

Identify the Microblogging Opinion Leaders Based on HITS-IMP Algorithm

Chengle Yu, Bingxiang Liu and Xing Xu

*School of Information Engineering, Jingdezhen Ceramic Institute, Jiangxi, China,
333403
whuxx84@aliyun.com*

Abstract

People who actively provide people with their own subjective comment on information in the process of microblogging transmission, and can influence others thought to a certain action, are named as the microblogging opinion leaders. Therefore it's very important to identify microblogging opinion leaders. In this paper, the HITS algorithm is improved based on micro-blog user's influence and liveness analysis. According to the weight calculation of micro-blog user hierarchy analysis, the micro-blog user comment, forwarding behavior; we construct the interactive network to identify micro-blog opinion leaders based on the user's weight micro-blog. The improved HITS-IMP algorithm can identify opinion leaders effectively.

Keywords: *micro blogging opinion leaders, Interactive network, The weight, HITS-IMP algorithm*

1. Introduction

Weibo is short for micro-blogging. The earliest and most famous microblogging is the Twitter of the United States, which is as an Internet platform. In Weibo, participants are free to edit content you want to tell others and the followers can view the message and comments in time. Thus a kind of free, autonomous, interactive, simple and rapid way of information dissemination is formed. Opinion leaders' thoughts originally come from the "public opinion" walter lippmann. The formal concept formed in the 1944 raza Mansfield's "people's choice", referring to those who have special influence in information dissemination. They have the power to change individual or group thought and behavior.

With the rapid development of Internet technology in China, weibo has penetrated into our daily life and becomes an indispensable platform where we acquire and exchange information. According to CNNIC (China Internet network information center) statistics, until December 2013, the scale of Internet users in China is near 618 million among which 281 million is weibo users. Microblogging opinion leaders who have a strong excitation force can produce a powerful social mobilization. To prevent the negative results of the dangers of the inappropriate words and deeds to the state and society, the government shall reasonably regulate the microblogging opinion leader's words and deeds to create a harmonious and healthy environment for weibo. Therefore, how to define and identify the opinion leaders is an important problem that is worth studying.

2. Related Work

In recent years, many algorithms have been developed to identify the opinion leaders in a virtual network, which can be roughly divided into the static information statistics method and information network structure method.

(1) Static information statistics method: The method is based on the basic

characteristics of opinion leaders, through the establishment of the corresponding index system. The users make use of good data sample clustering analysis and the same category of the individual as far as possible with high homogeneity and high heterogeneity between categories to screen out the opinion leaders. Simple and intuitive clustering analysis is widely used in current practice. Xue and Chen (2010) use Shanghai jiaotong university campus BBS "jiaotong university development" forum "threads" and "post ID" as the research object. From the activity attention, the three dimensions are esteemed by clustering analysis method using ID postings, regeneration diffusion, and four indicators, to extract the opinion leaders in BBS. But such static statistical analysis ignores the dynamic interaction behavior of information exchange between users.

(2) The network structure analysis: Network analysis is a kind of quantitative analysis method based on social relations between the users to build network, and then the calculation of extroverted degree of member's centrality, introverted centrality and intermediary index are used to calculate the importance of the user for identifying the opinion leaders. Yu, *et al.*, analyzed characteristics of opinion leaders in "online community" via leader Rank algorithm based on emotional weight. But these methods from a user's interaction network structure to identify the opinion leaders ignore the user's own characteristics.

Based on the above elaboration, the users own characteristics, network structure, analysis of the interaction between users are combined to identify microblogging opinion leaders in this paper. First, we analyze the influence and active degree of user characteristics and calculate the user feature weights by AHP (analytic hierarchy process) method. According to user's forward, and comments on weibo, user interaction network is built. Based on the user character and the user's authority value, HITS-IMP algorithm is used to calculate the weight to identify the microblogging opinion leaders, which effectively improves the accuracy of the identification of opinion leaders.

