2010 International Conference on P2P, Parallel, Grid, Cloud and Internet Computing

A Credibility Assessment for Message Streams on Microblogs

Yu Suzuki
Information Technology Center, Nagoya University
Furo, Chikusa, Nagoya, Aichi 464-8601, Japan
Email: suzuki@db.itc.nagoya-u.ac.jp

***Abstract*—Recently, many messages are submitted and read using microblog Web services, such as twitter, facebook, and others. These messages are written by unspecified users, these messages consists of unreliable contents. Therefore, to specify which contents are reliable or not, we should analyze messages. However, these messages are too short, we are hard extract meaningful, useful features from these messages. In this paper, we propose a method to assess credibility values of messages using remaining ratio of ReTweets. We assume that if a high credibility messages are retweeted, the original messages are remain. However, if a low credibility messages are retweeted, several terms are added about opinion of users. Therefore, if there are many retweets of a message, and these retweets are remain, the message should be credible. Using this assumption, we propose a method to calculate credibility degrees of messages using added and deleted messages of retweets.**

***Keywords*-Microblog; stream; credibility; short message; twitter**

I. Introduction

Recently, short message services called microblogs, such as Twitter[1] and facebook[2], are widely used to exchange messages between users. These services are used by many kinds of users, such as politicians, famous persons, governments, companies, etc. These users posts a lot of messages to microblogs, which consist of valuable messages, and also consist of useless messages.

Re-Tweet (RT) is one of the major functions, which is a message reposted by the other users. This function is mainly used to share interesting messages with other user groups. However, the problem is that the origially posted messages and re-tweeted messages are not always the same. In addition, several users manually change these messages to different contexts. Therefore, we need to detect which re-tweeted messages are adequate or indequate, which means which re-twitted messages remain original contexts or not.

We think that the deletion and addition of original messages as re-tweets are not always inadeqate. This means that, even if many terms are changed, the re-tweitted messages are adequate if original contexts remained. Therefore, if the messages

In this paper, we propose a calculation method of credibility values for re-twitted messages. In our system, we assume that if a high credibility message is posted to microblogs, this message is retweeted by many users with a small number of edits. Therefore, if a message is not credible, the message is ignored and not retweeted by the other users. On the other hand, if a message is credible, the message is frequently retwitted by the other users. Using this assumption, we contruct a credibility degree calculation algorithm.

Our proposed system consists of the following five steps:

1) Extract re-tweets from Twitter timelines using Twitter Streaming API[3].
2) Extract users name and their statistical values from re-tweets.
3) Calculate credibility values of users for each re-tweets.
4) Calculate credibility values of re-tweets using users' credibility values.

We use a reputation-based credibility degree assessment method. Adler et al. [1]–[3], Hu et al. [4] proposed a reputation-based credibility degree assessment method for Wikipedia[4] articles. In these methods, if the lifetime of versions is long, the versions should be treated as credible. This idea is similar to our study. However, intuitively, the credibility degrees of twitter and Wikipedia are slightly different. Therefore, we should discover whether the system can distinguish credible retweets or not.

II. Credibility value calculation system

For assessing credibility values for all versions of twitter messages, we first assess the appropriateness of edits. Next, we assess credibility values of users. Then, we assess credibility values of versions using the users' edit histories. In this section, we first present our analysis model. Based on this model, we describe the meaning and measurements of edit appropriateness, editors, and versions.

