

The Algorithm Study about the Information Dissemination Probability in Micro-blog Community

Bo Wang, Yuwang Zhang, Li Shao

School of Humanities and Social Sciences

Harbin Institute of Technology, Harbin, China

wangbo@hit.edu.cn, Zhangyw2122@163.com, shaolihit@163.com

Keywords: micro-blog community; information dissemination probability; Initial dissemination probability; specific time dissemination probability

Abstract. According to the reality of information dissemination in micro-blog community, the information dissemination probability is divided into two part: initial information dissemination probability and specific time information dissemination probability, As consideration of their impacts, algorithm of information dissemination probability is built by algorithm of initial information dissemination probability and algorithm of information dissemination probability in specific time. With the algorithm of information dissemination probability, we can provide a set of theoretical foundation and practical instruments to further study the process of information dissemination in micro-blog community.

Introduction

In recent years, the micro-blog community grew rapidly as a convenient network of social media and gets the social acceptance and recognition gradually. The research of information dissemination probability in micro-blog community has great significance to depict information dissemination process and information transmission mechanism. Many scholars introduce the epidemic model on the study of information dissemination process in micro-blog community, because there are many similarities between the dissemination and spread of infection. Users are abstracted as network node by and the most in-depth one is SIR model [1]. Xu Xiaodong etc. [2] establish a SIR model of micro-blog community by abstracting users as network nodes and study the dissemination mechanism of rumor. Yang Gang etc. [3] based on an infectious disease model; information transmission probability is redefined by taking into consideration of the networks weights. Zhang Yanchao etc. [4] discover that the greater degree the initial node of information transmission presents, the easier information spreads in network by means of simulating the process of information transmission in online social network. On the basis of elements called infection delay and non-uniform transmission, Zhao Jing etc. [5] bring forward a novel SIR model, by means of which the fact that the threshold value of propagation can be greatly lowered by infection delay and that critical value of propagation can be increased by transmission in the heterogeneous networks is found. Wang Hongliang etc. [6] built a SIR model about the dissemination of micro-blog information and analysis it with pathogen, infectivity, and immunity features that comes from epidemiology. YANG Zilong etc. [7] made some improvements on SIR model based on the law of diminishing reposting probability and proved the validity of the model, found that this model can better fit the process of information propagation when considering the time-effectiveness and information obsolescence of the message. From the current research achievement, many documents involved the study of the information dissemination process of the infectious disease model, but the research about structure of the specific algorithm for the micro-blog community information dissemination probability in is still lack. To study the process of information dissemination probability in micro-blog community in-depth, we have to clear obstacles above. Therefore, this article will based on the SIR model and combine the actual situation of micro-blog community constructs the algorithm of micro-blog community information dissemination probability and provides the theory basis and calculation tools for more in-depth research of tweet dissemination

process.

The establishment of information dissemination probability

In the research of SIR model about information dissemination process in micro-blog community, information dissemination probability is initially valuable to the depiction of information dissemination process. The information dissemination probability means the evaluation of effect in information dissemination process. Normally, the higher the information dissemination probability is, the better effect of information dissemination in micro-blog community is. On the contrary, the worse effect of information dissemination in micro-blog community is.

From the reality of information dissemination process, we can divide specific dissemination probability into initial dissemination probability and specific time dissemination probability. If the dissemination was represented by λ , the initial dissemination probability was represented by λ_0 and the specific time dissemination was represented by $\lambda_{kk'}$. Then the dissemination probability could be expressed by the following function:

$$\lambda = \lambda_0 + \lambda_{kk'} \quad (1)$$

Initial dissemination probability

The initial dissemination probability of means a dissemination probability when it is new published in micro-blog community. It influenced by information content itself and social authority of publisher. On the one hand, if content has high quality or easy to catch the eye of others, it has a high probability be accepted and continue to spread. We use p to denote the probability of information itself can be accepted, and then we have:

$$p = p(k_0) \quad (2)$$

In which, the range of value of $p(k_0)$ is .

When the publishers have high impact on society, they have higher social authority and always have more followers. When their information is released, there will be a higher level of initial dissemination probability. If we use θ to denote publisher's social authority, there is the function:

$$\theta = \frac{f_{i0}}{\sum f_i} \rho \quad (3)$$

f_{i0} is the quantity of fans of this publisher. $\sum f_i$ Denotes how many fans all of the publishers have in micro-blog community. ρ is the influence degree among the publishers of his fans. The range of value of ρ is .

