Credibility Ranking Methods Analysis of Users in Social Network Based on Relation Graph

Cai Xiang[1,a] , Weihong Han[2,b], Yan Zhou[3,c], Wenxiang Han[4,d], Shudong Li[5,e] Xiaobo Wu[6,f]

[1,2,3,4,5]Graduate school of National University of Defense Technology, Changsha, 410073, China

[6]School of software engineering Yantai Vocational College, Yantai, 264670, China

[a]email: ealgefly2017@163.com, [b]email: hanweihongnudt@139.com,[c]email: candle531@126.com,

[d]email: wenxiang2018@126,[e]email: lishudong@126.com, [f]email: happywxb@163.com

Keywords: Credibility; Ranking Methods; Social Network; Relation Graph

Abstract. The increasing proliferation of social media results in users are forced to ascertain the truthfulness of information that they encounter from unknown sources using a variety of indicators. In this paper, we propose a novel method to rank credibility of users in microblog social network. Microblog social network site provides lots of functions, like sharing, organizing and finding content and contacts. We build a weighted user relation graph model and design a relation graph based ranking algorithm, called Relation Graph Credibility Ranking (RGCR) algorithm, which can be applied to analyse user's credibility by their relationship. We perform experiments on three datasets, they are "The Continent", "Sangfor Tournament" and "Sina Campaign" crawled by API from the Sina microblog open platform. The experimental results show that our proposed method works well to analyse credibility of users in microblog social network.

Introduction

Microblog social network allow users to interact and collaborate with each other as creators of tremendous amounts of user-generated content or information [1,9]. Despite many users using it to get positive purpose, there is also a potential misuse of this technology to perform various malicious activities, such as spreading rumors or false messages on an inflammatory topic, creating accounts for false identities, and other; users engage in these activities to achieve high influence, instigate chaos, or even sabotage public safety [2,4]. Therefore, it is very important to detect the sources of misinformation; this can be done by analysing and ranking the credibility of users in social network. Different from others, in our experiment datasets we filter the users whoes blogs less than 50 but the followers more than 1000, for they either harmless or exemption, and we classify the users as "public account" who marked by "V" in Sina microblog and "Interest Homepage", because they are more easier to be followed and impact less on credibility of users[14] .

Researchers in the field of user credibility have been studied in conjunction with several other topics, including information diffusion, trust, user recommendations, and reputation. Most of these previous works propose two general types of methods [6,8,13]. Machine learning-based methods can be supervised or unsupervised, and they are based on building classifiers that determine credibility scores for blogs as a measure of their factuality. Examples of this type are presented in[5,11,12]. These methods require recursive processing of data to train and test the algorithms to achieve the desired accuracy. They talk about the "effective communicate degree", but did not fully use the user's relationship to evaluate its credibility[7,10,12].

Therefore, we propose a method based on users' relation graph to analyse their credibility. In the microblog social networks, a user can register to receive dynamic information automatically from another user by 'following' the user of interest. The follow relationship is asymmetrical on microblog: the user being followed does not necessarily follow his/her follower. Microblog is considered to be a news medium as well as a social networking medium, because its user relationships, or links, are directed on the social graph [3,7]. Nevertheless, symmetric relationships

Copyright © 2017, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

can exist on microblog: two users can follow each other. We defined a relation digraph G =(U, E), 'U' is the node of the graph, represents the users; 'E' is the edge of the graph, represents the relationship between users, and the 'E' have it weight which refered to the mutual-following relationship.

Design of the Relation Graph Credibility Ranking (RGCR) algorithm

We defines a user's weighted relation digraph G =(U, E), 'U' is the node of the graph, represents a user; 'E' is the edge of the graph, represents a relationship between users, and the 'E' have it weight which refered to the mutual-following relationship. An ego-centric unit graph as shown in Figure 1. Where we proposed an simplest design, black node represents the users who have mutual followed user, gray node represents the user who unidirectional following the black user, red node represents the public account that unidirectional followed by the black user, blue node represents the user who unidirectional followed by the black user. And, in the social network there have lots of users, including all of four kinds users above, so we got the proliferated cluster graph as shown in Figure 2.