A. Modeling Re-Tweet message History

We define several notations used in the following sections. A set of retweets includes messages $i = 1, 2, \cdots M$, and

[1] http://www.twitter.com/
[2] http://www.facebook.com
[3] http://dev.twitter.com/pages/streaming_api
[4] http://ja.wikipedia.org/

978-0-7695-4237-9/10 $26.00 © 2010 IEEE
DOI 10.1109/3PGCIC.2010.90

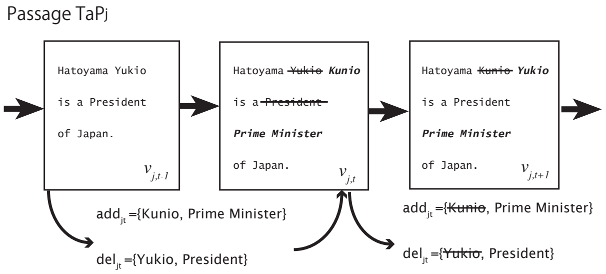

Figure 1. Example of added and deleted contents in retweet history

when a user retweets articles $v_{i,j}$. i is an retweet index and j is a version number. $v_{i,1}$ is the original version of message i. We define $v_{i,0}$ as an message with no content.

We define the users $e = 1, 2, \cdots, E$ who retweet messages more than once. $A_e = \{v_{i,j} | v_{i,j} \in V$ and $v_{i,j}$ is edited by $e\}$ is the set of retweets by e. One retweet message is edited by the other users. We do not set the users of the articles with version number $j = 0$.

B. User's Contribution to ReTweets

In our system, we first assess whether the version from $v_{i,j-1}$ to $v_{i,j}$ is appropriate, and calculate the user's contribution values of retweets $\tau(v_{i,j})$. We define the appropriateness values of retweets as the ratio of unchanged edits and non-reverted deletions after other users' edits. This is because, when the user adds appropriate terms to the retweets, these terms should remain and not be deleted by other users. In the same way, when a user removes inappropriate terms from articles, these terms should be deleted by other users.

We show an example of appropriate and inappropriate retweets in Figure 1. First, we identify the edits from $v_{i,j-1}$ and $v_{i,j}$ using diffs. $add_{i,j}$ is the set of added terms and $del_{i,j}$ is the set of deleted terms. In this example, $add_{i,j}$ consists of "Kunio" and "Prime Minister", and $del_{i,j}$ consists of "Yukio" and "President".

Next, we identify the remaining edits of $add_{i,j}$ and $del_{i,j}$ at $v_{i,j+p}$, where $p = 1, 2, \cdots, N_i - j$. We define $\delta(add_{i,j}, p)$ as the set of terms $add_{i,j}$ in $v_{i,j+p}$, and $\delta(del_{i,j}, p)$ as the set of terms $del_{i,j}$ in $v_{i,j+p}$. When we identify $\delta(add_{i,j})$, we compare $v_{i,j}$ and $v_{i,j+p}$, and extract the same parts of retweets.

We define the contribution ratio of additions $R^{add}(i, j, p)$ and that of deletions $R^{del}(i, j, p)$ as follows:

$$R^{add}(i, j, p) = \frac{|\delta(add_{i,j}, p)|}{|add_{i,j}|} \tag{1}$$

$$R^{del}(i, j, p) = \frac{|\delta(del_{i,j}, p)|}{|del_{i,j}|} \tag{2}$$

where $|\delta(add_{i,j}, p)|$, $|\delta(del_{i,j}, p)|$, and $|add_{i,j}|$, $|del_{i,j}|$ are the number of terms in $\delta(add_{i,j}, p)$, $\delta(del_{i,j}, p)$, $add_{i,j}$, and $del_{i,j}$.

We show how to calculate $R^{add}(i, j, 1)$ and $R^{del}(i, j, 1)$ at $p = 1$ using Figure 1. In this case, $add_{i,j}$ consists of 18 letters without spaces, and $\delta(add_{i,j}, p)$ consists of 13 letters. Therefore, $R^{add}(i, j, 1)$ is $\frac{13}{18} = 0.72$. In the same way, $del_{i,j}$ consists of 14 letters, and $\delta(add_{i,j}, p)$ consists of 9 letters. Then, $R^{del}(i, j, 1)$ is $\frac{9}{14} = 0.64$.

