Send Orders for Reprints to reprints@benthamscience.ae

The Open Cybernetics & Systemics Journal, 2014, *8,* 749-753

Open Access

New Method for Micro-blog User Influence Evaluation based on Transfer Entropy

Lihong Song[1], Haiyan Wang[1], Yijing Zhang[1], Peiguang Lin[1,2,*] and Peiyao Nie[1]

[1]*School of Computer Science & Technology, Shandong University of Finance & Economics, Jinan, 250014, China;* [2]*Research Center of Financial Information Engineering of Shandong Province, Jinan, 250014, China*

Abstract: Micro-blog is essentially a kind of web service platform, and it had a wide space with the development of mobile Internet. As an important part of social network, influences among the micro-blog users are becoming a hot spot in the research of the micro-blog. This work has a very important theoretical and practical significance for monitoring public opinion. Through the analysis of user's behavior patterns using the transfer Entropy theory, this paper set up an improved model for better evaluating micro-blog users' influences. We processed and analyzed micro-blog users' behaviors based on time series, and focused on the users' "tweet", "retweet", "comment" and "call"(@) behavior patterns. In turn, this would allow us to gain a better understanding of the characteristics of the new web service platform, and at the same time to find its potential values for researches and applications. To validate our method, we crawled data from the Sina weibo, and these data was processed and analyzed by our method in this paper. The experiment result showed that our method performed well on evaluating users' influences.

Keywords: History behavior, micro-blog user's influence, transfer entropy.

1. INTRODUCTION

With the development of mobile information, the functions of micro-blog became more mature, and the number of micro-blog's user grew rapidly in recent years. Micro-blog plays a more important role in people's daily life and the social network platform. Now more and more social events are released on the micro-blog platform in the first time, and soon become a focus, such as "3.01 incident" in Kunming, Yunnan province, "MH370 disappearances", etc. According to the latest report issued by the China Internet Network Information Center, by the end of December 2013, micro-blog user's scale was for 477 million in China. Among these years, more and more scholars have carried out a large number of micro-blog researches related to theory and practice in various fields on the basis of the social network.

In the social network platform, each user is not only the disseminator, but the audience as well. It is playing the two roles at the same time that greatly improves convenience and interactive communications. As important sources of information, some users often provided valuable information. We call such users influential ones who often published novels and original ideas, or they could attract other users to participate in discussions, and could affect the people's view point. Micro-blog users as an important part of social network, their influences among each others were becoming a hot spot in the research of the micro-blog. This work has very important theoretical and practical significance for monitoring public opinion. It is conducive to timely track the key characters of hot topics, and to provide support for micro-blog public opinion analysis.

Current academic researches on micro-blog users' influences mainly involved a variety of factors, such as the number of users' followers, the number of comments, and the number of forward, etc. In this paper, we proposed a new evaluation method of micro-blog users' influences, which was based on the transfer entropy theory and combining the above factors.

2. RELATED WORK

With the rapid development of social network, it gradually becomes a hot academic research in recent years. Many literatures studied in micro-blog users' influences, but the definition of "influence" had different methods and angles.

In previous studies, the number of a user's followers was often used to indicate his influences. By this method, the number of followers was regarded as the indicator directly to measure the influence of micro-blog users. The more followers a user has, the bigger the influence is. This way of measurement was more based on people's first impression of daily experience, but in many cases was inaccurate.

Considering the differences among social network users, more and more researchers cared about the user's individual characteristics, such as preferences, the relations among different users. Literature [1] concluded some basic characters of micro-blog through the theory of small world by analyzing the network features, degrees and central degree of the micro-blog, and some problems were proposed, including asymmetrical information, uneven micro-blog network.

Literature [2] studied the users' influences by analyzing the three indictors: the number of followers, times of "forward" and times of "mentioned". It indicated that the number of followers cannot reflect the overall influence of micro-blog users, although it can reflect the user's concerned de-

1874-110X/14

gree. At the same time, this literature proposed that the times of users mentioned and blogs forwarded can reflect the influence of the users and the blogs respectively.

About the research of the way of information dissemination, existing work can be divided into two aspects. One is by the network topology and the other by the individual difference and interaction in the transmission of rules [3-5]. In addition, some researchers analyzed and evaluated [6-8] the influence of micro-blog users by use of the method of web link analysis in data mining, such as the Page-Rank or HITS.

