Study on the Influential Factors of Micro-blog Forwarding

Keliang JIA[1,a] Zhinuo Li[2]

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

Abstract—The secondary forwarding of micro-blog can realize the fission transmission of micro-blog information, so the researches on Influential Factors of micro-blog retweeting has become a research hotspot of current scholars. Based on Lasswell's 5W model, this paper comprehensively analyzes the factors influencing micro-blog forwarding from the standpoint of disseminator, content, channel and audience, and puts forward 16 research hypotheses, then constructs an influential factors model of micro-blog retweeting, and finally verifies the hypothesis based on binary logistic regression analysis model. Data analysis results show that the count of the fans of a disseminator, count of micro-blogs of a disseminator, mean count of micro-blogs being forwarded of a disseminator,   mean count of micro-blogs being commented and thumbed of a disseminator, length of micro-blog, popularity of micro-blog, followees count of an audience, activity of audience retweeting, an audience's retweeting concern to his followee show a positive effect on the behavior of retweeting, while richness of micro-blog, emotion words, count of fans of the audience have a negative impact on the retweeting behavior.

1 Introduction

With the mushroom growth of social platforms, social applications have become a significant part of Internet applications. Micro-blog is a kind of social application which network structure is on the basis of users' concerned relationships. It adopts the information push-retweet mechanism. The micro-blog published by a user can be tweeted to his followers. Through the retweeting of his followers, the information is pushed to the fans of his followers. The secondary retweeting realizes the fission transmission of micro-blog information [1]. Therefore, the study on influential factors of micro-blog retweeting has become a research hotspot of current scholars. The research results can be applied in the following fields: micro-blog public opinion monitoring, rumor control, enterprise product marketing, hot spot mining, government micro-blog policy promotion.

2 RELATED WORK

Scholars have studied the micro-blog retweeting behavior on Twitter and Sina Weibo platform. Zhang et al. [2] selected information topic category, text length, whether to add V, whether to have a profile, count of users' followers and count of micro-blogs to analyze the impact of these factors on micro-blog retweeting by using logistic regression method. David R. Bild et al. [3] found that the types of the information, sources of the information, objectivity of the information, people and places mentioned in the twitter platform affect the information

retweeting. Based on SVM method, Li Zhiqing [4] selected micro-blog features, users' characteristics and micro-blog topic features to study the influential factors of micro-blog retweeting, which shows that topic features can significantly enhance the precision of micro-blog forwarding prediction. Zhao Rongying et al.[5] found that count of followers of a user is the basis of micro-blog retweeting. Guo Ya et al. [6] found that micro-blog author information, user interest and micro-blog popularity are important factors affecting user's retweeting behavior. Tang Xiaobo et al. [7] introduced emotional differences and users' interest features to realize micro-blog forwarding prediction based on SVM, which improves the effect of its retweeting prediction on Sina Weibo dataset. Research by Zhang et al. [8] shows that the local social network can produce an effect on the forwarding behavior of fans. Wang Shaoqing et al. [9] constructed multiple trust relationships on the basis of the social network relationship between micro-blog users to forecast users' forwarding behavior, and achieved better performance.

The users in micro-blog platform are not only the publisher of micro-blog information, but also the receiver and retweeter of micro-blog information. There are few studies on distinguishing the characteristics of the two kinds of roles of micro-blog users, and less comprehensive consideration of interfering determiners of micro-blog retweeting from the standpoint of the whole communication process. Based on Lasswell's 5W model. From the perspective of the whole communication process, this paper analyzes the influence of the four comprehensive factors of "who is saying", "saying what", "communication channel" and "to whom" on micro-blog information dissemination effect (i.e. whether the

[a]sdjiakeliang @qq.com

information is retweeted or not), puts forward hypotheses and verifies the hypotheses through empirical analysis.

3 Research hypothesis and theoretical model

According to the 5W communication model[10], a complete communication process consists of five elements: "who is saying", "saying what", "communication channel", "to whom", and "with what effect". The 5W communication model reveals the elements of information transmission and the links of information transmission from five dimensions: communicator, content, channel, audience and effect of communication.

3.1 Influence of publishers on micro-blog retweeting

In the 5W model, "communicator" is the first communication element. In micro-blog platform, a blogger acts as the "communicator", and the number of fans represents the audience coverage, which is the basis of micro-blog forwarding. Tags of a blogger indicates his content coverage, and count of the posted micro-blogs indicates his publishing activity. Social psychology researches show that people have conformity psychology, and users of micro-blog platform are also vulnerable to the influence of conformity psychology and forward the micro-blogs that has been forwarded in large quantities [11]. Therefore, the number of tweets posted by bloggers that are forwarded, commented and liked will also affect their fans' retweeting behavior. Therefore, this paper assumes that:

- H1.1 The number of fans of a blogger (X1) affects the retweeting of his micro-blogs.

