

# Markov Chain Model of Rumor Situation Prediction

Hongfang Man<sup>1, a</sup>, Pu Li<sup>2, b</sup>

<sup>1</sup>Library, Liaocheng University, Liaocheng, 252000, China

<sup>2</sup>School of Management Science and Engineering, Shandong Normal University, Jinan 250014, China

<sup>a</sup>manhongfang@lcu.edu.cn, <sup>b</sup>23176461@qq.com

**Keywords:** Network rumors, Microblog rumors, rumors situation prediction, rumor spreading

**Abstract:** Based on the study of the concept, characteristics, current situation and influencing factors of Microblog rumors, this paper observes and values the number of Microblog's posts, comments, and relays related to rumors, and establishes the prediction model of network rumor situation based on Markov chain, which can make a better prediction of the rumor spreading situation, so as to provide reasonable and effective suggestions for the governance of rumors.

## 1. Introduction

As a new type of broadcast social networking platform capable of real-time information sharing, publishing, and exchange, Microblog provides users with an environment to express their own ideas at any time and place, and its timeliness and arbitrariness make Microblog increasingly hot, and is gradually forming Microblog effect among Internet users. Microblog's diversity of information, freedom of expression, fissile dissemination, high user frequency, and high popularity have provided an excellent platform for the rumor spreading, which has greatly promoted the dissemination and spread of online rumors [1]. How to manage Microblog's rumors and create a healthy Internet environment is particularly important.

In order to better control and govern Microblog rumors, on May 12, 2016, Sina Microblog and the Ministry of Public Security launched the first "national platform for rumor removal". Users can upload the false information on any platform on the platform. The public security department and the network police will check the false information uploaded by the user and publish the rumors. However, due to the different criteria for judging rumors, the verification process is more difficult and often relies entirely on manpower, which is time-consuming and laborious [2].

Based on the Markov chain, this paper observes the network rumors with Microblog as the medium of communication, and designs situation prediction model for the development of rumors. Through the analysis of netizens' attention on rumors, predict the situation of rumors spread in the next period of time in order to take further appropriate measures to control Microblog rumor and provide a scientific and effective basis for rumors governance.

## 2. Related Works

With the increasing harm and influence of internet rumors on society, scholars at home and abroad have begun to study internet rumors from different perspectives [3], especially in the identification and dissemination analysis of Microblog rumors.

The earliest and most famous social networking site in foreign countries is Twitter in the United States. Mendoza [4] observed the user behavior data on Twitter in the Chilean earthquake in 2010 and found that the spread of rumors is not the same as the dissemination of news, the former is more likely to be questioned by social networks. Castillo [5] used the decision tree to classify microblog information as trustworthy and untrustworthy, and the study was based on the credibility of microblog to identify rumor information. Takahashi et al. [6] ranked the credibility of each piece of Microblog information, and then selected the micro-blog which has the most frequency to be spread as a rumor, in order to further determine whether it is a rumor. Morris et al. [7] found that the

information receiver was often difficult to identify the information based on the content of the rumor. But according to the nature of the publisher and the spread nature of the information release network can promote information receivers to judge the authenticity of information. Yibo Xue et al.[8] added rumors positive infection and negative infection in the current widespread use of rumor spreading model SIR, designed a SPNR model which was more suitable for describing rumors, and designed the SPNR algorithm to describe the mode of rumor spreading more accurately . Yan Xiong [9] started with the content of rumors and the activity of rumors, studied the causal relationship between the two, and found out the influence factors of rumors' content characteristics on the rumors' activity. Jianmin He et al. [10] used hidden Markov theory to design the Microblog public opinion evolution model and demonstrated the evolution rule of Microblog public opinion. Jiabin Liu [11] mainly studied the generation, dissemination and governance of Internet rumors in the perspective of communication, and put forward the control and governance opinions on Microblog rumor, which provided a certain theoretical value and reference significance for the management of Internet rumors.