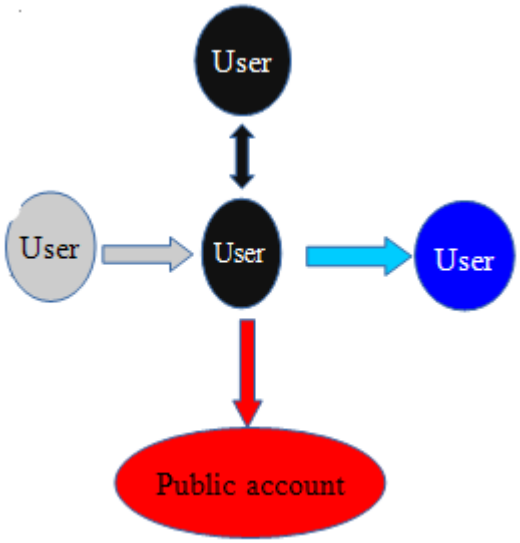

Fig.1. Ego-centric unit relation digraph

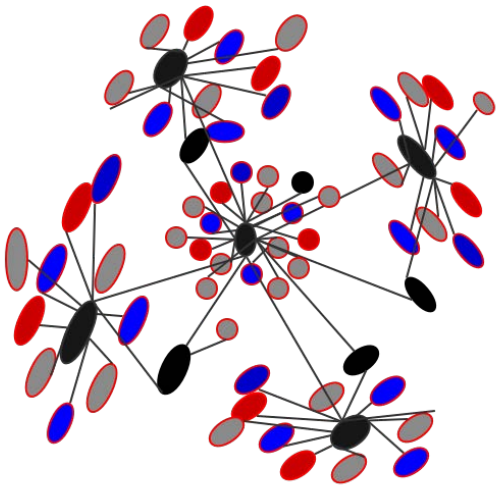

Fig.2. Ego-centric proliferated cluster digraph

Except the node, the edge is also a very important parameter. Black edge represents the relationship between black user, gray edge represents the relationship between black user and gray user, red edge represents the relationship between black user and red user, blue edge represents the relationship between black user and blue user. We know that the more friends and followers they have, the stronger the interaction between the users, the higher the influence in the social network, the higher the user's credibility. So the four kinds edges impact on the users credibility is decreasing in turn, and the specific impact would be discussed later. For given users, their Credible Values defined as:

$$U_i,\ U_i \in U \tag{1}$$

For the given users, the four kinds Edges Amount vector define as:

$$E_i = (E_{bk}(i),\ E_{gy}(i),\ E_{rd}(i),\ E_{be}(i)) = (A_i,\ B_i,\ C_i,\ D_i) \tag{2}$$

And different edge have its given weight, the Weighting Coefficient vector define as:

$$W_c = (\alpha,\beta,\gamma,\delta) \tag{3}$$

Table.1 highlights the most important parameters that we need for the algorithm.

Parameters	For short	Explanation
nodes	U	Users
U_i	U_i	Credible Values of given users
edges	E	Relationship between users
E_i	E_i	The four kinds Edges Amount vector of given users
W_c	W_c	The Weighting Coefficient vector for different edge
Black nodes	U_{bk}	The users who have mutual followed users
Gray nodes	U_{gy}	The users who unidirectional following the black users
Red nodes	U_{rd}	The public account that unidirectional followed by the black users
Blue nodes	U_{be}	The users who unidirectional followed by the black users
Black edges	E_{bk}	The relationship between black users
Gray edges	E_{gy}	The relationship between black users and gray users
Red edges	E_{rd}	The relationship between black users and red users
Blue edges	E_{be}	The relationship between black users and blue users

Tab.1. Parameters that we need for the algorithm

The user's credibility ranking can be quantitatively assessed using the Relation Graph Credibility Ranking (RGCR) algorithm, the equation for Ui is bellow:

General method: $U_i = A_i + B_i + C_i + D_i$ (4)

Coefficient condition: $U_i = \alpha * A_i + \beta * B_i + \gamma * C_i + \delta * D_i$ (5)

Ratio method: $U_i = (A_i + B_i)/(A_i + B_i + C_i + D_i)$ (6)

Coefficient condition: $U_i = (\alpha * A_i + \beta * B_i)/(\alpha * A_i + \beta * B_i + \gamma * C_i + \delta * D_i)$ (7)

product method: $U_i = A_i * B_i * C_i * D_i$ (8)

Coefficient condition: $U_i = \alpha * A_i * \beta * B_i * \gamma * C_i * \delta * D_i$ (9)

Choice method: $U_i = A_i * B_i - C_i * D_i$ (10)

Coefficient condition: $U_i = \alpha * A_i \beta * B_i - \gamma * C_i * \delta * D_i$ (11)

Which is the best method, should be evaluated via experiment, with constructed Dataset and Weighting Coefficient vector group.