3. HITS - IMP Algorithm for Feature Weighting

3.1. Analysis of Characteristics of the User

3.1.1. The Influence by the User: Users refer to the people who own public opinion influence on other users in microblogging platform. In this paper three characteristics are mainly discussed: microblogging certification, weibo comments and weibo forwarded number. You need to provide valid id card and the corresponding certification conditions for Weibo certification. So the authentication user comments are more authenticity. The more deeply forwarded, the more broad impact is accounted. So we can use the following expression definition:

$$R_1 = w_1 E_1 + w_2 E_2 + w_3 E_3 \quad \textcircled{1}$$

Among them, the R_1 represents user influence, E_1 weibo user authentication, E_2 weibo comments, E_3 weibo is forwarded, w_1 , w_2 , w_3 are the coefficients of the different characteristics.

3.1.2. The User Activity: Due to the existence of some high powerful but actually low tweeting frequency users who do not affect other users of comprehensively the identification of opinion leaders need to consider the user's activity. In weibo, starting from the original number of microblogging and reply, the two features are considered. Aiming at an event, opinion leaders should have their own ideas. The original weibo spreads their views extensively. The more responds to the comments of weibo around other topics to communicate and reply, the users pay more attention to the comments by others. Therefore, the following expression are defined:

$$R_2 = w_4 E_4 + w_5 E_5 \quad \textcircled{2}$$

Among them, R_2 represents user activity, E_4 originality of weibo, E_5 reply of weibo, w_4 ,

w_5 are coefficients of different characteristics.

3.2. User Quantitative Analysis

AHP (Analytic Hierarchy Process) is the elements related to the decision, which is always broken down into objectives, principles, scheme, such as level, on the basis of the decision-making methods of qualitative and quantitative analysis.

3.2.1. Create a Hierarchical Structure: Using AHP (analytic hierarchy process), according to analysis of the characteristics of weibo users, we established the hierarchy structure of weibo user's feature weights:

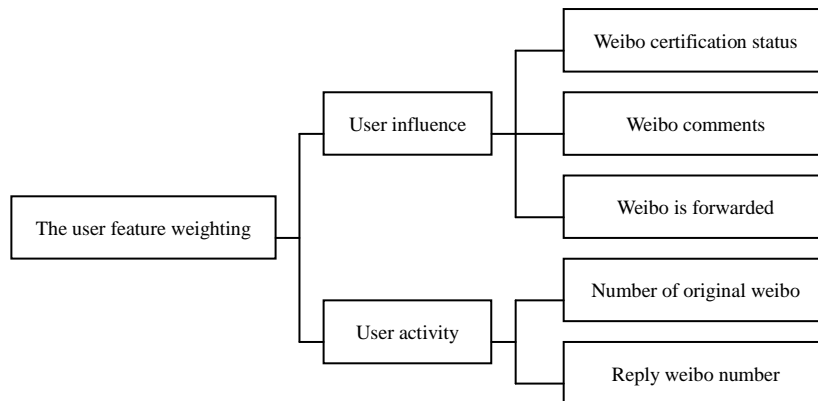


Figure 1. User Feature Weighting the Layer of the Structure

Weibo user feature weighting summarizes the influence of the user activity and the calculation formula is:

$$W=R_1+R_2 \quad (3)$$

3.2.2. The Characteristic Value of Relative Weight: Due to the characteristics of the value cannot be measured by the numbers, the value of each feature normalization is to facilitate the weights by using the following formula to map to interval [0, 1]:

$$Y = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (4)$$

Where X represents characteristic value, Y features. Since the importance of each factor is different, in this paper we use 1, 2, 3, 4, 5, 6, 7 to scale the importance of different factors between the intermediate state by 2, 4, 6, and legally obtained by using the characteristic values of relative weight and consistency inspection, the results are as follows:

Table 1. Characteristics Coefficients and the Consistency Check

Characteristics of the indicators	index property
Weibo certification status	0.23
Weibo comments	0.23
Weibo is forwarded	0.28
Number of original weibo	0.15
Reply weibo number	0.11

Consistency check: CR = 0.018 < 0.1

4. Microblogging Opinion Leaders Recognize

HITS algorithm was proposed by Dr .Jon Mr. Kleinberg in 1997, which is a web page importance analysis of the algorithm. The basic idea is to use the page references between chains to mining implied useful information such as authority. HITS algorithm has the characteristics of simple calculation and high efficiency. HITS algorithm should separate the authority and its content link authority on the web page content and the authority is evaluated on the basis of the links on the page against authority evaluation, then it gives the page comprehensive evaluation. Content authority and the page which directly provide information related to the quality of the content that was quoted by the web page. The higher the content of its authority is, the higher authority associates with the quality of web page.