### 3. Markov Prediction Model

#### 3.1 Situation prediction model

The Internet rumor situation prediction model based on Markov chain is a three tuple  $\lambda$ ,  $\lambda = (S, A, \pi)$ :

$S$  is the set of states,  $S = \{S_1, S_2, \dots, S_n\}$ ;

$A$  is the probability transfer matrix of the state,  $A = (a_{ij})$ ,  $a_{ij} = P(X(n+1) = s_j | X(n) = s_i)$ ,  $1 < i, j < n$ ;

$\pi$  Is the initial probability distribution,  $\pi = \{\pi_i\}$ ,  $\pi_i$  is the probability that the initial time is  $s_i$ ,  $\pi_i = P(X_1 = s_i)$ .

#### 3.2 Count the observed variable values

The trend of rumor spread is described by the increasing degree of netizens' attention to rumors. There are four development trends, namely, sharp rise, sharp decline, slow rise and slow decline. The degree of attention of the Microblog rumors is obtained by counting the number of Microblog's posts, comments, and relays related to rumors. During the spread of rumors, there will be many comments and relays after publishing each Microblog rumor, and as time goes by, the number of microblogs related to the rumor and the number of comments and relays under the Microblog are increasing. In the four stages of the formation, outbreak, fluctuation, and disappearance of rumor spread, the total attention value of rumors gradually increases with the increase of time, and when it reaches the stage of disappearance, the total attention value of rumors is the greatest.

Assume that the number of posts  $M$ , the number of relays  $R$ , and the number of comments  $U$  of the relevant Microblog are fixed data at the time of data acquisition. Therefore, through the web crawler software crawling data and obtain the number of Microblog's posts, comments, and relays related to rumors within a certain period of time.

The formula for calculating the attention degree the rumor is as follows:

$$X_t = \sum_{i=1}^{m(t)} w_1 * m + w_2 * U + w_3 * R$$

Among them,  $w_1, w_2, w_3$  is the weight, which is determined by the combination of subjective assignment and objective assignment.

#### 3.3 Determine the state space

Definition The rumor attention value is recorded as  $X$ ,  $X = [X_1, X_2, \dots, X_t]$ , the increase degree

of the rumor attention is recorded as:  $\Delta X$ ,  $\Delta X = [\Delta X_1, X_2, \dots, X_t]$ , and  $\Delta X_i = X_{i+1} - X_i$ ,  $\Delta X_i (i = 1, 2, \dots, t)$  indicates increase degree of the rumor attention at time  $i$ .

The state space  $S$  is classified into four states according to the increase degree of the rumor attention: slow increase  $s_1$ , sharp increase  $s_2$ , sharp decline  $s_3$ , and slow decline  $s_4$ .  $S = \{s_1, s_2, s_3, s_4\}$ , among them,

$$\begin{aligned} s_1 &= (0, \Delta \|X\|_{\max} / 4) \\ s_2 &= (\Delta \|X\|_{\max} / 4, \Delta \|X\|_{\max}) \\ s_3 &= (-\Delta \|X\|_{\min}, -\Delta \|X\|_{\min} / 4) \\ s_4 &= (-\Delta \|X\|_{\min} / 4, 0) \end{aligned}$$

### 3.4 Determine model parameters

The key to solving the model  $\lambda$  is to solve the state space and the state transfer matrix. In this paper, we select several groups' data which is defined as rumors as the training samples, select the hourly interval as the minimum observation statistical interval, and organize the rumor data of the training samples, and make the weighted average. Finally, the best state space and state transfer matrix for rumor spreading is as follows:

$$\begin{aligned} s_1 &= (0, 858.2083), \quad s_2 = (858.2083, 3432833), \\ A &= \begin{bmatrix} 0.368421 & 0.1052632 & 0.157895 & 0.210526 \\ 0 & 0.25 & 0.75 & 0 \\ 0.666667 & 0.0833333 & 0.25 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \\ s_3 &= (-101.5, -25.375), \quad s_4 = (-25.375, 0) \end{aligned}$$