Above all, previous research works regarded micro-blog users as the homogeneous, without considering the differences among users, ignoring the fact that the different users will have different preferences for different information. For example, Epidemic model [9, 10] or the rumor model [11, 12], which was based on complex network, marked the public information network node as two states, susceptible and infected. Each of the infected nodes could infect the adjacent susceptible nodes under a certain probability, simulating the spread of the virus, and eventually all nodes became infected. At the same time, the previous research works measured users' influences by using fewer features of micro-blog users or combining some features simply. In fact, the development of micro-blog network is dynamic, and the micro-blog users' influences is mainly manifested in the information dissemination in network [13-15]. If the information generated by a user was forwarded or commented by a large number of micro-blog users, we called the information had higher influences. Accordingly, the greater the influence of micro-blog users had, information released by him relatively will become more valuable and affect more people. Although these methods can evaluate users' influences, it can be affected by man-made factors on the evaluation results, and cannot truly reflect the actual situation of the users. At the same time, because sample size was small or the samples could not represent the universal set, some researches had weaker persuasion.

Therefore, this paper proposed a new method for micro-blog users' influences evaluation which was based on the model of transfer entropy. This method can not only explain the actions in the social network better, including "tweet", "retweet", "comment" and "forward" etc., but also reflect the relationship among the users and dig out the influential users.

3. ANALYSIS OF MICRO-BLOG USERS' INFLUENCES

3.1. Definitions of Micro-Blog Users' Influences

Literature [16] proposed that the user's influence was essentially of the interaction among micro-blog users. And corresponding to Chinese dictionary, the definition of "influence" is the "effect" for someone else's ideas or behavior. In other words, the influence can be considered as the ability of a user's behavior to impact on others.

On the micro-blog platform, the influence can be embodied in the effect of the action, including the "tweet", "retweet", "comment", "forward", on other's idea of behaviors. And the other's idea can also be embodied on the action mentioned before. Therefore, this paper proposed that the influences of micro-blog users can be evaluated by the users' behaviors.

3.2. Transfer Entropy

Transfer entropy theory was proposed in 2000 Daniel Schriever [17]. Transfer entropy, a non-parametric measure of co-dependence, is identical to (conditional) mutual information measured in nats (using the natural logarithm) [18]. Assume that data from two sequences X and Y are simultaneously available at k time-stamps: $t_{n-k+2:n+1} \equiv \{t_{n-k+2}, t_{n-k+2}, ..., t_n, t_{n+1}\}$. Then we express transfer entropies as:

$$TE^{(k)}_{x\to y} = I(y_{n+1}; x_{n-k+2:n} \mid y_{n-k+2:n}) = H(y_{n+1} \mid y_{n-k+2:n}) - H(y_{n+1} \mid y_{n-k+2:n}, x_{n-k+2:n}) \quad TE^{(k)}_{y\to x} = I(x_{n+1}; y_{n-k+2:n} \mid x_{n-k+2:n}) = H(x_{n+1} \mid y_{n-k+2:n}) - H(x_{n+1} \mid x_{n-k+2:n}, y_{n-k+2:n})$$

Each of the two transfer entropy values $TE_{x\to y}$ and $TE_{y\to x}$ is nonnegative and both will be positive (and not necessarily equal) when information flow is bi-directional.

3.3. Factors of Micro-blog Users' Influences Evaluation

Micro-blog users' influences were composed by multiple factors, and the intuitive one of them was the number of users' followers. The number can reflect the influence of micro-blog users to some extent. In general, a bigger number indicates a bigger influence. Moreover, micro-blog users' behaviour was the important factors of micro-blog users' influences. In general, if the blog released by a user was retweeted more times, or was commented more times, we thought it had higher influence.

3.4. Introduction of Our Evaluation Model

This model was constructed by a directed graph, G=<U,E>, in which, U presented micro-blog users and E presented the relationship between the users.

On the graph G, for $\forall u_1, u_2 \in U$, if $<u_1,u_2> \in E$, it indicates a relationship between u_1 and u_2. U_1 is u_2's followee, u_2 is u_1's follower. In the figure is expressed as $u_1 \rightarrow u_2$.

On the micro-blog platform, if u_2 is u_1's follower, u_2 can see the blogs released or forwarded by u1, but u_2 may not forward these blogs. This paper focused on the probability of u_2 forward the blogs released by u1, or in other words, focused on the influence of u_1 to u_2 aiming at the specific blogs.

