

Evaluation of User's Influence Based on Micro-blog Information Forwarded Network

Ke Liang Jia^{*} and Xi Wang

School of Management Science and Engineering, Shandong University of Finance and Economics,
Shandong, China

jiakeliang@sina.com, 908609064@qq.com

Keywords: Sina Micro-blog; User influence; Micro-blog user's attraction; User's activity; Forwarding coefficient

Abstract. In order to evaluate the influence of the users in micro-blog platform, the paper defined three concepts user's attraction, user's activity, forwarding coefficient and described how to calculate them, and then proposed a method to evaluate the influence based on micro-blog information dissemination mechanism. To compared our method with the method based on the number of fans, the results showed that the method had significant correlation with the baseline methods. Through the comparative analysis of the experiment, the provided micro-blog user influence evaluation method was effective.

Introduction

According to CNNIC thirty-eighth statistical report on the development of Internet in China, As of June 2016, there were 242 million Micro-blog users with 34% utilization ratio. Micro-blog platform is still the mainstream platform for Chinese Internet users, it plays an important role in spreading of Internet public opinion. The unique information dissemination mechanism of micro-blog makes the information recipient also become the disseminator of the next round of information transmission. High-impact micro-blog is the basis for rapid dissemination of micro-blog information, the influence of micro-blog is greater, the likelihood that micro-blog information will be forwarded is greater, the information which will be forwarded is more, the magnitude of Microblogging is wider, the speed of the information spreading is more quickly. So many scholars are working on the influence of micro-blog users.

According to the characteristics of micro-blog's information dissemination, Yu Hong etc.[1] collected a single micro-blog's broadcast / comment information and pre-processed them, and then defined a formal description of micro-blog's propagation network. Based on the network, they defined the influence of nodes from two aspects which were local and global, and then analyzed the micro-blog information dissemination path model. Kwak etc. [2] took Twitter as the experimental object, and made detailed comparison of the three different sorting methods of the influence which were based on the number of fans, based on micro-blog forwarded number and based on PageRank, analyzed the correlation between different sorting results. Weng etc. [3] considered the similarity of topics discussed among users and the network structure, extended the PageRank algorithm and proposed a Twitter user impact ranking method, TwitterRank algorithm, which computed the ranking score of the micro-blog users based on the number of micro-blog information released by each micro-blog user and the amount of information received by the fans. Ye etc. [4] calculated and compared the influence based on the fans, the influence based on replied information, the influence based on forwarded information, and considered that the micro-blog user influence value was most stable from the perspective of the maximum replied response. Mao Jiabin etc. [5] analyzed the relationship between user behavior factors and putted forward a method which could analyze and measure the user's social influence by predicting the ability of the user's information dissemination. Yuan Fuyong etc. [6] integrated user's activity and micro-blog influence as micro-blog user's influence factors and synthesized a micro-blog user's influence measurement model which could reduce the impact of micro-blog zombie fan users.

Webberley W etc.[7] pointed out that the information dissemination and diffusion in micro-blog platform mainly relied on user's forward. So this paper evaluated the influence of micro-blog users

based on micro-blog's information forwarding mechanism. Webberley W etc.[7] pointed out that the consecutive forwarded times of an micro-blog information was limited and the probability that the user forwarded the information would decrease with the increase of the length of the forwarding chain. This paper defined the forwarding coefficients of micro-blog information to describe the probability of the information forwarding in order to evaluate the influence of cascade diffusion of micro-blog users.

In the micro-blog social network, typical pushing and forward information mechanism was adopted for information dissemination, so the rapid dissemination of information was achieved through the continuous forwarding of information among different micro-blog users. The paper divided the micro-blog users into information publishers, information forwarders, information terminator according to the users' regular operations in the micro-blog platform. The information publisher is a user who actively released information. The information forwarder referred to the user who forwarded the micro-blog information. The other users who did not forward the information were classified as the information terminator. Among them, the former two were the main force of micro-blog's information dissemination. This paper mainly aimed at the former two users to evaluate the influence of the users and collectively referred to them as "information communicators". In the micro-blog platform, there were many users rarely to publish micro-blog information and rarely to forward information, such users were defined as the information terminator who basically did not work in the spread of micro-blog information.

The organizational structure of this paper is as follows. The second part described the relevant definitions. The third part introduced the research method of this paper in detail. The fourth part used Sina micro-blog data to test the research method proposed in this paper. The fifth part summarized the research work of this paper and pointed out the research direction in the future.

Definitions of the Concepts

Micro-blog User's Attraction. The influence of micro-blog users referred to the audience who affected by released micro-blog information and forwarded information. The spread of micro-blog's information was achieved by the continuous forwarding of micro-blog users. So the forwarding behavior of micro-blog users was an important evaluating indicator of the influence of micro-blog users. In this paper, according to the role of micro-blog users in the micro-blog information dissemination process, micro-blog users were divided into information disseminators and information terminator. In order to distinguish these two types of micro-blog users, this paper defined the concept of micro-blog users' attraction. The users with low attraction would not cause to the fans to forward their micro-blog messages and was identified as the information terminators and had no effect on the spread of micro-blog's information.

