Journal of Theoretical and Applied Information Technology
31st January 2013. Vol. 47 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A SURVEY OF RESEARCHING ON MICRO-BLOGGING

[1]LEINA JI, [2]WEIJIANG LI, [3]HUI DENG, [4]FENG WANG

[1] Ms., Computer Technology Appliance Key Lab of Yunnan Province, KUST, Kunming, CHINA

[2]Assoc. Prof., Computer Technology Appliance Key Lab of Yunnan Province, KUST, Kunming, CHINA

[3]Assoc. Prof., Computer Technology Appliance Key Lab of Yunnan Province, KUST, Kunming, CHINA

[4] Prof., Computer Technology Appliance Key Lab of Yunnan Province, KUST, Kunming, CHINA

E-mail: [1]jileina@gmail.com , [2]hrbrichard@gmail.com , [3]denghui@cnlab.net , [4]wangfeng@cnlab.net

ABSTRACT

Micro-blogging is an emerging internet communication platform based on the social relationships. It can be used to share, disseminate and obtain all kinds of real-time updated information with about 140 words. Currently, micro-blogging becomes more and more popular since users exploit it to achieve a wide variety of social purposes. According to relevant public data, the earliest and most famous micro-blogging -- Twitter already has 500 million registered users in the world. Thus we survey the considerable literature on Twitter and show masses of research theoretical results in the area obtained over the past few years. By reading the paper, we hope authors can gain an in-depth understanding of the current research focus and tendency related to the results of the Twitter. Specifically, in this paper, we first introduce micro-blogging and its difference with traditional blogging. Then we list several different research directions about Twitter as well as their research purposes, methods and conclusions. Finally, we show the gaps of previous researches and prospect the future work of micro-blogging.

Keywords: *Twitter, Micro-blogging, Influence*

1. INTRODUCTION

Micro-blogging, an emerging form of communication medium, provides a new channel for people to publish short messages updates, including daily conversations, URLs sharing and information news, by instant messages, email, the web and cell-phone.

Actually, micro-blogging is a new informal communication medium in the form of blogging. But it differs from a traditional blogging, in which its content is typically smaller in both form and scale. By encouraging shorter posts, micro-blogging lowers users' requirement of time and thought investment for content generation [1]. The other important difference is the frequency of update [1]. Blogging is equivalent to electronic diary, which allows people to insert pictures, flash, and music into text. However, Blogging, as a close personal space, is lack of the use of the social relations. Micro-blogging, as an open personal space, can be more convenient and quick to express their views and comment the views of others, integrating the SNS social relations with the Blogging mode. Specific differences between them are shown in table 1.

Table 1: Difference Between Micro-Blogging And Traditional Blogging

	Micro-Blogging	**Blog**
Content	One word	A full blown article
Word Limit	140 characters	Long pages
Interface	Mobile device, IM software, API interface	Computer, web
Characteristic	Timely, speedy, broad convenient	Poor timeliness
User Experience	The use of topic changes and widgets	Concise page
Methods of Propagation	Retweet followers, initiative, speedy	Recommended sites, passive
User Group	All kinds of people	Elite crowd
The Frequency of Update	Several updates in a single day	Once every few days

One of the most notable Micro-blogging services is Twitter, which was created in March 2006 and launched inJuly. It allows twitter users to send a read text-based post of up to 140 characters, known

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

as "tweets". The service rapidly gained worldwide popularity, with over 500 million active and handling over 1.6 billion search queries per day, which makes Twitter one of the fastest-growing sites in the world [2]. Twitter employs a social networking model called "following", in which each user can choose whom she wants to "follow" to receive tweets form without requiring permission [3]. So Twitter not only gains huge popularity among people, but also drawing increasing interest from research community.

In this paper, we first introduce micro-blogging and its difference with traditional blogging in section 1. Then we list several different research directions about Twitter as well as their research purposes, methods and conclusions in section 2. In Section 3, we put forward more focused on some possible directions and gaps of research in this area. Finally, section 4 concludes with directions for further research.

2. RELATED WORK

Millions of people around the world use Twitter to remain social connections to their friends, family members and co-workers through their computers and mobile phones [4]. So, Twitter gains huge popularity and draws increasing interest from research community. In this section, we will begin with the Sociology, then list parts of research work in different respects about Twitter.

A. Social psychological analysis

With the recent popularity of Twitter and similar micro-blogging systems, many researchers have proposed their question about "why so many people use it?" Understanding this question will help us improving and adding new features that would retain more users.