For any users u the micro-blog platform recorded all users' activities, such as releasing micro-blog (tweet, expressed by t), forwarding (retweet, expressed by rt), commenting (expressed by c) and calling other users (@, expressed by at). Therefore, we defined user u's sequence of activities:

$$S_u = \{a_t\}, t = 0, 1, ... \quad (1)$$

In (1), a_t denoted an event at the time t, a_t=t(id) denoted that, at time t, user u tweeted blog with id. a_t=rt(id,sid) denoted that, at time t, user u forwarded blog id, along with original micro-blog blog sid. a_t=at(uid) denoted that, at time

t, user u called user uid. a_t =c(id) denoted that, at time t, user u commented blog id.

At the same time, we give the following functions:

$id_u(a_t)$: denoted the blog with behaviour a_t,

$sid_u=(a_t)$: denoted the original micro-blog id with the user u's behaviour a_t,

$uid_u=(a_t)$: denotes the user u's id with behaviour at,

$c_u=(a_t)$: denotes the comment id with user's u behaviour a_t,

Fe(u): denotes all of the user u's followee,

Fr(u): denotes all of the user u's follower.

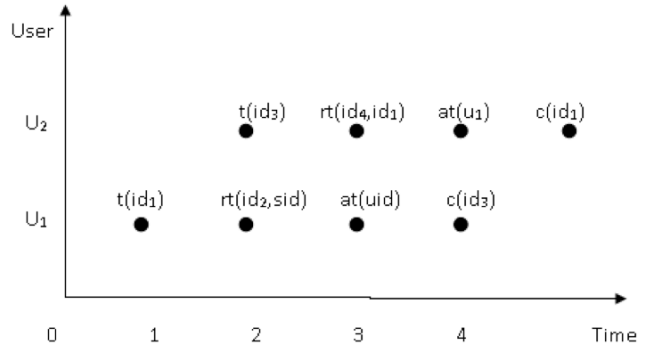

Fig. (1). Users' activities sequence diagrams.

Under the condition of not causing confusion, this paper also used id, sid, uid, cid to denote the current micro-blog id, the original micro-blog id, user id, comment id.

3.5. Solution to The Problem

In this section, we gave the computation of the influence of users u_1 to u_2.

Because transfer entropy theory described the influence between two objects and embodied the relationship between them, this paper fully considered the users' operation history and micro-blog semantic, and combined with entropy theory to our research.

Firstly, we defined the random variable that reflected the user's behaviours:

$$B_u(t_1,t_2)=\begin{cases}1 & if \quad \exists a_i \in S_u \wedge i \in [t_1,t_2] \\ 0 & otherwise\end{cases} \tag{2}$$

Hence, the probability of random variable can be expressed as:

$$P(B_u(t_1,t_2)=a_i)=\frac{1}{t_2-t_1}\int_{t_1}^{t_2} B_u(t_1,t_2)dt \tag{3}$$

Which can be simplified as $P_\alpha(t_1,t_2)$ or P_α.

At the same time, the paper proposed the joint random variable reflected the relationship between two users' behaviours.

$$B_{u_1,u_2}(t_1,t_2)=$$
$$\begin{cases}1 & if \quad \exists a_i=rt \in S_{u_1} \wedge i \in [t_1,t_2] \wedge \exists a_j=t \in S_{u_2} \wedge j \in [t_1,t_2] \wedge sid(a_i)=id(a_j) \\ 1 & if \quad \exists a_i=at \in S_{u_1} \wedge i \in [t_1,t_2] \wedge \exists uid(a_i)=u_2 \\ 0 & otherwise\end{cases} \tag{4}$$

Based on the two probability distribution, we can get the joint probability distribution $P_\alpha^{u1,u2}$ between the users.

Secondly, we gave the definition of $A_u(t_1,t_2)$ to denote the behaviours of user u at the time $[t_1,t_2]$:

$$A_u(t_1,t_2)=\{a_i \mid a_i \in S_u \wedge i \in [t_1,t_2]\} \tag{5}$$

Based on the related research of semantic similarity, this paper got the similarity between the blog id and the historical behaviours:

$$Sim_u(id,A_u(t_1,t_2))=\frac{1}{n}\sum_i sim(id,a_i) \tag{6}$$

In which, n is the number of $A_u(t_1,t_2)$, $a_i \in A_u(t_1,t_2)$.

On the basis of the above definition of similarity, we gave the following definition of random variables under the constraints to similarity threshold α_h.

$$S_n(t_1,t_2)=\begin{cases}1 & if \quad Sim(id,A_u(t_1,t_2))>\alpha_n \\ 0 & otherwise\end{cases} \tag{7}$$

Then we got the probability:

$$P_s(S_n(t_1,t_2))=\int_{t_1}^{t_2} S_n(t_1,t_2)dt \tag{8}$$

Which was simplified as $P_s(t_1,t_2)$ or P_s.