The attraction of micro-blog user's showed how attractive their released information was to its fans. The attraction of the users was greater; their released information was more likely to be forwarded by their fans. Micro-blog user's attraction was defined as the proportion of the information which were forwarded to all his released information, the formula was as follows.

$$f_{charm} = \frac{n_{rd}}{n_{sum}} \quad (1)$$

Here, f_{charm} represented the micro-blog user's attraction, n_{rd} represented the number of the being forwarded information by the fans, n_{sum} represented the number of all released information. The users with lower attraction, the released information was less likely to be forwarded by the fans. So in this paper a threshold was set to the micro-blog user's attraction. The paper thought that the micro-blog users whose attraction was below the threshold would not work for the spread of micro-blog information and was defined as information terminator.

Micro-blog User's Activity. Due to the dynamic and real-time nature of micro-blog information dissemination, this paper introduced micro-blog user's activity to distinguish different types of micro-blog users. The paper investigated the posting information, the comment, the forwarding

information of the user in the latest period of time and quantified them to calculate the activity, and then set a threshold value. In the paper a micro-blog user whose activity was less than the threshold value was recognized as a zombie user. The influence of zombie fans were ignored in micro-blog user influence evaluation.

Micro-blog user's activity. micro-blog users' behavior included browsing, posting, forwarding, comments. But the user's browsing behavior was difficult to obtain. So this paper defined the micro-blog user's activity as the total number of user's posting, forwarding and commenting information within a given period of time, the formula was as follows.

$$f_{active} = \frac{n_p + n_r + n_c}{t_e - t_b} \quad (2)$$

Here, f_{active} represented micro-blog user's activity. n_p was the number of the released information by the user. n_r was the number of the forwarded information. n_c was the number of comments. t_b 、 t_e represented the beginning and ending time of the micro-blog user's activity statistics.

Evaluating Method of Micro-blog User's Influence

The key of micro-blog's information dissemination was the push-forwarded mechanism. The message posted by micro-blog blogger could be pushed directly to the fans and brought a direct influence on them. The fans could forward whatever they were interested in. It made the information to be pushed to the fans' fans once again and then resulted in cascading influence. The more micro-blog information was forwarded, the information were pushed to the more users. Therefore, micro-blog's information which was released or forwarded would have an affected crowd. The size of the crowd was a clear indication of the strength of the of the user's impact on micro-blog's Web Platform. Therefore, the paper took the size of people who were effected and covered by the posted and forwarded information as the main criteria for the evaluation of the micro-blog user's influence. And then the paper defined the influence as the range of fans directly affected and potentially affected by micro-blog blogger posted information. That was the possible range of diffusion of the information that posted by the blogger.

$$Inf(u_i) = inf_f(u_i) + inf_r(u_i) \quad (3)$$

Here, $Inf(u_i)$ was the influence of the user u_i . $inf_f(u_i)$ was the direct influence of the user u_i , represented the effective fans of the user. $inf_r(u_i)$ was the cascading influence of the user u_i . Cascading influence referred that the posted information were pushed to more micro-blog users through the u_i 's fans multi-level forwarding. It could expanded the scope of the impact of information and the user's influence. It was defined as follows.

$$inf_r(u_i) = \sum_{j=1}^n \lambda_j \sum_{k=1}^m \alpha_{jk} num_{fan}(u_{jk}) \quad (4)$$

In the formula, λ_j was the cascading coefficient, n was the length of the cascading path, u_{jk} was the k th forwarding user in the j th cascading level. α_{jk} was the user u_{jk} 's forwarding coefficient which represented the forwarding probability of the user u_{jk} to the information posted by the user who was followed with interest by the user u_{jk} . $num_{fan}(u_{jk})$ represented the number of the user u_{jk} effective fans.

The forwarding coefficient represented the impact of a micro-blog user to another one. In order to describe how to determine the forwarding coefficient between two micro-blog users, the paper defined the forwarding attractiveness of one user to one of his fans and the forwarding interestingness of the latter to the former.

There were two micro-blog users u_i and u_j , u_j was the fan of u_i , and then the paper defined the forwarding attractiveness of u_i to u_j was the ratio of the number of information which forwarded by u_j from u_i accounting for all information published by u_i . The forwarding attractiveness of u_i to u_j was greater, the possibility that the information posted by u_i would be forwarded by u_j was greater.

$$f_{charm}(u_i, u_j) = \frac{n_r(u_i, u_j)}{n_{sum}(u_i)} \quad (5)$$

Here, $n_r(u_i, u_j)$ was the number of information which was forwarded by u_j from u_i . $n_{sum}(u_i)$ was the number of information posted by u_i .

The paper defined that the forwarding interestingness of u_j to u_i was the ratio of the number of information which were forwarded by u_j from u_i accounting for all information posted by u_j . The forwarding interestingness of u_j to u_i was greater meant that the possibility that the information posted by u_i would be forwarded by u_j was greater.

$$f_r(u_i, u_j) = \frac{n_r(u_i, u_j)}{n_{sum}(u_j)} \quad (6)$$

Here, $n_{sum}(u_j)$ was the number of information posted by u_j .