- H1.2 The number of blogger's tags (X2) affects the retweeting of his micro-blogs.

- H1.3 The number of blogger's micro-blogs (X3) affects the retweeting of his micro-blogs.

- H1.4 The average count of a blogger's micro-blogs being forwarded (X4) affects the retweeting of his micro-blogs.

- H1.5 The average count of comments on a blogger's micro-blogs (X5) affects the retweeting of the micro-blogs.

- H1.6 The average count of likes on a blogger's micro-blogs (X6) affects the retweeting of the micro-blogs.

3.2 Influence of content on micro-blog retweeting

As the carrier of information dissemination, micro-blog with different contents will lead to different communication effects. Compared with short text micro-blog, people prefer to get complete information. The longer the length of micro-blog, the more complete the information is, and it is easier to attract users. The emotional difference of micro-blog content[7] will cause

different emotional resonance of readers. Whether Micro-blog content information refers to other users or not, whether it contains links or not, whether it contains pictures and videos or not, whether it involves topics and other rich forms of expression or not, which are also easy to attract users' attention. In micro-blog social network, popular topics are more likely to arouse users' interest and attention. Therefore, this paper assumes that:

- H2.1 The length of micro-blog (X7) affects its retweeting.

- H2.2 The emotion of micro-blog (X8) affects its retweeting.

- H2.3 The rich information forms of micro-blog (X9) affect its retweeting.

- H2.4 The popularity of micro-blog content (X10) affects its retweeting.

3.3 Influence of social media on micro-blog retweeting

Compared with traditional media, micro-blog media platform shows more interactivity and complexity. A micro-blog user's acceptance of the information published by different blogger is not consistent, and shows different retweeting attention to different bloggers. Based on the perspective of social network analysis [12], a user directly links with many other users in social network. If a user has an outstanding degree of centrality, he will play a significant role in information diffusion. Count of fans and count of followees of a micro-blog user are the main measure of the centrality in the social network. Therefore, this paper assumes that:

- H3.1 Count of a user's followees (X11) affects his retweeting behavior.

- H3.2 Count of a user's fans (X12) affects his retweeting behavior.

- H3.3 Users' retweeting attention to a blogger (X13) affects their behaviors of retweeting the blogger's micro-blogs.

3.4 Influence of receiver on micro-blog retweeting

From the perspective of receivers, it is found that some micro-blog fans are accustomed to forwarding the information published by the followees, which is beneficial for the high-speed dissemination of the information, while some other fans seldom forward, so their impact on dissemination of the micro-blog can be ignored. The users' historical behaviors pattern on micro-blog platform has certain influence on users' retweeting behaviors, and users' retweeting behaviors are driven by users' interests [10]. Therefore, this paper assumes that:

- H4.1 Users' retweeting activity (X14) affects their retweeting behaviors.

- H4.2 Users' platform activity (X15) affects their retweeting behaviors.

- H4.3 The consistency of micro-blog information and users' interests (X16) affects users' retweeting behaviors.

3.5 Model

In accordance with the above hypotheses, an influential factors model of micro-blog retweeting based on 5W propagation model is constructed, as shown in Figure 1.
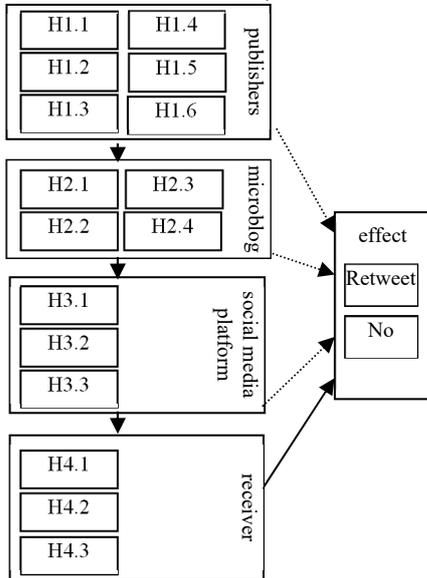

Figure 1. Influential factors model of micro-blog retweeting based on 5W propagation model

4 Logistic regression and result analysis

4.1 Samples data

We use network spider to grab micro-blogs from Sina Weibo. For the retweeting sample data, the micro-blog forwarded by the fans is extracted as the forwarding samples data. For non retweeting samples data, the method in reference [13] is used to extract micro-blog data that users see but not forward as non retweeting samples data. Finally, 194395 samples are obtained, including 107304 positive samples and 87091 negative samples. Using the method in reference [14], we calculate the characteristics variables of micro-blog publishers, micro-blogs, micro-blog communication platform and micro-blog receivers.