Based on the formula 3 and formula 8, we got the joint probability $P_{a,s}$ between the user's behaviours and the similarity.

Finally, we gave the following influence definition between user u_1 and u_2:

$$I_{u_1 \to u_2}=E(a_i=rt \mid H_{u_2})-E(a_i=rt \mid H_{u_1},H_{u_2}) \tag{9}$$

In which, the first part denoted the entropy of the blogs forwarded under the condition of the knowledge of the history of u_2:

$$E(a_i=rt \mid H_{u_2})=-\sum_{a_i,H_{u_2}} P(a_i,H_{u_2})\log(P(a_i \mid H_{u_2})) \tag{10}$$

The last part denoted the entropy by the supplement the history of u_1:

$$E(a_i=rt \mid H_{u_1},H_{u_2})=-\sum_{a_i,H_{u_1},H_{u_2}} P(a_i,H_{u_1},H_{u_2})\log(P(a_i \mid H_{u_1},H_{u_2})) \tag{11}$$

In theory, when the historical information of user u_1 was supplied, its entropy will be reduced, and then $I_{u1 \to u2}>0$.

Moreover, the bigger $I_{u1 \to u2}$ was, the greater the contribution of history information had. In other words, u_2 had the bigger probability to forward the blogs. Inversely, if $I_{u1 \to u2} \leq 0$, that meant u_1 had no influence to the u_2, or had a

Table 1. Statistics of sina weibo users' information.

	Send	Forward	Comment
$User_1$	7	5	10
$User_2$	10	3	11
$User_3$	5	7	31
$User_4$	32	10	7
$User_5$	10	2	3

Table 2. Sina weibo users' influences.

	$User_1$	$User_2$	$User_3$	$User_4$	$User_5$
$User_1$		0.01	0.015	-0.165	-0.121
$User_2$	-0.007		0.007	-0.175	-0.109
$User_3$	-0.186	-0.15		-0.011	-0.189
$User_4$	-0.435	-0.44	-0.11		-0.104
$User_5$	-0.373	-0.367	-0.233	0.418	

negative one, because the unknown knowledge did not reduce after supplying the history of u_1.

4. EXPERIMENT AND ANALYSIS

Based on the above model, we designed the following experiment to analyze and compare micro-blog users' influences quantitatively.

4.1. Data Preparation

Through the latest report of Sina weibo, the number of Sina weibo users had exceeded 500 million. Comparing with Tencent weibo and other platforms, Sina weibo had the advantages in the number of users, the number of active users, and the influence of platform and so on. So we selected the Sina weibo as a sample in our research.

There are two main ways to retrieve data from micro-blog: the first one is through the Sina micro-blog open API and the second one is using the web crawler.

Including the users' original blogs, the forwarded blogs, and the comments, the data was obtained by Java based on the open API. We collected the data from June 5, 2014 June 12, 2014, including 5 micro-blog users, a total of 203 micro-blog informations. In our process, in order to ensure the accuracy of analysis and feasibility, we eliminated invalid informations and divided the data into four groups ("tweet", "retweet", "comments" and "call"). Then we analyzed and computed the data according to the above method.

4.2. Result Analysis

We selected 5 users and calculated influence among them. We analyzed the 5 Sina users' informations, and the results were shown in the Table **1**.

Then, we calculated the influence among the above five users, and the results were shown in the Table **2**.

We can easily figure it out that $User_5$ had the highest influence on $User_4$ and $User_4$ had the maximum of negative influence on $User_2$. In other words, when $User_5$'s historical information were supplied, and its entropy decreased ($I_{u1\to u2>0}$) and $I_{User5\to User4}$ was the biggest. It indicated that $User_5$'s history of information had the largest contributions to $User_4$, and $User_5$ had the biggest influence on the $User_4$. Or, $User_4$ had the biggest possibility to forward the latest blogs. Consequently, in the actual situation, $User_5$ released a total of 15 Sina blogs during the period, in which $User_4$ forwarded 2 times, commented 3 times, and this number was significantly higher than the number of $User_4$ forwarded and commented others Sina blogs. This can show that the influence of the $User_5$ to the $User_4$ was the biggest when compared with other users. On the contrary, $I_{User4\to User2}\leq 0$ indicated that when the history information of $User_4$ were supplied, the unknown information did not reduce, or even increase the uncertainty. In practice, $User_4$ released 49 Sina blogs during the period, and $User_2$ forwarded and commented none of those blogs. This was significantly less than $User_2$'s behaviours towards other Sina weibo users which meant that $User_4$ had the least influence to $User_2$.