Experiments

In this section, we first introduce the datasets used for experiments and then discuss the result of the eight method of RGCR algorithm separately.

Datasets. We get three datasets, they are "The Continent", "Sangfor Tournament" and "Sina Campaign" from Sina microblog API. In order to protect privacy, we do not provide the crawling details, only provide some statistical data in these datasets. "The Continent" datasets have 72,064 users, "Sangfor Tournament" datasets have 16,364 users, "Sina Campaign" datasets have 53,062 users. We get the user's following relationship between the users and crawl the top 200 blog of each user, and we filter the users whoes blogs less than 50 and followers more than 1000, for they either harmless or exemption. The information including microblog user's ID, Nickname,the number of user's four kinds relationship amount. Partial users information by example as Table.2. Based on the Sina application programming interface (API, Application Programming Interface), information collection as follows: (1) Crawl microblog seed user information. Crawl the microblog seed user information from a certain user, get the related user's

relationship amount and the user’s list, these related users as seed user for next step. (2) Get users relationship information. From the seed users, step by step crawl and record the relationship between users. (3) In the three datasets the credibility ranking of users is known, we choose 450 users for each datasets to form three test sets, the untrusted users and credible users are (171,279), (351,99), (294,156).

Tab.2. Partial users information

Number	User ID	Nickname	Ai	Bi	Ci	Di
1	1005052936749205	xiaokang sunny	46	200	245	6
2	1005053043415970	jiutiaoyiou	26	173	2419	1
3	1005052920283142	cheshize1992	126	419	1712	64
4	1005052919430892	bogaoyi1988	151	290	1829	37
5	1005053044937320	Great womenzhouyi	47	235	2429	116
6	1005052737840481	Zhangqian Lucia_	58	128	93	24
7	1005055593495067	Signal 88351	2	23	35	19

Evaluating verification with different methods. In order to facilitate the observation and comparison, we normalized the experimental results, so that the value of “U_i” range between [0-1]. In view of the fact that the Weighting Coefficient have great impact to the experimental results, we constructed a Weighting Coefficient vector group as follow: W_c1=(2,1,0.1,0.01), W_c2=(10,2,0.1,0.01). With W_c1 we get the test profiles1, as shown in figure.3, with W_c2 we get the test profiles2, as shown in figure.4.

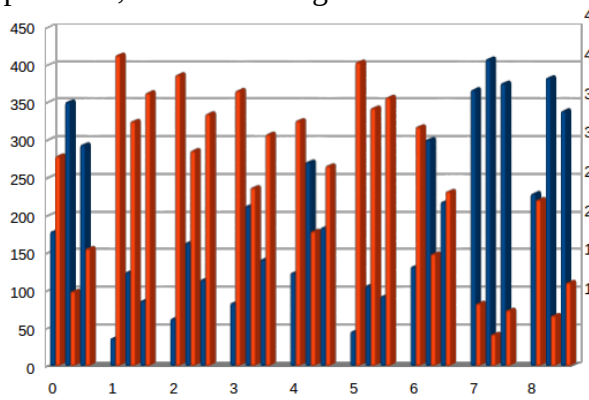

Fig.3. Profile of experimental results with W_c1

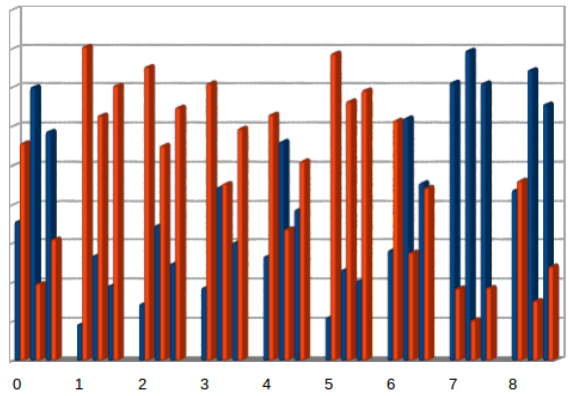

Fig.4. Profile of experimental results with W_c2

Blue icon represent the untrusted users and Red icon represent the credible users, “0” group icons represents the tested example, “1” to “8” groups represents eight method test results. Similarity of icon groups reveal the performance of algorithm. From the profiles we get that the coefficient condition method all better than its basic method, and the product method of coefficient condition is the most outstanding method of the them, ratio method of coefficient condition is secondly. Meanwhile, from Fig.4 we can see that credible users are more depending on its A_i than B_i, If we choose a adapt Weighting Coefficient vector, the RGCR algorithm will get a better performance.