HITS algorithm, introduced in this article, mines the opinion leaders in ordinary weibo users. The authority of opinion leaders creates values, as well as the information center value of middlemen, and gets the user authority value and center of final order. Because the user's own feature weights were added to the algorithm, the improved algorithm called HITS-IMP algorithm in this paper.

4.1 Build User Interaction Network

Weibo users between the main communication through forward and comment on weibo, based on the two kinds of behavior get set for edge M, with the user for the node set N, build user interaction network, user node feature weighting value calculated by the formula ③, and the weight value of interaction by the user's emotional tendency to determine.

4.2 Emotional Interaction Behavior Analysis

This article uses the IK Analyzer 3.0 to complete the word segmentation and then uses SVM (support vector machine) algorithm. The micro blog includes these three classes of positive, negative and neutral emotional value, respectively by 1、 -1、 0. Statistically the aggregation is greater than zero; emotional words used in the user reviews are identified as positive comments, less than zero as negative comments. So we can use the following formula:

$$m_{pq}=a+b \quad \text{⑤}$$

The m_{pq} represents user p of emotion tendency for user q, a user p of the number of positive evaluation to user q, b user p for user q of number of negative evaluation.

4.3. HITS-IMP Parsing Algorithm

Weighted HITS-IMP algorithm is based on user characteristics. After each iteration value and the center for authority to be normalized, we can get the final formula in the following:

$$\text{Authority value : } s(p)_k = \frac{\sum_{q:(q,p) \in M} t(q)_{k-1} * w_q * m_{qp}}{\sqrt{\sum_{x:x \in N} [\sum_{x:(x,y) \in M} t(y)_{k-1} * w_y * m_{yx}]^2}} \quad \text{⑥}$$

$$\text{Center value : } t(p)_k = \frac{\sum_{q:(q,p) \in M} s(q)_{k-1} * w_q * m_{qp}}{\sqrt{\sum_{x:x \in N} [\sum_{y:(y,x) \in M} s(y)_{k-1} * w_y * m_{yx}]^2}} \quad \text{⑦}$$

5. Experimental Results and Analysis

5.1. Experimental Data

This paper is based on a web crawler tool (SoukeyMiner) and use "electricity" as keywords, grabbed from March 1, 2013 to May 1, 2013, nearly thousands sina weibo micro. Through regular expression parsing, involves users, fetching user information are extracted as the experimental data.

5.2. Result of the Experiment and Analysis

HITS-IMP algorithm is used in our experiments. We set the number of iterations 20. Figure 2 and Figure 3 show jack ma, lee two weibo user authority value and the center of convergence respectively.

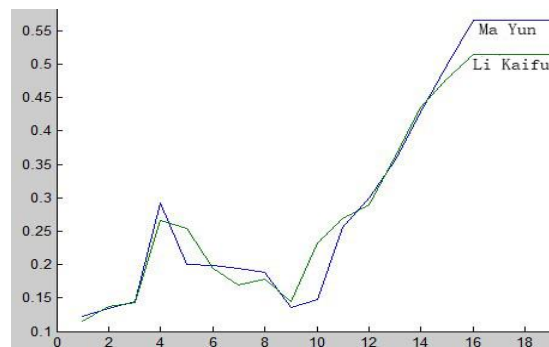


Figure 2. Authority Value of Convergence

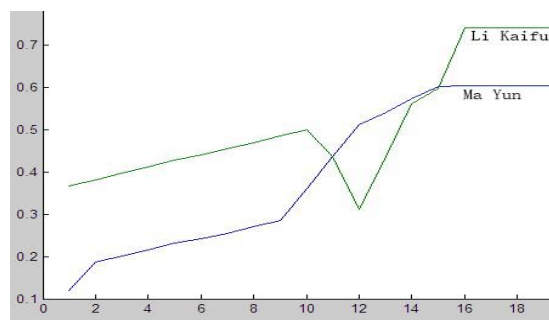


Figure 3. Center Value of Convergence

Above the two figures, we can see that HITS – IMP algorithm converges from the 16th iteration. In order to ensure accuracy, this paper only select authority value and the center are the first top five users as opinion leaders, which are ranked as follows:

