *PLoS One, 10*(6), e0129179.

Varshney, D., Kumar, S., & Gupta, V. (2017). Predicting information diffusion probabilities in social networks: A Bayesian networks based approach. *Knowledge-Based Systems, 133*, 66-76.

Wang, G., Zheng, Q., Wang, Y., Xiong, W., & Xie, H. (2014). Research on the characteristics and users' retweeting rules of top trending micro-blogs on sina. *Journal of Intelligence, 33*(4), 117-121.

Weng, J., Lim, E. P., Jiang, J., & He, Q. (2010, February). Twitterrank: Finding topic-sensitive influential twitterers. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining* (pp. 261-270). ACM, *New York City, USA*.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington, Massachusetts: Morgan Kaufmann.

Xiao, L.,& Qi, J. (2013). On the Evaluation System of the Social Influence of Enterprise Public Opinion on Internet Based on Microblog. *Journal of Intelligence*, 32(5), 5-9.

Yamaguchi, Y., Takahashi, T., Amagasa, T., & Kitagawa, H. (2010, December). TURank: Twitter user ranking based on user-tweet graph analysis. In *International Conference on Web Information Systems Engineering* (pp. 240-253). *Hong Kong, China*

Yang, Z., Guo, J., Cai, K., Tang, J., Li, J., Zhang, L., & Su, Z. (2010, October). Understanding retweeting behaviors in social networks. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (pp. 1633-1636), Toronto, Canada.

Ye, Li, Yang, Lee, Wu, (2018). The fear of Ebola: A tale of two cities in China. In Zhen, J., Shen, M., Li (Eds.), *Big Data Support of Urban Planning and Management* (pp. 113-132). Cham: Springer.

Ye, S., & Wu, S. F. (2010). Measuring message propagation and social influence on Twitter.com. In *International Conference on Social Informatics* (pp. 216-231). Springer, Berlin, Heidelberg.

Ye, T. (2012), *Trends Analysis and Prediction of Micro-Blogging Platforms*, Wuhan: Doctoral Dissertation of Wuhan University.

Zaman, T. R., Herbrich, R., Van Gael, J., & Stern, D. (2010, December). Predicting information spreading in twitter. Paper presented at *Computational Social Science and the Wisdom of Crowds (NIPS 2010 pp.599-601*), Whistler, Canada.

Zhai, X., Liu, Q., Cheng, Y., Hu, Q., & Li, H. (2015). Research on hotness prediction in Sina microblog based on forward level analysis. *Computer Engineering, 41*(7), 31-35.

Zhang, Y., Lu, R., & Yang, Q. (2012). Predicting retweeting in microblogs. *Journal of Chinese Information Processing, 26*(4), 109-114.

Zhao, R., & Zeng, X. (2014). Analysis of the influence factors of microblog's information dissemination. *Information Studies: Theory & Application, 37*(3), 58-63.