So the initial dissemination probability can be expressed as the following form:

$$\lambda_0 = p\theta = p(k_0) \frac{f_{i0}}{\sum f_i} \rho \quad (4)$$

So far, initial dissemination probability formula is constructed.

Dissemination probability in specific time

In general, the information of the micro-blog community can be scanned or spread by fans. When we observe a specific time, we will find this node requires a certain number of forwarding, comments and praises. Generally, the forwarding, comments and praises of a node have indirect influenced on information dissemination. It shows the probability of information can be spread is increase as the quantity of forwarding, comments and praises increase. So if we use $\lambda_{kk'}$ to denote information dissemination probability in the specific time and f_{ik} to denote the number of followers of the information publisher or forwarder, α_k , β_k , γ_k representing information forwarding quantity, comment quantity and praise quantity in specific time. The dissemination probability in specific time can be expressed as

$$\lambda_{kk'} = \frac{\alpha_k + \beta_k + \gamma_k}{f_{ik}} \quad (5)$$

It is clearly we should consider: the three behavioral pattern forwarding, comments and praises have different indirect influence degree to information. Forwarding node can forward original one to forwarder's own interpersonal circle and proceed with the second dissemination or many times dissemination. It can enlarge dissemination range to a great degree and it has the best communication effects. Comment can show commentator's opinions of publishers' or forwarders interpersonal circle. To some extent, comment can attract information publishers or forwarder's fans to browse or spread the information. It also has strong influence on information dissemination but lower than forwarding. Praise can only explain the level of the information be concerned. Thus praise has the lowest influence in the three. So we should give weight to the degree of comment influence, forwarding and praise. If we use x, y and z to denote the weight of forwarding comment and praise .dissemination probability in specific time can be improved as:

$$\lambda_{kk'} = \frac{\alpha_k x + \beta_k y + \gamma_k z}{f_{ik}} \quad (6)$$

Because of forwarding have higher influence degree than comment, comment have higher influence than praise. On the numerical from large to small order is x, y, z. But the value of x, y, z should be calculated by data.

According to reports, although there is a dramatic increase in micro-blog users, most of the users is not very active. After considering this factor, we find that the information dissemination probability in specific time is significantly undervalued. So we introduce micro blogging users' activity function here. We use φ to denote user activity function, f'_{ik} to denote active fans in specific time node .Then we have the following functional form:

$$\varphi = \frac{\sum f'_{ik}}{\sum f_{ik}} \quad (7)$$

After introduce the function of user activity, we can improve information dissemination probability in specific time as:

$$\lambda_{kk'} = \frac{\alpha_k x + \beta_k y + \gamma_k z}{\varphi f_{ik}} \quad (8)$$

Usually, the degrees of information in specific time whether continue to spread is related to the degree of closeness between this node and its fans, the impact of this node in its fans and their social positions. This factor is called as credibility and introduce credibility parameters $\xi_{kk'}$. In general, information dissemination probability will increase with the increase of credibility. On the contrary, it reduced. Therefore, information dissemination probability in specific time can be optimized as the following form:

$$\lambda_{kk'} = \frac{\alpha_k x + \beta_k y + \gamma_k z}{\varphi f_{ik}} \xi_{kk'} \quad (9)$$

Whether micro-blog users through the real-name authentication are an important factor influence on how well the information is accepted or it is accepted. When the other conditions is certain, real-name registration users have greater impact on dissemination probability than normal users. We introduce dummy variable D1 here. When there is a real-name authentication user in a specific time node, D1=1. When there is a normal user in a specific time node, D1=0. Under the assumption of the real-name authentication user has the degree of v affection while the normal user not, the scope of v equals . Then we can express information dissemination probability in another way:

$$\lambda_{kk'} = \frac{\alpha_k x + \beta_k y + \gamma_k z}{\varphi f_{ik}} \xi_{kk'} (1 + D_1 v)$$

$$D_1 = \begin{cases} 1, & \text{real-name users} \\ 0, & \text{normal users} \end{cases} \quad (10)$$