The problem is that the contribution ratio of addition $R^{add}(i, j, p)$ and that of deletion $R^{del}(i, j, p)$ is not equivalent when we change p, because if the number of edits after version $v_{i,j}$ increases, the contribution of addition and deletion should be high. Therefore, we normalize $R^{add}(i, j, p)$ and $R^{del}(i, j, p)$ using the standard addition and deletion ratio. We define the standard addition/deletion ratio as the ratio of edits that remains/reverts after p versions.

The normalized contribution ratio of addition and deletion, $\overline{R^{add}(i, j, p)}$ and $\overline{R^{del}(i, j, p)}$, is define as follows:

$$\overline{R^{add}(i, j, p)} = \frac{R^{add}(i, j, p)}{S^{add}(p)} \tag{3}$$

$$\overline{R^{del}(i, j, p)} = \frac{R^{del}(i, j, p)}{S^{del}(p)} \tag{4}$$

where $S^{add}(p)$ and $S^{del}(p)$ is the average value of $R^{add}(i, j, p)$ and $R^{del}(i, j, p)$ for each p in all i and j. In our preliminary experiments, we show that $S^{add}(1) = 0.93$ and $S^{del}(1) = 0.99$. Therefore, we can calculate that $\overline{R^{add}(i, j, 1)} = \frac{0.72}{0.93} = 0.77$, and $\overline{R^{del}(i, j, 1)} = \frac{0.64}{0.99} = 0.64$.

Next, we calculate the user's contribution degrees of retweets $\tau(v_{i,j})$ using $R^{add}(i, j, p)$ and $R^{del}(i, j, p)$ as follows:

$$\tau(v_{i,j}) = \sum_{q=1}^{N_i - j} R^{add}(i, j, p) + \sum_{q=1}^{N_i - j} R^{del}(i, j, p) \tag{5}$$

The value of N_i depends on article i; however, we do not normalize $\tau(v_{i,j})$ using N_i because we assume that if the number of edits increases, the credibility values of edits should increase. In the example in Figure 1, $\tau(v_{i,j}) = \overline{R^{add}(i, j, 1)} + \overline{R^{del}(i, j, 1)} = 0.77 + 0.64 = 1.41$.

Finally, we normalize the values $\tau(v_{i,j})$ for converting the average value of all $\tau(v_{i,j})$ to 0. This is because almost all edits are small, and these edits do not change the credibility values of retweets. Therefore, if the credibility value of an retweet is lower than the average value of $\tau(v_{i,j})$, we should determine that the edit decrease the credibility value of the article. Using this normalization, the credibility values for these retweets are converted to negative values. We convert the normalized value of $\overline{\tau(v_{i,j})}$ follows:

$$\overline{\tau(v_{i,j})} = \tau(v_{i,j}) - \frac{\sum_{i=1}^{M}\sum_{j=1}^{N_i}\tau(v_{i,j})}{\sum_{i=1}^{M}N_i} \tag{6}$$

Using this credibility value of retweets, we calculate the credibility values of users.

C. Credibility values of editors

In this section, we calculate the user's credibility values using the credibility values of retweets. First, we set A_e as the set of versions edited by e. Then, we calculate the credibility values of retweets as follows:

$$U_e = \frac{\sum_{v_{i,j}\in A_e}\overline{\tau(v_{i,j})}}{|A_e|} \tag{7}$$

where $|A_e|$ is the number of retweets in A_e, which means the number of retweets by e.

III. Related Work

To our best knowledge, we cannot find credibility ratio for twitters and other microblogs. However, there has been much research on the credibility of Wikipedia articles [5]. In this section, we focus on the automatic or semi-automatic credibility value calculation methods for Wikipedia. When we extract data from Wikipedia articles, three methods are mainly used, a user voting system of articles, version analysis using natural language processing techniques, addition and deletion of edit history.

A popular method of using user voting was proposed by Kramer et al. [6]. For this method, they developed MediaWiki[5] and a user voting system was added. In this system, users can directly vote on which articles have better quality. However, many votes are needed to calculate the credibility of articles. Moreover, this system cannot calculate new articles' credibility values, which have not been read and not voted on by users. In our system, we do not need a voting mechanism, and our system can calculate credibility values even if the articles have not been read.