Table 2. Before Five Values the User Authority

rank	user name	Authority Scores
1	Ma Yun	0.566
2	Li Kaifu	0.514
3	Chen Nian	0.498
4	Li Guoqing	0.484
5	Liu Qiangdong	0.437

Table 3. Center Top Five Values to the User

	rank	user name	central value
	1	Li Kaifu	0.741
	2	Chen Nian	0.622
	3	Ma Yun	0.604
	4	Liu Qiangdong	0.58
	5	Li Guoqing	0.435

First of all, a true opinion leader must have the very high prestige and influence; secondly, inside the circle he is very active and plays the role of the information hub. This paper requires that user must have the authority of the high value and the center can be identified as opinion leaders. We can see from Table 2 and Table 3, the HITS - IMP algorithm concluded that the five people in the center of the authority value and value are ranked in the top. So we can assume they are opinion leaders in this field. Also it suggests that the real opinion leader indeed owns these two characteristics: "influential" and "information hub".

6. Conclusion

Through the improved HITS algorithm, this paper analyzed the characteristics of the user interaction and emotional tendency with the process of "electricity" as topic at the same time. By using the analytic hierarchy process, the Chinese word segmentation technology and emotional analysis of SVM (support vector machine) algorithm, we explored the relevant user of the sina weibo. The results show that the opinion leaders are identified more accurately. Also we found that the opinion leaders are usually influential information intermediaries through microblogging certification.

Acknowledgements

This paper was supported by the National Science and Technology Support Plan (2013BAF02B01), State Key Laboratory of Software Engineering (SKLSE2012-09-35), the National Natural Science Foundation of Jiangxi Province (20122BAB211036, 20122BAB201044, 20142BAB217024), the Science Foundation of Jiangxi Provincial Department of Education(GJJ14639), the National Nature Science Foundation of China (61202313), Students Innovation and Entrepreneurship Education Project of Jiangxi Province.

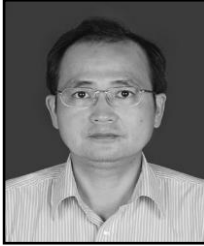
References

- [1] L. Jing, "Based on the characteristics and the user interaction network of microblogging opinion leaders recognize", Southeast University, (2014).
- [2] L. Zhirming, "Microblogging opinion leaders identified within the network public opinion and analysis", Journal of systems engineering, vol. 29, no. 6, (2011), pp. 8-16.
- [3] W. Yu and Z. Jianping, "BBS based on clustering analysis of network opinion leaders find method", Computer engineering, vol. 5, no. 5, (2011), pp. 44-49.
- [4] W. Yin and W. Xinling, "Chinese weibo sentiment analysis method research", Guangdong technology normal university, (2014).
- [5] J. Weng, E. P. Lim and J. Jiang, "Twitterrank: finding topic-sensitive influential witters", Proceedings of the third ACM international conference on Web search and data mining. ACM, (2010), pp. 261-270
- [6] X. Yue and H. Tingzhu, "Web page sorting of PageRank algorithm and HITS algorithm research", University of electronic science and technology, (2014).

Author



Chengle Yu, Born in 1988 Received his Bachelor degree in Jingdezhen Ceramic Institute. Currently, he is a Master graduate student in Jingdezhen Ceramic Institute Information Engineering Institute. His main research interests include data mining and swarm intelligence algorithm.



Bingxiang Liu, Born in 1966 Received his PhD degree in Southeast University, Professor in Jingdezhen Ceramic Institute Information Engineering Institute since 2002. His main research interests include data mining and knowledge engineer.



Xing Xu, Born in 1984 Received his PhD degree in Wuhan University in 2010. Now He is the Associated Professor in Jingdezhen Ceramic Institute since 2013. His main research interests include data mining and evolutionary computation.