CONCLUSION

In this paper, we proposed a new method for micro-blog users' influences evaluation based on Transfer Entropy. Our method considered several factors including the number of micro-blog users' followers, the number of comments, the forwarding, etc. Our experiments showed that the method had good perform. In the future, we would evaluate the influence of multi-users to a single micro-blog user, and the influence to micro-blog users' fans.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This paper belongs to the project of the "National Science Foundation of Shandong Province", No. 2009ZRB019PF.

EXPLANATION

Lihong Song, Yijing Zhang and Haiyan Wang were the undergraduates of SDUFE. Song and Zhang's major were e-Commerce, and Wang's major was Computer Science and Technology.

REFERENCES

[1] ZHAO LING, ZHANG JING. "Analysis of Micro - blog User Behaviour Based on Complex Network" [J]. Journal of Modern Information, 2013, 5: 015.

[2] CHA M, GUMMADI K P. "Measuring User Influence in Twitter: The Million Follower Fallacy" [J]. Artificial Intelligence, 2010, 146(1):10-17

[3] SAITO K, KIMURA M, OHARA K, etc. "Learning Continuous Time Information Diffusion Model for Social Behavioural Data Analysis" [J]. Advances in Machine Learning, Springer, 2009(5828) :322-337.

[4] HADDADI H, BENEVENUTO F, GUMMADI K P. "Measuring User Influence in Twitter: Million Follower Fallacy" [C]. Proceedings0fInternational AAAI Conference on Weblogs and Social. Washington, DC: ICWSM, 2010:97-105.

[5] BHARATHI S, DAVID K, MAYHARS. "Competitive Influence Maximization in Social Networks Internet and Network Economics" [J].Springer, 2007(2315):30f-311.

[6] WENG JIANSHU, LIM EE-PENG, JIANG JING. "TwitterRank: Finding Topic-sensitive Influential Twitterers" [J]. New York, 2010, Paper 504(1):261-270.

[7] HAVELIWALA T H. "Topic-sensitive PageRank" [C]//Proc of the eleventh international conference on World Wide Web. New York: ACM Press, 2002:517-526.

[8] ROMERO D M, GALUBA W, ASUR S, *et al.* "Influence and Passivity in Social Media" [J]. Information Systems Journal, 2010, abs/1008.1(4):1-9.

[9] V M EGUILUZ, K KLEMM, "Epidemic Threshold in Structured Scale-Free Networks" [J], Physical Review Letters, 2002, 89 (10): 108701 1-4.

[10] C Y JI, D Q JIANG, N Z SHI, "Multigroup SIR epidemic model with stochastic perturbation"[J], Physica A: Statistical Mechanics and its Applications, 2011, 390(10): 1747-1762.

[11] MORENO Y, NEKOVEE M, PACHECO A F. "Dynamics of Rumor Spreading in Complex Networks" [J]. Physical Review E, 2004, 69(6) :066130 1-7.

[12] ZHOU J, LIU Z H, LI B W. "Influence of Network Structure on Rumour Propagation" [J]. Physics Letters A,2007,368(6) :458-463.

[13] BAKSHY E, HOFMAN J M, WATTS D J, *et al.* "Identifying "Influencers" on Twitter" [J]. Communication, 2011, 1-10.

[14] DAVE K, BHATT R, VARMA V. "Identifying Influencers in Social Networks"[C]//Proc of the Fifth International Conference on Weblogs and Social Media. Palo Alto, CA: AAAI Press, 2011, 1-9.

[15] KIMURA M, SAITO K, NAKANO R, *et al.* "Extracting influential nodes on a social network for information diffusion" [J]. Data Mining and Knowledge Discovery, 2009, 20(1):70-97.

[16] LIU YAOTING. "Research on Social Network Structure" [D]. Hangzhou: Zhejiang University, 2008.

[17] SCHREIBER, T. "Measuring Information Transfer" [J]. Phys. Rev. Lett. 2000, 85, 461–464.

[18] HAHS D W, PETHEL S D. "Transfer entropy for Coupled Autoregressive Processes: [J]. Entropy, 2013, 15(3): 767-788.

Received: September 16, 2014 Revised: December 23, 2014 Accepted: December 31, 2014