## 4. Experimental simulation

### 4.1 Data collection and processing

This paper selects Sina Microblog as the experimental sample data collection source and research platform. Select March 13, 2013 to March 15, 2013 as the observation time, and select the "be dead" rumor such as "Liu Xiaolingtong was dead" as the research object. Related the information data about Microblog can be determined and can be used as experimental samples. Taking into account the rapid spread of microblog rumors, in order to accurately record the state of rumor spreading, this paper selects the hourly interval as the minimum statistical interval, and records the change of the observations in hourly.

Set the data collection keywords as "Liu Xiaolingtong", "death", collect the required data through the web crawler software, and import the obtained data records into the corresponding attribute items of the Excel table, the fields include: user name, release Time, the number of comments, the number of replies, Microblog content, etc.

### 4.2 Calculation results

According to the model, the state transition matrix, the state space, the weight  $\omega$  of the model can be determined and the weights  $\omega$  can be calculated by a combination of subjective assignment and objective assignment. The results are shown in Table 1. To avoid the influence of errors on the calculation results, this paper screens the data with excessive deviations.

Table 1. List of rumor attention and the increasing degree of rumor attention

Number	The number of posts	The number of comments	The number of relays	The degree of the rumor attention	The increasing degree of rumor attention
1	39	16	5	208	151
2	35	183	11	550	342
3	37	26	7	228	-322
4	59	44	11	368	140
5	55	200	12	668	300
6	49	17	7	258	-410
7	23	8	4	124	-134
8	8	0	3	44	-80
9	2	1	1	14	-30
10	1	0	0	4	-10
Number	The number of posts	The number of comments	The number of relays	The degree of the rumor attention	The increasing degree of rumor attention
11	3	1	0	14	10
12	4	7	1	34	20
13	7	3	3	46	12
14	24	7	2	118	72
15	29	22	0	160	42
16	18	2	2	84	-76
17	23	23	3	150	66
18	16	3	1	74	-76
19	15	4	0	68	-6
20	7	3	2	42	-26
...	...	...	...	...	...

Among the selected 48 first time numbers, there are 22 in the state  $s_1$ , 0 in state  $s_2$ , 12 in state  $s_3$ , and 14 in state  $s_4$ . The State probability distribution is:

$$\pi = (0.458333, 0, 0.25, 0.291667)$$

Substituting this initial probability distribution into the rumor spreading situation prediction model  $\lambda$ , we get the state matrix for predicting the degree of rumor attention in the 49th time number and that is:

$$\pi' = (0.627193, 0.069079, 0.134868, 0.096491)$$

From the predicted matrix, it can be seen that in the 49th time number, the degree of the rumor attention has a high probability to be in state of  $s_1$ . That is, in the 49th time number, the degree of rumor attention will increase slowly.

In order to the accuracy of the prediction, the data that is collected in the 49th time number is listed in Table 2:

Table 2. Data for the 49th time number

Number	The number of posts	The number of comments	The number of relays	The degree of the rumor attention	The increasing degree of rumor attention
49	3	3	0	18	14

The degree of the rumor attention is 14, which is in the state space  $s_1 = (0, 858.2083)$  defined by the model, and proves that in the 49th time number, the degree of the rumor attention will increase slowly.

The results of real data are consistent with the prediction results of the prediction model, which verifies the correctness and applicability of the model in predicting the spreading situation of Microblog rumors.