Java et al. [1] tackle this problem by studying the micro-blogging phenomena and analyzing different types of user intentions. It describes the social network of Twitter users and investigates the motivation of Twitter users. [5] seeks to build a rich understanding of why people use micro-blogs and explores how the characteristics of their micro-blogging behaviors enable informal communication. In the Figure 1, they have organized the benefits discussed for informal communication into relational and personal beneficial consequences. Emotionally, people seem to use micro-blogging to achieve a level of cyberspace presence, being "out there" and to feel another layer of connection with friends and the world [6] . Finally, a conclusion was given. The main types of user intentions are: (1) keeping close connection with friends and colleagues; (2) building one's influence among followers; (3) collecting useful and interesting information; (4) asking for help and opinions; (5) discharging emotional pressure. Furthermore, users play different roles of information source, friends or information seeker in different communities [1].

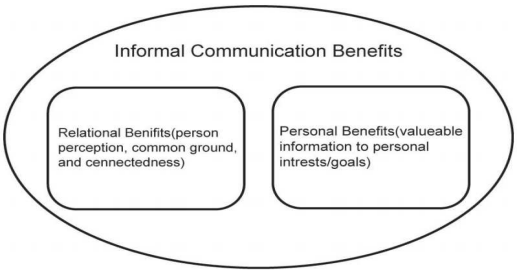

Figure1. Proposed Benefits Of Informal Communication

B. Social network & Information network

As we all know, Twitter can be used to update information about their daily lives, follow many people they like, rebroadcast what they are font of, and interact with each other. Just like studies [7, 8] have shown that Twitter users use Twitter as a new tool of network communication; making it look like a hybridization between conventional blogging and an online social network. So, many researcher think twitter is a social network.

However, the Twitter follower graph [9] does not conform to the usual characteristics of social networks, which exhibits much higher reciprocity and far less skewed degree distributions [10], but instead resembles more the mixture of one-way mass communications and reciprocated interpersonal communications [11]. Kwak et al. [9] made a large number of research on the topological features of the Twitter follower graph. They drew a conclusion from the highly skewed nature of the distribution of followers and the low rate of reciprocated ties that twitter is more similarly to be an information communication network than a social network. Just like Evan Williams, one of the founders of Twitter.com, said that they think of Twitter as it's not a social network, but it's an information network, it tells people what they care about as it is happening in the world (see Figure 2).

Journal of Theoretical and Applied Information Technology
31st January 2013. Vol. 47 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

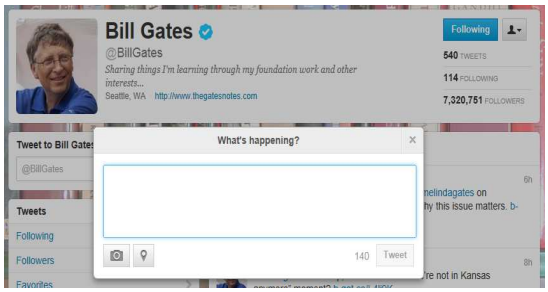

Figure 2 Status Update Has Change To "What's Happening?"

This result made our research aspect from social network to information network and how the information diffusion on it.

C. Opinion leaders

Aside from the literature surveyed above, a number of papers have examined information diffusion on Twitter. Mid-century researchers[12-15] argued that the mass media influenced the public only indirectly, via what they called a two-step flow of communications, where the critical intermediate layer was occupied by a category of media-savvy individuals called opinion leaders [11].

It is clear to all the users that Twitter has a section on its main page entitled "Trending Topics", which displays the top 10 mentioned terms on Twitter at any given moment. And one of the hottest tasks is to search for the opinion leader of this Trending Topics, in order to detect a newly emerging topic. Aiming at this problem, [16] applied visualization techniques in conjunction with artificial intelligence-based data mining methods to classify messages to deal with the trend topic. This is potentially useful in many areas, such as marketing, epidemic research and other related work. [3] proposed a novel approach to measure the influence of twitterers, known as TwitterRank. The framework of the proposed approach is shown in Figure 3. Experimental results show that TwitterRank outperforms the one Twitter currently uses and other related algorithms. However, it ignores the possibility of r users to interact with the content in Twitter. Meeyoung et al. [17] presented an indepth comparison of three measures of influence: in-degree, retweets, and mentions. They found in-degree alone reveals very little about the influence of a user and most influential users can hold significant influence over a variety of topics.

