Recent Researches on Application and Influence of Twitter

Hye-Jin Kim and Yong-Hyuk Kim[*]

*Department of Computer Science and Engineering,
Kwangwoon University, Seoul, South Korea.*

*Corresponding author

Abstract

We introduce and summarize the researches that are categorized into parts on Twitter studied for the past five years. The circulators in Twitter can be divided into two types, mass media source and evangelist, except normal users. If a certain leader has a huge number of retweets, he/she may be seen as an influential leader. The less difference between the number of 'following' and 'follower' accounts there is, the more credible are tweets in Twitter. Researchers have distinguished spam from tweets based on contents, and have detected spammer and tweets by analyzing the behavior of Twitter accounts. Twitter has been applied variously on marketing, education, etc.

Keywords: Twitter, Twitter application, Twitter influence, information diffusion

INTRODUCTION

Twitter is one of the most popular services that began in March 2006. We can receive concerned information through 'following'. Owing to such convenience, the number of new subscribers on Twitter is increasing day by day and has been used to give a variety of applications and a great impact on individuals, organizations, and societies. The number of papers during the last five years from Google Scholar, which include a word 'Twitter' in the title, is approximately 1,150,000. Park et al. [25] summarized studies in the field of information retrieval by 2010, we piece the rest together. The organization of this paper is in the following. Section 2 presents researches on diffusion and flow of Twitter. We can know the characteristics and kinds of circulators through the studies. We can confirm why Twitter has been used in information diffusion. Section 3 introduces studies on determining reliability of tweet contents. Section 4 deals with about social influence of Twitter. Section 5 introduces various studies of detecting spam based on relationship and contents. Section 6 presents studies that show Twitter used in a variety of applications. Finally, we draw our conclusion in Section 7.

INFORMATION DIFFUSION

Twitter is a social network service characterizing that it has a huge diffusion effect. For example, president Obama's account (@BarackObama) has about 70 million followers, of which the number is close to the population of France. If that account posts a tweet, it spreads to about 70 million users immediately. Furthermore, we can expect Twitter's diffusion effect. If we follow an account related to the subject that we are interested in, we can get information quickly and easily.

Topics for major/minor international news headlines are spreading to various users via Twitter. In relation to the connection of information, we can divide circulators into mass media source, grassroots, and evangelist [11]. 'Mass media source' that plays an important role to transmit to public key topics includes e.g., '@nbc' that is an account operated by the broadcaster NBC. 'Grassroots' is a group that can be configured as a normal user, such as the Twitter account of the authors (@_kimagine_ and @yhdfly), and helps to spread the news, occupying for 98% of the total users, but it is a relatively passive group. Finally, 'evangelist' is a group, like the '@taylorswift13' account of a singer, Taylor Swift. It consisted of leaders, politicians, celebrities, businessmen, and so on. Unlike mass media source that conveys only major topics, evangelists convey small topics. Moreover, they convey the topics to users that mass media source cannot reach.

About why rumors spread fast in social networks, Derre et al. [6] simulated a process spreading simple rumors. The simulation was based on a graph with the structure of the existing large social networks (see Figure 1). As a result, the rumors on social networks were mathematically proven to be spreading much faster than other network topologies.
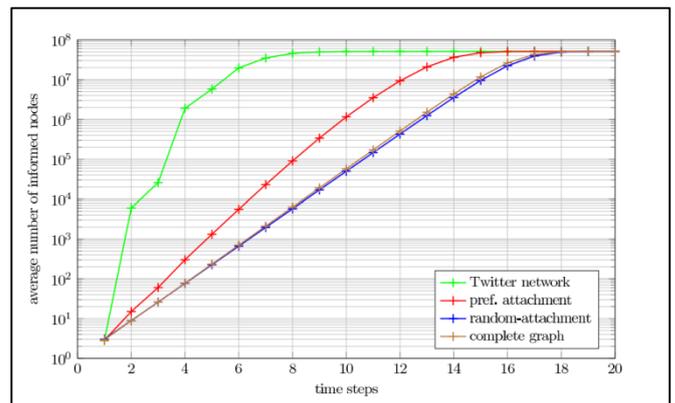

Figure 1: A graph showing how rumors spread starting at a random node [6]

Gruzd [10] explicated the reason for Twitter being used in the diffusion of scholar information. Twitter is free for all and easy to use, and from Twitter researchers can obtain ideas, feedback in real time, etc.