The total attention curve of the rumors collected during the crawling period is shown in Figure 1:

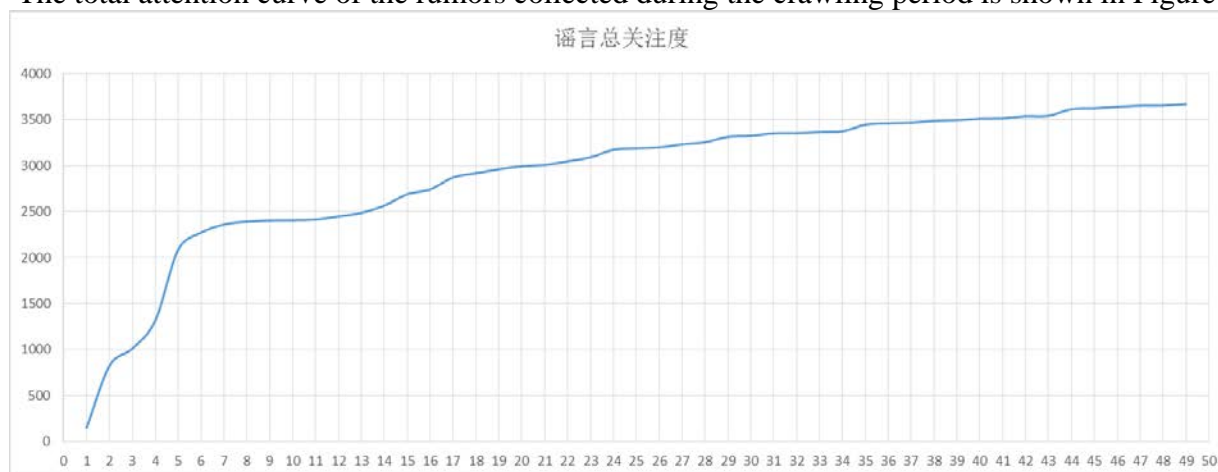


Figure 1. The trend of the total attention of the rumors during the crawling period

According to the model predictions, the degree of the rumor attention in the 49th time number will increase slowly in the next time period, that is, the increase is smaller. This proves that the spread of rumors is in a phase of volatility, and netizens have a certain degree of cognition and understanding of the rumor and hold a sustained attitude. At this time, the rumor management department should use its own resources to encourage social forces to actively participate in refuting the rumor, open up a diversified channel to refute the rumor, ensure the circulation of official information, and effectively monitor and control negative information, and effectively avoid the rise of the rumor heat.

## 5. Conclusion

### 5.1 The prevention measures of the Microblog rumors

In the formation period of rumors, fundamentally purifying the original source is an effective way to prevent the spread of Microblog rumors. Specific measures can be implemented from the following aspects;

Strengthen the supervision of news media networks. Microblog rumors are mainly disseminated and disseminated through online media. Therefore, it is imperative to strengthen the supervision of online media. The legislature, the judiciary, and the supervisory authorities should quickly formulate and improve relevant network management laws and regulations, and allow the network supervision to be in accordance with the law as soon as possible.

Strengthen supervision of Internet V. Implementing a real-name system for registering the network as soon as possible, so that the Internet V will have some concerns in publishing, or replying or commenting on information. It will also be able to take into consideration the legal responsibilities that will be borne when publishing and replying Microblog rumors. This will play an important role in restraining the spread of Microblog rumors. The law enforcement department also needs to focus on educating, monitoring and controlling the Internet V who disseminates rumors, and supervises the extremely active Internet V, guides the Internet V to output positive energy information, and jointly resists the diffusion and propagation of Microblog rumors.

### 5.2 The governance measures of the Microblog rumors

The life cycle of rumors can usually be divided into four stages: formation period, outbreak period, fluctuation period, and extinction period. When network rumors are in the outbreak period and the fluctuation period, it is a key stage for taking measures to manage. According to the heat and trend of

rumor, the management department should predict the next stage of the rumor spreading and take appropriate measures to deal with it.

When it is predicted that the heat of rumors has risen sharply, management departments at all levels should respond immediately and verify the authenticity of rumors as soon as possible. The false information that has already been verified should be clearly identified as “rumors” in the information and restrict its dissemination and visibility. The users who spread rumors should be actively informed that the information being disseminated is rumors, be asked to delete and discourage them continue to disseminate. If the user still insists on broadcasting, the user should be punished by legal means. If it is still not possible to verify whether the false information is a rumor, a “warning” approach should be used to remind netizens that the authenticity of the information is not yet clear, and do not trust and spread it completely, so as to avoid further deterioration of the situation.

