Research on the evaluation of user's influence in micro-blog platform

Zhi Nuo Li[1, a] and Ke Liang Jia[2, b *]

[1]School of Law, Shandong University of Finance and Economics, Jinan, China

[2]School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan, China

[a]lizhinuo@163.com, [b]jiakeliang@sina.com

Keywords: User's influence; Micro-blog; Social networks; User's characteristics; Micro-blog topics;

Abstract. The paper elaborates and analyzes several methods of the evaluation of user's influence in micro-blog platform basing on extant researches, and generalizes three kinds of methods: the methods based on user's characteristics, the methods based on social networks topology and the methods based on topics. Then it reviews related documents and discusses the deficiencies of the current research methods and future trend of the research.

Introduction

In 2014 July CNNIC released the thirty-third times "China Internet development statistics report", as of 2014 June, China micro-blogging users scale was 0.275 billion. Micro-blog has some characteristic such as large amount users, fast spread, high real-time, which makes micro-blog platform in the exchange of information convenient for people at the same time, it also brings a series of problems, such as garbage information, false information, network rumor spread etc. Information dissemination in micro-blogging social network uses pushing and forwarding mechanism. Information could be quickly speeded by many forwarding. The influential users are opinion leaders, they play a key role in the rapid spread of information．They have direct or indirect impact on large numbers of users of micro-blogging within a short time．Their original or forwarded information could be paid attention to and be forwarded. So the information could be propagated rapidly in the micro-blog platform. How to effectively measure the user's influence in the Chinese micro-blog platform has become a hot research problem in the fields of the discovery and early warning of public opinion and the other related area．This paper evaluates the extant ranking method of user's influence and classifies them to three kinds, discusses their deficiencies and proposes the idea and the direction of future research.

User's Influence Ranking Method Based on User's Characteristics

In the micro blog, forum, Face book, Google plus and other social public network, many scholars provided many methods to rank the user's influence. Some scholars use the user's characteristics to do that, such as the fans and the friends.

In order to identify influentials on Twitter, Kwak[1] .etc ranked users' influence by the number of followers, by PageRank and by retweets three different methods and compared the results. The results showed that the two former rankings to be similar and ranking by retweets differed from the previous two rankings. It indicated a gap in influence inferred from the number of followers and that from the popularity of one's tweets. Cha[2] .etc thought directed links in social media could represent anything from intimate friendships to common interests. So they collected a large amount of data from Twitter and then compared in depth the three measures of influence: in-degree, retweets, and mentions. The results showed popular users who had high in-degree were not necessarily influential in terms of spawning retweets or mentions. Leavitt[3] .etc represented a new methodology based on the content and responses of 12 popular users to determine the measurements of relative influence on Twitter, but didn't compare with other methods.

The disadvantage of these methods is only to consider the attribute information of the user, and not to consider the interaction between users in a social network. In the micro blog social network,

the user's influence is reflected by the information transmission between the users and the information transmission depends on the linked relationship of the users.

User's Influence Ranking Method Based on Network Topology

In the micro blog social network, the relationship between users is constructed by concerning other users and then a complex network is built. So some researchers did some work according to the network topology structure, ranked the user's influence based on the classical random walk algorithm such as PageRank. The notable one was TunkRank[4] which was proposed by extending and improving the PageRank algorithm. It used a constant to represent the forwarding probability by users, combined with the people whom the user concerned and the fans who concerned the user to rank the user's influence. The user's influence was the expected number of the people influenced by the released information in TunkRank. The core formula of TunkRank was as following, more details in document[4].

$$Influence(X) = \sum_{Y \in Followers(X)} \frac{1 + p * Influence(Y)}{|Friends(Y)|}$$

(1)

Lv .etc[5] devised an adaptive and parameter-free algorithm, the LeaderRank, to quantify user influence by revising the traditional Google's PageRank algorithm. The results showed that LeaderRank outperformed PageRank in terms of ranking effectiveness, as well as robustness against manipulations and noisy data and suggested that leaders who were aware of their clout could reinforce the development of social networks, and thus the power of collective search. The core formula were as following, more details in document[5].