Conclusion

In this paper, we build a weighted graph G =(U, E) and design a Relation Graph Credibility Ranking (RGCR) algorithm, use eight methods to analysis the credibility ranking of users in social network. In experimental process, we perform experiments on three datasets crawled from the Sina microblog with a constructed Weighting Coefficient vector group. The comparison of results on average credibility of users in each set and each method demonstrates that the RGCR algorithm is

an efficient credibility ranking algorithm. In a nutshell, the results of the experiments show that our proposed 4th and 6th methods works better than other six methods to rank the credibility of users in the social network.

Acknowledgement

The authors acknowledge the financial support by State Key Development Program of Basic Research of China (No.2013CB329604), the research was sponsored by NSFC(Grant No.61672020,61662069,61472433), Project funded by China Postdoctoral Science Foundation:(2013M542560,2015T81129), A Project of Shandong Province Higher Educational Science and Technology Program(No.J16LN61), project of Yantai science and technology program(2016ZH054).

References

[1] Ceolin D, Groth P, Hage W R V, et al. Trust Evaluation through User Reputation and Provenance Analysis[C]// International Workshop on Uncertainty Reasoning for the Semantic Web. 2012.

[2] Chen B C, Guo J, Tseng B, et al. User reputation in a comment rating environment[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, Ca, Usa, August. DBLP, 2011:159-167.

[3] Li B, Li R H, King I, et al. A topic-biased user reputation model in rating systems[J]. Knowledge & Information Systems, 2015, 44(3):581-607.

[4] Wang Z, Tan Y, Zhang M. Graph-Based Recommendation on Social Networks[C]// Advances in Web Technologies and Applications, Proceedings of the, Asia-Pacific Web Conference, APWEB 2010, Busan, Korea, 6-8 April. DBLP, 2010:116-122.

[5] Briscoe E J, Appling D S, Mappus R L, et al. Determining credibility from social network structure[C]// Ieee/acm International Conference on Advances in Social Networks Analysis and Mining. ACM, 2013:1418-1424.

[6] Al-Qurishi M, Aldrees R, Alrubaian M, et al. A New Model for Classifying Social Media Users According to Their Behaviors[C]// Web Applications and NETWORKING. 2015:1-5.

[7] Li J, Ma S, Hong S. Recommendation on social network based on graph model[C]// Control Conference. 2012:7548-7551.

[8] Nie Y, Jia Y, Li S, et al. Identifying users across social networks based on dynamic core interests[J]. Neurocomputing, 2016, 210:107-115.

[9] Wei W, King I. Measuring credibility of users in an E-learning social network[C]// International Conference on Computer Science and Education. IEEE, 2010:418-424.

[10] Gao Y, Li Y, Zhang R. Multi-dimensional User Credibility Analysis on Review Content[M]// Web Technologies and Applications. Springer International Publishing, 2015.

[11] Alrubaian M, Al‐Qurishi M, Al‐Rakhami M, et al. Reputation‐based credibility analysis of Twitter social network users[J]. Concurrency & Computation Practice & Experience, 2016.

[12] Lin L, Xu L, Zhou S, et al. The Social Feature-Based Priority Relation Graph of Mobile Social Networks[C]// IEEE, International Conference on Computational Science and Engineering. IEEE Computer Society, 2014:1921-1926.

[13] Nie Y, Huang J, Li A, et al. Identifying Users Based on Behavioral-Modeling across Social Media Sites[M]// Web Technologies and Applications. 2014:48-55.

[14] Wang X, Yu X, Zhou B, et al. Mining Personal Interests of Microbloggers Based on Free Tags in SINA Weibo[M]// Web-Age Information Management. Springer International Publishing, 2015.