When it is predicted that the popularity of rumors had dropped sharply, it is proved that netizens have already had a clearer and more thorough understanding of the rumors. The rumors management department can focus on the “remedial work” of rumors management, and give a certain degree of commendation to the netizens who report rumors, which can guide netizens to actively participate in the process of governing rumors. At the same time, we must also do a good job in recording the evolution of the rumor and information content in order to prevent such rumors from happening again and eliminate potential hazards.

When it is predicted that the heat of rumors is rising slowly or decreasing slowly, it is proved that rumors spread is in a stage of fluctuation. Netizens have already had a certain level of awareness and understanding of rumors and maintained a wait-and-see attitude. Rumors management department should use their own resources to encourage social forces to participate in refuting the rumor and open up diversified channels, ensure the circulation of official information, and monitor and control effectively negative information to avoid the rise of the heat of rumor at the same time.

## References

- [1] China industry survey network. Analysis and interpretation of the 2016 microblog user development report [EB/OL].
- [2] Gang Him, Xueqiang Lv, Zhuo Li, and Liping Xu. Automatic Rumor Identification on Microblog [J]. Library and Information Service, 2013, (23): 114 - 120.
- [3] Xiangdong He, Yining Zhu. Internet Rumors Identification Methods and It's Future [J]. Cyberspace Security, 2016, (Z2): 9 - 12.
- [4] Mendoza M, Poblete B, and Castillo C. Twitter under Crisis: Can we trust what we RT? [C]// Social Media Analytics, SOMA, KDD workshop. 2010: 71 - 79.
- [5] Castillo C, Mendoza M, Poblete B, Information credibility on twitter [C]//Proceedings of the 20th international conference on Worldwide web. ACM, 2011: 675 - 684.
- [6] Takahashi T, Igata N. Rumor detection on twitter [C]// Joint, International Conference on Soft Computing and Intelligent Systems.2012: 452 - 457.
- [7] Morris M R, Counts S, Roseway A, et al. Tweeting is believing?:understanding microblog credibility perceptions [C]// Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. ACM, 2012: 441 - 450.
- [8] Yibo Xue, Yuanyuan Xue, Chengqi Yi. SPNR: A Novel Rumor Propagation Model on Social Networks [J]. Netinfo Security, 2014, (01): 5 - 9.
- [9] Yan Xiong. A new method of rumor research: activity factor prediction [J]. Modern Communication (Journal of Communication University of China), 2013, (03): 70 - 76.
- [10] Jainmin He, Xue Li. A Hidden Markov Model Research in the Microblog Public Opinion

Evolutionary Analysis [J]. Information Science, 2016, (04): 7 - 12.

[11] Jiaxin Liu. The Spreading Mechanism and Control of Network Rumor [D]. Northeast Normal University, 2014.

[12] Changhe Fu, Su Deng. Application of Markov Chain in Infectious Disease Forecasting [J]. Journal of Shenyang Normal University (Natural Science Edition), 2009, (01): 28 - 30.

[13] Gang Zhu. A study on the screening mechanism of Internet rumors[C]// Forum on decision making: Symposium on regional development and public policy research. 2016: 3.

[14] Yu Hu. Mass communication effect [M]. Xinhua Publishing, 2000.

[15] Xiaoyong Tang. Algorithm Research for Markov Chain [D]. University of Electronic Science and Technology of China, 2015.

[16] Chenggang He. The Research of Markov Model Forecasting Method and Application [D]. Anhui University, 2011.

[17] Xiujuan Gu, Chao Li. Research on House Price Prediction Based on Markov Chain [J]. Consumer Economics, 2012, (05): 40 - 42+48.