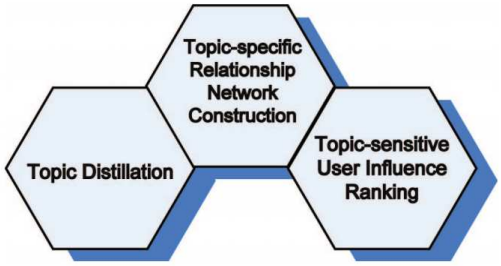

Figure 3 Framework Of The Proposed Approach

D. Sentiment analysis

Sentiment Analysis is also called opinion mining. It aims to classifying the polarity of a given text at the document, sentence, or feature/aspect level [18] and mining user's attitude to some topic or a document.

Ming people's sentiment expressed has attracted more and more attention. However, the main problem is that the messages posted on Twitter, so-called tweets, are very short. The maximum size of a tweet is 140 characters. [19] used the tag and emoticons in Twitter as feature to training a classifier that is similar to KNN for sentiment polarity analysis and show that their framework successfully identifies sentiment types of untagged sentences. [20] leveraged sources of noisy labels which were provided by a few sentiment detection websites over twitter data as their training data, and proposed a two-step sentiment analysis classification method for Twitter, namely: first classifies messages as subjective and objective, and further distinguishes the subjective tweets as positive or negative. [21] also adopted this method, but they propose to improve sentiment classification by using both target-dependent and context-aware approaches. In addition, they also proposed to incorporate the contexts of tweets into classification. By considering the sentiment labels of the related tweets, they further boosted the performance of the sentiment classification, especially for very short and ambiguous tweets.

E. Topic detection

Topic detection which refers to resolve the information-overload problem is an important research direction in text mining. With the most popular of twitter, topic detection on Twitter data has attracted much attention recently.

Takeshi et al. considers Twitter as a social sensor for detecting large scale events like earthquake, typhoons and traffic jams. They investigated the real-time nature of twitter and developed an earthquake reporting system [22], which is a novel

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

approach to notify people promptly of an earthquake event. Sasa and Miles [23] put forward a novel modified topic detection method that can used for fast processing abundant Short messages which contain more than 1.6 Billion news Without losing accuracy. [24] propose a novel topic detection technique that permits to retrieve in real-time the most emergent topics expressed by the com-munity. They [24] extract the contents of the tweets and model the term life cycle according to the new method to mine the emerging ones. And they used Page Rank to determine the authority of the users to analyze the social relationships in the network. Finally, they leveraged a navigable topic graph which connects the emerging terms with other semantically related keywords, allowing the detection of the emerging topics, under user-specified time constraints.

3. OPEN ISSUES

Micro-blogging is a fresh internet commu-nication platform based on the social relationships that had already attracted many researchers' attentions as a novel research branch in the area of information processing. And it brings forward higher request for the research methods, due to its huge bursts data and special "following" relationship.

In Section 2, we have listed parts of research work in different respects about Twitter, i.e. social psychological analysis, social network & infor-mation network, opinion leaders, sentiment analysis and topic detection. From these works, we can see that the researches in this field are very useful, but each of them is in its early stage, and needs to be developed. Specifically, previous works on infor-mation retrieval on Micro-blogging still exploit traditional technologies such as filter, classification, cluster etc., based on the statistic strategy, which ignores Twitter itself has such characteristics as the similarity of friends' attention, point of interests, bursts of trending topics and so on. Aim at this question, current research trends on Twitter are to integrate multiple traditional methods with the characteristics mentioned above on Twitter, such as social psychological analysis, opinion leaders, topic detection and so on. Therefore, for future work, the research trends on Twitter will mainly focus on the followings: (1) build a statement model which collects a set of messages related to news events from Twitter; (2) integrate machine learning with natural language processing effectively; (3) adaptive learning and update strategy for burst event detection.

4. CONCLUSION

In this paper, we first introduce micro-blogging and its difference with traditional blogging and then survey the research of the Twitter with the purpose of giving authors an in-depth understanding of the current research focus and tendency as well as the related results of the Twitter. By this paper, we can see that the research in this field is in its early stage, and needs to be developed. The tremendous data of message have exerted a profound influence on various aspects, in future, besides the characteristic of the Twitter itself, more research may be done with the combination of the Twitter and other relevant fields, just like [25] and [26]. In China, the most possible future work encompasses an implementation of a platform overlooking various aspects of the Sina micro-blogging, which is the most popular site domestically. We recently learned that Qihoo has diverted its attention to the fields of Social Media Monitor by using the technology in information retrieval. It is believed that the paper related is worth studying and attractive results should come out.