RELIABILITY OF TWEET

The more accounts we actively follow, the more we get a lot of information on Twitter. However, among the wealth of information, reliable data and unreliable ones coexist in Twitter. Twitter is emerging as an ideal of dialog platform. Meanwhile a tweet is growing concern about whether or not it is credible. Thus there have been ways to determine the reliability of tweets according to a certain rule [8]. Castillo et al. [13] proposed a way to determine information reliability in Twitter automatically. They found features in an article posted by users, and then the article is classified into reliable post and unreliable one with precision and recall reaching 70%-80%. Westerman et al. [18] experimented the effect of connections related to the reliability of Twitter. They determined the account posting reliable tweets by analyzing the ratio of the numbers of followers and following users. As a result, if the difference between the numbers of follower and following users is small, the account's tweet is reliable.

INFLUENCE OF TWITTER

Twitter largely influences the whole society, as well as it gives useful information to individual users. A company 'Burson Mastller' analyzed the number of retweets. Then they ranked leaders of the world via influence on Twitter. Willson and Supa [15] showed how practitioners and journalists each use Twitter for relations with professions unlike their own affairs. In their study, it was confirmed that Twitter can play an important role during the formation of the relationship between people with different duties. However, Twitter has been seen as means of communication rather than real relationships. Byrne [17] studied an improved effect of Twitter as developing policies related to human rights.

SPAM DETECTION

Because Twitter is a social network service with a variety of users, it is characterized as with high diffusivity. We can often see repeatedly spam tweets containing URLs passed to illegal gambling sites or pornographic ones.

Wang [21] examined user's followers and friend relationships, and he proposed a detection system that reached 89% precision based on the spam policies of Twitter. Zang et al. [20] presented a framework consisting of three steps to detect spam and promotional campaigns at Twitter. At first, they connected accounts that posted URLs with similar purpose. Secondly, they chose candidate posts having purpose of spamming or promoting. Finally, they decided whether or not it is spam through contents of the posts selected as candidate spam tweets. Also, they introduced a URL-based estimation method for measuring the degree of similarity between purposes of published URLs, and proposed a function based on machine leaning to separate posts estimated as spam. Wang and Pu [9] introduced a function for detecting URL spam through analyzing account's behavior. One of the function's features was 'investigating' which is the behavior of people who send unusual spam message. In their study, they found a number of URL spam that cannot be detected by conventional methods, and investigated unusual patterns of spammer's behavior unlike ordinary users.

Chen et al. [23] detected spam based on retweet relationship. They focused on tweets that record the highest number of retweets. Then, they decided which tweet spreads spam contents.

Song et al. [7] filtered spam through relationship between sender and receiver. To determine whether or not a message is spam, they took advantage of the relationship between the message sender and receiver. As a result, they found that most of spam come from accounts via more than three accounts which a user does not follow directly. There is a spammer detecting method using relationship-based features, to solve inconvenience of users due to harmful contents and mass production from spammer accounts that were authenticated in e-mail [16].

Santos et al. [19] referred that Twitter blacklist system does not always detect all spammers. So they filtered out spam tweets through machine learning algorithms and compression of tweet contents.

Nagy and Kanzi [22] proposed a method by analyzing the content of tweet messages for spam detection consisting of three steps. The steps are as follows: firstly, they appeared to differentiate and marked between spammers and normal users through classification like words of post, character of user, etc. Secondly, messages from normal user and spammer were used in analyzing behavior pattern of spam messages. Then they distinguished spammers from users and ranked users. Thirdly, they classified message's functions and evaluated performance.