Combined with the above two definitions, the paper defined that the forwarding coefficient of u_j to u_i was the geometric mean of the forwarding attractiveness of u_i to u_j and the forwarding interestingness of u_j to u_i . The formula was as follows.

$$\alpha(u_i, u_j) = \sqrt{f_{charm}(u_i, u_j) \bullet f_r(u_i, u_j)} \quad (7)$$

To calculate the cascading coefficients, the paper crawled the micro-blog data and extracted the micro-blog information propagation paths, counted the number of the paths with different forwarding length $n_{l1}, n_{l2}, n_{l3} \dots n_{lj} \dots n_{lm}$. The cascading coefficient was defined as the ratio of the next forwarding paths to the total forwarding paths.

Here, λ_i was the i-level cascade coefficient. n_{lsum} was the number of all cascading paths. n_{lj} was the number of cascading path which length was j, and m was the maximum length of the cascading paths.

Data Acquisition and Experimental Analysis

Data Acquisition. In order to verify the effectiveness of the micro-blog user influence evaluation method proposed in this paper, the paper selected the largest micro-blog data platform "Sina micro-blog" as the source of the test data, used the Python language to design the web crawler. The crawler started from the set seeds of micro-blog user ID, used the URL link to access the user's micro-blog pages to crawl the micro-blog data, and then got their fans' ID and the followers' ID, added the ID to the queue and used breadth priority strategy to crawl the needed micro-blog data. This paper got users ID as a seed from two plates Sina micro-blog special recommendation and 24 hours of popular bloggers users and crawled micro-blog data from May 20, 2016 to June 20, 2016. The data set mainly consists of micro-blog bloggers' information, fans information, followers' information, posting information and so on.

Comparative Analysis of Experimental Results. In order to verify the validity of the proposed method, this paper used a micro-blog user influence evaluation algorithm based on the fans to compare with it. The Spearman rank correlation coefficient [8] was used for comparative analysis.

The experimental results showed that the Spearman rank correlation coefficient of the two methods was 1.000 and the p value was less than 0.05. It meant that there was a positive correlation between the results based on the two methods and the influence calculated based on the users' fans and the methods in the paper were significant relevance. So the method proposed in the paper was effective.

Conclusion

The paper mainly examined several evaluation factors that affected the influence of users, including user's activity, user's attraction, and forwarding coefficient. According to the micro-blog user's forwarding behavior, the paper divided the micro-blog users into three roles: information publishers, information forwarders, and information terminators. The main component of micro-blog communication influence power was information publishers and information forwarders. The paper set the threshold of user's attraction to determine whether the user was the role of information forwarding, and set the threshold of user's activity to exclude the influence of micro-blog zombie fans. Then the paper defined the micro-blog forwarding cascade coefficient to describe the probability of micro-blog multi-level forwarding, and used the data which already had removed the zombie fans data to calculate the influence of micro-blog users. The comparative analysis of the experiment showed that the evaluation method in the paper was effective.

The future work was to choose more other methods and do more experiments in order to verify the method proposed in the paper.

Acknowledgements

This work is supported by Humanity and Social Science foundation of Education Ministry of China (No. 14YJC860011), Shandong province key R & D projects (No. 2015GGB01518).

References

- [1] YU Hong, YANG Xian, Studying on the node's influence and propagation path modes in microblogging, *Journal on Communications*. Vol.33(2012)No.z1, p.96.
- [2] Kwak H., Lee C., Park H., Moon S.. What is twitter, a social network or a news media? [C]. Proc. of the 19th international conference on World wide web (WWW' 2010), New York, NY, USA, p.591.
- [3] Weng J., Lim E. P., Jiang J., He Q.. Twiterrank: finding topic-sensitive influential twitterers[C]. Proc. of the third ACM international conference on Web search and data mining (WSDM '10), ACM, New York, NY, USA, p.261.
- [4] Ye S Z, Wu S F. Measuring Message Propagation and Social Influence on Twitter.com[C]. Proceedings of the 2nd International Conference on Social Informatics. Heidelberg: Springer-Verlag(Laxenburg, Austria, October 27-29, 2010), p.216.
- [5] MAO Jia-Xin, LIU Yi-Qun, ZHANG Min, MA Shao-Ping, Social Influence Analysis for Micro-blog User Based on User Behavior, *Chinese Journal of Computers*. Vol.37(2014)No.4, p.791.
- [6] Yuan Fuyong, Feng Jing, Fu Qianqian, Influence Index Model of Micro-blog User, *XIANDAI TUSHU QINGBAO JISHU*. (2012)No.6, p.60.
- [7] Webberley W, Allen S, Whitaker R. Retweeting: A study of message-forwarding in twitter[c]. 2011 Workshop on Mobile and Online Social Networks. Milan: IEEE(2011), Italy, p.13.
- [8] Myers J L, Well A D. *Research Design and Statistical Analysis*(2nd Edition). New Jersey, USA: Lawrence Erlbaum Associates Publishers, 2003.