$$s_i(t+1) = \sum_{j=1}^{N+1} \frac{a_{ji}}{k_j^{out}} s_j(t)$$

(2)

$$s_i = s_i(t_c) + \frac{s_g(t_c)}{N}$$

(3)

Bakshy[6].etc investigated the attributes and relative influence of 1.6M Twitter users by tracking 39 million diffusion events that took place on the Twitter follower graph over a two month interval in 2009 and found that the largest cascades tended to be generated by users who had been influential in the past and from URLs that were rated more interesting and/or elicited more positive feelings by workers on Mechanical Turk. Limitations of this study are that the scope of the information dissemination is not only related to the information releaser but also related to the information content itself. So the scope of the information dissemination reflects the comprehensive influence of the user and the information, and does not accurately reflect the influence of the releaser. Liang Ping[7].etc adopted the methods of centrality and central potential indexes, combined with the micro blog users network topology of "concerning" and "concerned" and analyzed the centrality of the micro-blog network from point centrality, betweenness centrality and closeness centrality finally researched the power of the influentials.

These algorithms such as TunkRank were designed for the all network users, reflected the overall influence in the network. Their shortcomings are not to take the correlation of the topic between users into account. It is whether the topics are similar. In general, the topic similarity between two users is greater; the mutual influence between them is also more.

User's Influence Ranking Method Based on Topic

In Micro-blog social network, user's influence didn't only stand out in their own attributes and network structure features，it was also highly relevant to the topics in which the user participated．

Weng[8] .etc observed that (1) 72.4% of the users in Twitter followed more than 80% of their followers, and (2) 80.5% of the users had 80% of users they were following follow them back and revealed that the presence of "reciprocity" could be explained by phenomenon of homophily. Based on this finding, they proposed a new algorithm TwitterRank which was extended based on PageRank to measure the influence of users in twitter. TwitterRank measured the influence taking both the topical similarity between users and the link structure into account. Experimental results showed that TwitterRank outperformed the original PageRank and topic-sensitive PageRank. Haveliwala[9] .etc proposed a Topic-sensitive PageRank algorithm to rank the user's influence, its difference to TwitterRank was whether the influence of transition probability was equal to different topics. The core formula of TwitterRank were as following, more details in document[8].

$$\overrightarrow{TR_t} = \gamma P_t \times \overrightarrow{TR_t} + (1-\gamma)E_t \tag{4}$$

$$\overrightarrow{TR} = \sum_t r_t \bullet \overrightarrow{TR_t} \tag{5}$$

Wu Xianhui[10] .etc built a topic-related graph model of micro-blogging according to users' attributes and mutual information among them, then adopted the idea of random walk in the algorithm to evaluate the user's influence by finding central nodes of the graph model.

Feng Shi [11] .etc proposed an algorithm combined LDA model and HowNet to classify the short texts in micro-blogs based on their underlying subtopics, then introduced an influence measurement containing the criteria on explicit，implicit and user features and employed an analytic hierarchy process method to assign different weight of each parameter．Experiment results showed that the proposed short text classification method outperformed the traditional SVM based method，and the proposed influence measurement model could effectively detect the user with strong influence in the Chinese hot topic micro logs.

Xiong Xiaobing [12] designed and implemented a topic-sensitive influence ranking algorithm named WTSIRank. WTSIRank first used a node-weighted graph to illustrate the social network. Weight value was mainly determined by the content and the number of the published posts. A method was designed and impletemented to calculate the node weight based on correlation measuring and originality distinguishing. The transition probability between individuals was calculated using the correlation coefficient between the published posts and the concerned topic, and used the node weight to fix the influence score of each individual. With two real-world data sets from Twitter, he compared the algorithm of WTSIRank with DegreeRank, PageRank, LeaderRank and TwitterRank. The comparing results showed that the highest correlation coefficient between WTSIRank and other algorithms was 0.71, and WTSIRank outperformed all other methods and could exploit the key individuals with more accuracy.