Kim et al. [14] showed a statically significant improvement in performance of rule-based automatic classification using a dictionary for filtering of harmful spam tweets. In addition, Maccord and Chuah [24] explicated difference between ordinary user and spammer by illustrating features based on user and contents.

APPLICATIONS OF TWITTER

In the field of advertising and promoting, applications of Twitter have been much enabled. Also, the importance of advertising using a social network is growing day by day. Starbucks (@Starbucks) is an example that operates an account for marketing. Through the account, its officials hear customer's opinions and proceed various events and advertisements. As well as promoting field, Twitter has been applied in the field of education.

Shiffman [2] said that Twitter can be a powerful tool for environmental education and public relations. Although there are many professionals in this field, important information they explained on environmental conservation got less attention from the public. However, the information can be shared via Twitter. Rinaldo et al. [1] said that Twitter can be a great teaching tool for marketing educators. In addition, they proposed a Twitter application as an easy and quick way for educators to post announcements to students and for students to solve problems. Fife et al. [3] studied Twitter applications in high education classroom. When an instructor used Twitter in a classroom, it was evaluated as more effective than not to use that. Moreover, they noted the necessity of more researches focused on education using Twitter in high education classroom. Veletsianos [4] has shown positive

results of self-directed learning participation on social networks.

Twitter boasts a powerful marketing tool that is disseminative through a short text and retweet features. Yoon and Hwang [12] designed a system to analyze the performance of sales promotion marketing existing on Twitter. The system established an emotion valuation dictionary suitable for online, extracted qualities that represent feelings, and then recognized an emotion contained in tweets. In addition, a performance indicator was proposed that incorporated into a network with respect to subject, considered the reaction of customers on Twitter, and then quantified them through the social relations of reasoning tweets[12].

Eftekhar and Koudas [5] alluded to advertisement on Twitter. Twitter advertising platform includes advertising promotional tweets/accounts, advertising options/keywords for trends, and user targeting methods based on keywords, interests, location, etc. Likewise, they reported the following three challenges of advertising by Twitter. Firstly, when it comes to advertisers, they select a less expensive one as an alternative favorite theme for advertising. Second one was to identify user groups with different interests. The third was to classify experts on the basis of different specialized topics.

CONCLUSION

We could see that there have been many studies conducted during recent 5 years on various applications and impacts of Twitter. Moreover, we have summarized these studies by dividing them into fields of information diffusion, information reliability, spam detection, and Twitter application. Besides, it was possible to categorize diffusion features on Twitter and to know a diffusion process through simulation. Recent studies dealt with about the spread of information through newly emerged hash tags on Twitter. Meanwhile, Twitter has given impacts such as policies, and then has brought social progresses. There was such use of the 'following' number and 'followers/following' ratio to determine the reliability of tweets. We expect to require a variety of researches on other methods for determining whether or not a tweet is reliable. In addition, based on the relationships in the contents of links and tweets, spam and spammers can be detected. Various methods will be needed for detecting spam tweets directly and blocking them. Studies on applications of Twitter being diversified subsequently and will be continually expanded.

ACKNOWLEDGEMENTS

REFERENCES

[1] S. B. Rinaldo, S. Tapp, and D. A. Laverie, Leaning by tweeting: using Twitter as a pedagogical tool, Journal of Marketing Education, 2011

[2] D. S. Shiffman, Twitter as a tool for conservation education and outreach: what scientific conferences can do to promote live-tweeting, Journal of Environmental Studies and Sciences, Vol. 2, No. 3, pp. 257-262 2012

[3] E. Fife, C. L. Nelson, and T. B. Clarke, Online technological media in the higher education classroom: an exploratory investigation of varied levels of Twitter use, International Journal of Online Pedagogy and Course Design, Vol. 4, No. 2, 2014

[4] G. Veletsianos, Higher education scholars' participation and practices on Twitter, Journal of Computer Assisted Learning, Vol. 28, No. 4, pp. 336-349, 2012

[5] M. Eftekhar and N. Koudas, Some research opportunities on Twitter Advertising, IEEE Data Engineering Bulletin, pp. 77-82, 2013