4.2 Logistic regression

Micro-blog retweeting behavior is a two category problem. In this paper, we take whether to forward as a dependent variable, the characteristics influencing micro-blog retweeting as independent variables, and then use binary logistic regression analysis in SPSS software to verify the factors affecting micro-blog retweeting. After further screening the samples data, the samples 15134 are obtained after removing the samples with missing values, including 8171 retweeting samples and 6963 non retweeting samples. For the average count of comments (X5) and the average count of likes (X6) on micro-blog

with colinearity, the two variables are combined into one variable (X56), and the value of the new variable is the average of the original two variables. The results of logistic regression based on the new samples are shown as following tables.

Results of significance test of the overall model coefficients in TABLE 1 show that the chi square of the global goodness of the model constructed by the 15 variables in this paper is 19359, which is greater than the critical value of 25, and the corresponding p value is less than 0.05, which meets the level of significance. The test indicates that at least one of the 15 independent variables can effectively explain and predict the results of micro-blog retweeting.

Table1. omnibus test

		Chi square	df	Sig.
Step 1	Step	19359.236	15	.000
	Module	19359.236	15	.000
	Model	19359.236	15	.000

Table2. model summary

Step	-2 Log Likelihood	Cox & Snell R^2	Nagelkerke R^2
1	1524.417[a]	.722	.964

a. The estimation is terminated at iteration 9 because the change range of parameter estimation is less than 0.001.

In TABLE2, the model summary table shows the correlation strength test between 15 independent variables and dependent variables. The maximum likelihood square logarithm is greater than the critical value of 25, the goodness of fit of Cox & Snell R^2 is 0.722, and the goodness of fit of Nagelkerke R^2 is 0.964, which indicates that the dependent variable has a powerful association with these 15 independent variables.

Table3. Classification results

			predicted value		
	Observation value		re_result		Accuracy (%)
			0	*1*	
Step 1	re_result	0	6851	112	98.4
		1	137	8034	98.3
	Overall Accuracy(%)				98.4

a. The split value is .500

In TABLE3, there are 6963 non retweeting micro-blogs in the samples data. According to the logistic regression prediction results, 6851 micro-blogs are classified as not forwarded (predicted correctly), 112 are classified as forwarded (predicted wrong); among 8171 forwarded micro-blogs, 8034 are predicted to be forwarded (predicted correctly), and 137 are predicted to be not forwarded (predicted wrong). The accuracy of the overall classification prediction reaches 98.4%. It means that the established logistic regression prediction model has a high overall goodness of fit, and the independent variables have a great influence on the dependent variable.

Based on omnibus test, model summary test, and classification prediction result test, we believe that the logistic regression model in this paper has higher goodness of fit.

Table4. variables in the equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1ᵃ	X9	-.288	.128	5.083	1	.024	.750
	X7	.084	.006	199.852	1	.000	1.088
	X8	-.386	.076	25.864	1	.000	.680
	X10	.766	.166	21.373	1	.000	2.151
	X2	.028	.022	1.626	1	.202	1.028
	X1	.598	.116	26.444	1	.000	1.818
	X3	2.143	.144	222.010	1	.000	8.525
	X56	1.593	.152	110.105	1	.000	4.918
	X4	1.177	.138	72.829	1	.000	3.246
	X12	-.406	.058	49.135	1	.000	.666
	X11	.869	.102	72.351	1	.000	2.384
	X13	1.012	.292	12.017	1	.001	2.750
	X14	1.437	.275	27.257	1	.000	4.207
	X15	-.168	.568	.088	1	.767	.845
	X16	1.172	.738	2.525	1	.112	3.230
	Constant	-21.412	1.034	428.672	1	.000	.000

4.3 Result analysis

In TABLE4, for the characteristics of the publisher, the number of publishers' tags has no significant effect on micro-blog retweeting ($P > 0.05$), so hypothesis H1.2 is false. Count of fans, count of micro-blogs, mean count of publisher's micro-blogs being forwarded and mean count of likes and comments on the micro-blogs have significant impact on micro-blog retweeting ($P < 0.05$). Therefore, hypotheses H1.1, H1.3, H1.4, H1.5, H1.6 are true.

Among the characteristics of the micro-blog, the rich forms of content, length, emotion words and popularity of micro-blog have significant effects on micro-blog retweeting ($P < 0.05$). Therefore, hypotheses H2.1, H2.2, H2.3, H2.4 are true.