These algorithms didn't care whether the users were active in micro-blog network. There are a lot of "stiff fans" in the micro-blog platform, they didn't release and forward any information. So they have no effect on the spread of information and their influence could be ignored.

Conclusion and Future Trend

Conclusion. We summarize deficiencies of the current research methods as following.

The methods based on user characteristics made full use of the characteristics of the user itself, considered the importance of the user itself, but ignored the user as a node of micro-blogging social networks which was affected by the neighboring nodes. If the influence of an adjacent node is greater, it will bring the user more influential.

The algorithms based on topological structure paid attention the relationship of users in a social network, but ignored whether the users were interested in the forwarding information．If the user was not interested in the information, he could ignore it and couldn't forward it. So to introduce the

topic relevance between the users to rank the user's influence will make the evaluation results more accurate.

The methods based on the topology and the topic improved the ranked results and made the evaluation results more objective, but ignored the existence of "stiff fans" in the micro-blog platform.

Micro-blog platform for social network is a dynamic network, the above research methods were studied in static perspective, without considering the dynamic change of network node.

Future Trend. We prospect the future trends of the research as following.

The Further research should introduce the user's liveness to rank the user's influence in order to avoid the effect of "stiff fans" and achieve a better performance. It could improve the algorithm accuracy and anti-jamming capability of the new algorithm.

The future research should consider the dynamics of the micro-blog platform networking and research the dynamic change rules, study the effect of new users and new " stiff fans ", further research micro-blog network evolution mechanism.

Acknowledgments

References

[1] H. Kwak, C. Lee, H. Park and S. Moon: What is twitter, a social network or a news media?. *Proc. of the 19th International Conference on World Wide Web* (WWW' 2010)(New York, NY, USA, 2010), p. 591-600.

[2] M. Cha, H. Haddadi, F. Benevenuto and K. P. Gummad: Measuring user influence on twitter: The million follower fallacy. *Proc. of the 4th Int'l AAAI Conference on Weblogs and Social Media*, (Washington, DC, 2010), p.10-17.

[3] A. Leavitt, E. Burchard, D. Fisher and S. Gilbert: The influentials: New approaches for analyzing influence on twitter, Information on http://www.twylah.com/StilesCarrie/tweets/ 245590535031631872.

[4] T. Daniel: TunkRank, Information on http://tunkrank.com/.

[5]L.Y.Lv, Y.C. Zhang, C.H. Yeung and T. Zhou: Leaders in social networks, the delicious case, Information on http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0021202.

[6] E. Bakshy, J. Hofman, W. Mason and D. J. Watts: Identifying Influencers on Twitter, Information on http://thenoisychannel.com/2011/04/16/identifying-influencers-on-twitter/.

[7]L. Ping and L.Y. Zong: Research on Microblog Information Dissemination Based on SNA Centrality Analysis-A Case Study with Sina Microblog, Document, Information & Knowledge. (2010) No. 6, p.92-97 (In Chinese) .

[8] J. Weng, E. P. Lim, J. Jiang, Q. He: Twitterrank: finding topic-sensitive influential twitterers, *Proc. of the third ACM International Conference on Web Search and Data Mining* (WSDM '10), (New York, NY, USA, 2010), p. 261-270.

[9] T. H. Haveliwala: Topic-sensitive PageRank, *Proc. of the 11th International Conference on World Wide Web*, (New York, NY, USA, 2002), p. 517-562.

[10] X.H. WU, H. ZHANG, C.C. YANG, B. LI and X.J. ZHAO: An Algorithm of Topic-related Microblogging Opinion Leaders Mining, Journal of Chinese Computer Systems, Vol.35(2014)No.1, p.2296-2301 (In Chinese).

[11] S. Feng, S. Jing, Z. Yang and D.L. Wang: Detecting Topical Opinion Leaders Based on LDA Model in Chinese Microblogs, Journal of Northeastern University(Natural Science), Vol.34(2013) No.4, p. 490-494 (In Chinese).

[12] X.B. Xiong: *Research on Key Issues of Spreading Behavior in Microblogging Network*(Ph.D., PLA Information Engineering University, China 2013) , p.59-72(In Chinese) .