Adler et al. [1]–[3], Hu et al. [4] and Wilkinson et al. [7] proposed a system to calculate credibility values using edit history. These authors implemented this system to FireFox and MediaWiki plug-ins as the WikiTrust module [8] [6]. This system calculates the credibility values in real time. Our system is almost the same as Adler's. However, their system's performance is very slow because of high calculation costs. In our system, we improve the calculation cost problem by identifying important editors.

IV. Conclusion

In this paper, we proposed a method for calculating credibility degrees of retweets. In our system, we assumed that if a high credibility message is posted to microblogs, this message is retweeted by many users with a small number of edits. Therefore, if a message is not credible, the message is ignored and not retweeted by the other users. On the other hand, if a message is credible, the message is frequently retwitted by the other users. Using this assumption, we contructed the credibility degree calculation algorithms.

In the near future, we implement our proposed system, and confirm the accuracy of our propsoed method. Moreover, we will compare our proposed method with the other methods such as TwitterRank [9].

ACKNOWLEDGES

This work is partially supported by the Grant-in-Aid from Ministry of Education, Science, Sports and Culture, Japan (No, 20300036, 20500104, 21013026, 20700101).

References

[1] B. T. Adler and L. de Alfaro, "A content-driven reputation system for the wikipedia," in *WWW '07: Proceedings of the 16th international conference on World Wide Web*. New York, NY, USA: ACM, 2007, pp. 261–270.

[2] B. T. Adler, K. Chatterjee, L. de Alfaro, M. Faella, I. Pye, and V. Raman, "Assigning trust to wikipedia content," in *WikiSym '08: Proceedings of International Symposium on Wikis*. ACM, 2008.

[3] B. T. Adler, B. T. Adler, I. Pye, and V. Raman, "Measuting author contributions to the wikipedia," in *WikiSym '08: Proceedings of International Symposium on Wikis*, 2008.

[4] M. Hu, E.-P. Lim, A. Sun, H. W. Lauw, and B.-Q. Vuong, "Measuring article quality in wikipedia: models and evaluation," in *CIKM*, M. J. Silva, A. H. F. Laender, R. A. Baeza-Yates, D. L. McGuinness, B. Olstad, Ø. H. Olsen, and A. O. Falcão, Eds. ACM, 2007, pp. 243–252.

[5] B. Stvilia, M. Twidale, L. Smith, and L. Gasser, "Information quality work organization in wikipedia," *J. Am. Soc. Inf. Sci. Technol.*, vol. 59, no. 6, pp. 983–1001, 2008.

[6] M. Kramer, A. Gregorowicz, and B. Iyer, "Wiki trust metrics based on phrasal analysis," in *WikiSym '08: Proceedings of International Symposium on Wikis*. ACM, 2008.

[7] D. M. Wilkinson and B. A. Huberman, "Cooperation and quality in wikipedia," in *WikiSym '07: Proceedings of the 2007 international symposium on Wikis*. New York, NY, USA: ACM, 2007, pp. 157–164. [Online]. Available: http://dx.doi.org/10.1145/1296951.1296968

[8] K. Chatterjee, L. de Alfaro, and I. Pye, "Robust content-driven reputation," in *AISec*, D. Balfanz and J. Staddon, Eds. ACM, 2008, pp. 33–42.

[5] MediaWiki is a Wiki program used for Wikipedia, see `http://www.mediawiki.org/`

[6] `http://en.wikipedia.org/wiki/WikiTrust`

[9] J. Weng, E.-P. Lim, J. Jiang, and Q. He, “Twitterrank: finding topic-sensitive influential twitterers,” in *WSDM ’10: Proceedings of the third ACM international conference on Web search and data mining*. New York, NY, USA: ACM, 2010, pp. 261–270.