Among the characteristic variables of micro-blog platform, network centrality of the receiver, that is, count of his fans and count of his followees significantly affect the micro-blog retweeting ($P < 0.05$). Therefore, hypotheses H3.1 and H3.2 are true. The retweeting attention of the receiver to his followees significantly affects micro-blog retweeting ($P < 0.05$), so H3.3 is true.

Among the receiver characteristic variables, the receiver's platform activity and retweeting interest consistency have an impact on micro-blog retweeting, but the statistical results aren't remarkable ($P > 0.05$), so the hypotheses H4.2 and H4.3 are false; the receiver's retweeting activity significantly affects micro-blog retweeting ($P < 0.05$), hypothesis H4.1 is true.

5 Conclusion

From the perspective of Lasswell's 5W model, this paper constructs an influential factors model of micro-blog retweeting. After crawling Sina micro-blogs, the influential factors of micro-blog forwarding are analyzed by logistic regression model. The results of logistic regression analysis show that count of fans of a blogger, count of blogger's micro-blogs, mean count of a blogger's micro-blogs being forwarded and mean count of comments and likes on a blogger's micro-blogs have a remarkable positive influence on micro-blog retweeting. Length and popularity of micro-blog have significant positive influence on micro-blog retweeting, while the richness of content and emotion words in micro-blog have a remarkable negative influence on micro-blog retweeting. To a retweeter, count of his followees and his retweeting attention to his followees have remarkable positive influence on his retweeting behavior, while count of his fans has a significant negative impact on micro-blog retweeting. Retweeting activity of a receiver has a remarkable positive influence on his retweeting behavior. The work provides a novel research viewpoint of the research on micro-blog retweeting information dissemination, and has certain reference significance for micro-blog retweeting prediction, network public opinion control, network environment purification and enterprise micro-blog marketing.

Acknowledgment

References

1. Zhang Sai, Xu Ke, Li Haitao. "Measurement and analysis of information propagation in online social networks like microblog," Journal of Xi'an Jiaotong University, 2013,(2):124-130.

2. Zhang L,Peng T Q,Zhang Y P,et al. "Content or Context:Which Matters More In Information Processing on Microblogging Sites," Computers In Human Behavior,2014,31:242-249.

3. David R. Bild, Yue Liu, Robert P. Dick．"Aggregate characterization of user behavior in Twitter and analysis of the retweet graph," ACM Transactions on Internet Technology(TOIT),2015,15(1): 4.1-4.24.

4. Li Zhiqing, "Predicting Retweeting Behavior Based on LDA Topic Features," JOU Ｒ NAL OF INTELLIGENCE，2015,34(9): 158-162.

5. Zhao Rongying, Zeng Xianqin, "Research and Analysis on Influential Factors of micro-blog information dissemination," Information theory and Practice,2014,37(3): 58-63.

6. Guo Ya, Gong Yeyun, Zhang Qi, Huang Xuanjing, "Retweet behavior prediction using topic model," Journal of Chinese Information Processing,2018,32(4): 130-136.

7. Tang Xiaobo, Luo Yingli, "Integrating Emotional Divergence and User Interests into the Prediction of Micro-blog Retweeting," Library and Information Service, 2017,61(9):102-110.

8. Zhang J, Tang J, Li J, etal. "Who influenced you? predicting retweet via social influence locality, "

ACM Transactions on Knowledge Discovery from Data, 2015,9(3):1-26．

9. WANG Shaoqing,LI Cuiping,WANG Zheng,CHEN Hong, "Prediction of retweet behavior based on multiple trust relationships," J Tsinghua Univ (Sci& Technol),2019,59(4):270-275.

10. Harold Lasswell. The structure and function of social communication (translated by Zhan Jiang, He Daokuan), Beijing: Communication University of China Press, 2013，35-36.

11. Kim E,Sung Y,Kang H. "Brand Followers' Retweeting Behavior on Twitter:How Brand Relationships Influence Brand Electronic Word-Of-Mouth," Computers In Human Behavior,2014,37: 18-25.

12. Reza Zafarani,Mohammas Ali Abbasi,Huan Liu, "Social media mining an introduction [M] (translated by Liu Ting, Qin Bing, Zhao Yanyan). Beijing: Posts & Telecom Press, 2015, 40-41

13. Cao Jiuxin, Wu Jianglin, Shi Wei, Liu Bo，et al．"Sina Micro-blog Information Diffusion Analysis and Prediction，" Chinese Journal of Computers.2014,37(4): 779-790.

14. Jia Keliang, Zhang Xiaotian，"Micro-blog Retweeting Prediction Based on Combined-Features and Random Forest," Communications in Computer and Information Science(CCIS), 2019,v 1042, p429-440.