5. ACKNOWLEDGMENT

Thank for supports of open project provided by computer technology appliance key lab of Yunnan province.

REFRENCES:

[1] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities," *in Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007*, San Jose, California , USA, 2007, pp. 56-65.
[2] Twitter. *http://en.wikipedia.org/wiki/Twitter*.
[3] J. Weng, E. P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," *in the WSDM'10: the Third ACM International Conference on Web Search and Data Mining*, New York, 2010, pp. 261-270.
[4] S. Milstein, Twitter and the micro-messaging revolution: Communication, connections, and immediacy--140 characters at a time: *O'Reilly Media*, Incorporated, 2008.
[5] D. Zhao and M. B. Rosson, "How and why people Twitter: the role that micro-blogging plays in informal communication at work," *in Proceedings of the ACM 2009: international conference on Supporting group work, Sanibel Island*, Florida, 2009, pp. 243-252.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[6] P. McFedries, "Technically speaking: All a-twitter," *Spectrum, IEEE*, vol. 44, pp. 84-84, 2007.
[7] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: An analysis of a micro-blogging community," *Advances in Web Mining and Web Usage Analysis*, pp. 118-138, 2009.
[8] E. Mischaud, "Twitter: Expressions of the whole self," *An in*, 2007.
[9] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," *in WWW 2010: the ACM proceedings of the World Wide Web*, Raleigh, North Carolin, 2010, pp. 591-600.
[10] G. Kossinets and D. J. Watts, "Empirical analysis of an evolving social network," *Science*, vol. 311, pp. 88-90, 2006.
[11] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts, "Who says what to whom on twitter," *in WWW 2011: Proceedings of the 20th international conference on World Wide Web*, Hyderabad, India, 2011, pp. 705-714.
[12] P. F. Lazarsfeld, B. Berelson, and H. Gaudet, *The people's choice: how the voter makes up his mind in a presidential campaign*, 3rd edition ed.: Columbia Univ. Press, 1965.
[13] J. Coleman, E. Katz, and H. Menzel, "The diffusion of an innovation among physicians," *Sociometry*, pp. 253-270, 1957.
[14] E. Katz and P. F. Lazarsfeld, "Personal influence: The part played by people in the flow of mass communications", *Transaction Pub*, 2006.
[15] R. K. Merton, "Patterns of influence: Local and cosmopolitan influentials," *Social theory and social structure*, pp. 387-420, 1957.
[16] M. Cheong and V. Lee, "Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base," 2009, pp. 1-8.
[17] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in twitter: The million follower fallacy," *in the 4th International AAAI Conference on Weblogs and Social Media*, Washington DC 2010, p. 8.
[18] M. de Haaff, "Sentiment Analysis, Hard But Worth It!, CustomerThink," ed: retrieved 2010-03-12, 2010.
[19] D. Davidov, O. Tsur, and A. Rappoport, "Enhanced sentiment learning using twitter hashtags and smileys," 2010, pp. 241-249.
[20] L. Barbosa and J. Feng, "Robust sentiment detection on twitter from biased and noisy data," *in Proceeding HLT'11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, 2010, pp. 36-44.
[21] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification," *in Proceeding HLT'11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, 2011, pp. 151-160.
[22] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," *in the 19th international conference on World wide web, Raleigh, North Carolina*, USA, 2010, pp. 851-860.
[23] S. Petrovic, M.Osborne, and V. Lavrenko, "Streaming first story detection with application to twitter," 2010.
[24] M. Cataldi, L. Di Caro, and C. Schifanella, "Emerging topic detection on Twitter based on temporal and social terms evaluation," *in Proceeding MDMKDD'10: Proceedings of the Tenth International Workshop on Multimedia Data Mining*, 2010, p. 4.
[25] A. M. Popescu, M. Pennacchiotti, and D. Paranjpe, "Extracting events and event descriptions from twitter," *in WWW'11: Proceedings of the 20th international conference companion on World wide web*, Hyderabad, India, 2011, pp. 105-106.
[26] L. Hong, O. Dan, and B. D. Davison, "Predicting popular messages in twitter," *in the WWW'11: Proceedings of the 20th international conference companion on World wide web*, Hyderabad, India, 2011, pp. 57-58.