[6] B. Doerr, M. Fouz, and T. Friedrich, Why rumors spread so quickly in social networks, Communications of the ACM, Vol. 55, No. 6, pp. 70-75, 2012

[7] J. Song, S. Lee, and J. Kim, Spam filtering in Twitter using sender-receiver relationship, Recent Advances in Intrusion Detection, Vol. 6961, pp. 301-317, 2011

[8] J. An, W. J. Li, L. N. Ji, and F. Wang, A survey on information credibility on Twitter, Applied Mechanics and Materials, Vols. 401-403, pp. 1788-1791, 2013

[9] D. Wang and C. Pu, BEAN: a BEhavior ANalysis approach of URL spam filtering in Twitter, IEEE International Conference on Information Reuse and Integration, pp. 403-410, 2015

[10] A. Gruzd, Discovery and Visualization of Scholarly Information Diffusion in Twitter Networks, Workshop on Mapping Ideas: Discovering and Information Landscapes, 2012

[11] M. Cha, F. Benevenuto, H. Haddadi, and K. Gummadi, The world of connections and information flow in Twitter, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, Vol. 42, No. 4, pp. 991-998, 2012

[12] J. Yoon and B. Hwang, Design of a productivity measuring system for Twitter marketing reflecting user's sentiment in social networks, Journal of the Korean Institute of Information Scientists and engineers: Database, Vol. 40, No. 6, pp. 377-385, 2013

[13] C. Castillo, M. Mendoza, and B. Poblete, Information credibility on Twitter, In Proceedings of the 20th International Conference on World Wide Web, pp. 675-684, 2011

[14] M. Kim, K. Shim, N. Han, Y. Kim, and M, Song, Automatic classification of malicious usage on Twitter, Korean Society for Library and Information Science, Vol. 47, No. 1, pp. 269-286, 2013

[15] D. Wilson and D. W. Supa, Examining modern media relations: an exploratory study of the effect of Twitter on the public relations–journalist relationship, Public Relations Journal, Vol. 7, No. 3 , 2013

[16]   C. Lee and J. Kim, Spammper detection using features based on user relationships in Twitter, Journal of the Korean Institute of Information Scientists and engineers, Vol. 41, No. 10, pp. 785-791, 2014

[17]   C. Byrne, The Twitter effect, Humanities and Social Sciences papers, pp. 96-99, 2011

[18]   D. Westerman, P. R Spence, and B. V. D. Heide, A social network as information: the effect of system generated reports of connectedness on credibility on Twitter, Computers in Human Behavior, Vol. 28, No. 1, pp. 199-206, 2012

[19]   I. Santos, Igor M. Miñambres-Marcos, C. Laorden, P. Galán-García, A. Santamaría-Ibirika, and P. G. Bringas, Twitter content-based spam filtering, International Joint Conference, Vol. 239, pp. 449-458, 2014

[20]   X. Zhang, S. Zhu, and W. Liang, Detecting spam and promoting campaigns in the twitter social network, IEEE 12th International Conference on Data Mining, pp. 1194-1199, 2012

[21]   A. H. Wang, Machine learning for the detection of spam in Twitter networks, e-Business and Telecommunications, Vol. 222, pp. 319-333, 2012

[22]   G. I. Nagy and S. Kazi, Filtering noise from stock related Twitter messages, 9th Conference of PhD Students in Computer Science, pp. 41-42, 2014

[23]   P. C. Chen, H. M. Lee, H. R. Tyan, J. S. Wu, and T. E. Wei, Detecting spam on Twitter via message-passing based on retweet-relation, Technologies and Applications of Artificial Intelligence, Vol. 8916, pp. 66-65, 2014

[24]   M. Mccord and M. Chuah, Spam detection on Twitter using traditional classifiers, Autonomic and Trusted Computing, Vol. 6906, pp. 175-186, 2011

[25]   S. Y. Park, Y. H. Ha, and Y. H. Kim, Recent studies on Twitter in the field of information retrieval, Korea Information Science Society, pp. 25-29, 2010