Base on previous study about the information transfer process, in the life cycle of information dissemination, information transmission probability will rise in the first stage by a changing process. So if we want to describe information transmission more realistic, we should take time variation into consideration. Here we introduce time-shifted function $\varphi(t)$, Ψ is the maximum value of information dissemination probability in its lifecycle, t is the length of time information spread in the specific time node, π represents the length of time during which the dissemination probability of information reaches threshold values of its life cycle. T represents the market of the entire life cycle of information transmission. η is the parameter and it is also an even number. So the time-shifted function can be expressed as:

$$\varphi(t) = \Psi - \left(\frac{t - \pi}{T} \right)^\eta \quad (11)$$

After set the time-shifted function, we can improve information dissemination probability into the form as follows:

$$\lambda_{kk'} = \frac{\alpha_k x + \beta_k y + \gamma_k z}{\varphi f_{ik}} \xi_{kk'} (1 + D_1 \nu) \phi(t) \quad (12)$$

$$D_1 = \begin{cases} 1, & \text{real-name users} \\ 0, & \text{normal users} \end{cases}$$

In addition to above factors, other factors can also influence information dissemination probability. For example, whether the content will be referred to accidents, or how many times acceptors have received the information. In comparison, we won't go to details because the probability of those factors is smaller. However, we can't ignore these factors' impact on information dissemination probability, so we can include a stray parameter from those factors. Finally, we can get a function about specific time node information dissemination probability:

$$\lambda_{kk'} = \frac{\alpha_k x + \beta_k y + \gamma_k z}{\varphi f_{ik}} \xi_{kk'} (1 + D_1 \nu) \phi(t) \mu \quad (13)$$

Therefore, we just finish the dissemination probability construction of formula of in specific time node.

The information dissemination probability is composed of initial information dissemination probability and dissemination probability of specific time node. By the algorithm of this two parts detail, the function form of the information dissemination probability could be included:

$$\lambda = \lambda_0 + \lambda_{kk'} = P(k_0) \frac{f_{i0}}{\sum_i f_i} \rho + \frac{\alpha_k x + \beta_k y + \gamma_k z}{\frac{\sum_{ik} f_{ik}}{\sum_{ik} f_{ik}}} \xi_{kk'} (1 + D_1 \nu) \left[\Psi - \left(\frac{t - \pi}{T} \right)^\eta \right] \mu \quad D_1 = \begin{cases} 1, & \text{real-name users} \\ 0, & \text{normal users} \end{cases} \quad (14)$$

So we just finish the construction work of dissemination probability in micro-blog community.

Conclusion

In summary, an algorithm of dissemination probability in micro-blog community is constructed on classical SIR model. From the reality of information dissemination in micro-blog community, we divide information dissemination probability into two parts which conclude initial information dissemination probability and dissemination probability in specific time, and make specific portray of each part, relative comprehensively considering possible influential factors. Thus providing relative amount of theoretical foundation and practical calculating tools for further depicting the dissemination process of information. However, this kind of research is not established by large quantities of data collection analysis; some parameters do not conferred with clear value. Therefore, much more in-depth researches are needed in the research of parameters and weight in algorithm.

Acknowledgment

Supported by the Humanities and Social Science Foundation Youth Project of Ministry of Education of China (Grant No. 13YJC630158, 14YJC860022); The China Postdoctoral Science Foundation funded project (Grant No. 2012M510932); The natural science foundation of Heilongjiang Province of China(Grant No. F201209).

References

- [1] MURRAY J D, "Mathematical biology" Berlin, Springer-Verlag, 1993.
- [2] XU Xiao-dong, XIAO Yin-tao, ZHU Shi-ru, "Simulation Investigation of Rumor Propagation in Micro blogging Community", Computer Engineering, vol. 37, 2011, pp. 272-274.
- [3] YAN Gang, ZHOU Tao, WANG Jie, et al. "Epidemic Spread in Weighted Scale-free Networks", Chinese Phys, Lett, vol. 22, 2005, pp. 510-513.
- [4] Zhang Yan-Chao, Liu Yun, Zhang Hai-Feng, et al, "The model of information dissemination based on on-line social network", Acta Phys, Sin, vol. 60, 2011:050501.
- [5] ZHAO Jing, XIA Chengyi, SUN Shiwen, et al, "A novel SIR model with infection delay and no uniform transmission in complex networks", CAAI Transactions on Intelligent Systems, vol. 08, 2013, pp. 128-134.
- [6] Wang Hongliang Zhou Haiwei, "Research on Microblog Opinion Dissemination Law and Controlling Strategy of Sudden Natural Calamity", Journal of Intelligence, vol. 32, 2013, pp. 23-28.
- [7] YANG Zilong, HUANG Shuguang, WANG Zhen, et al, "Study on Micro Blog Reposting Model Based on Characteristics of Information Obsolescence", Computer Science, vol. 41, 2014, pp. 82-85.