User Perception of Information Credibility of News on Twitter

Shafiza Mohd Shariff, Xiuzhen Zhang, and Mark Sanderson

School of Computer Science and IT, RMIT University, Australia
{shafiza.mohdshariff, xiuzhen.zhang, mark.sanderson}@rmit.edu.au

Abstract. In this paper, we examine user perception of credibility for news-related tweets. We conduct a user study on a crowd-sourcing platform to judge the credibility of such tweets. By analysing user judgments and comments, we find that eight features, including some that can not be automatically identified from tweets, are perceived by users as important for judging information credibility. Moreover, distinct features like link in tweet, display name and user belief consistently lead users to judge tweets as credible. We also find that users can not consistently judge or even misjudge the credibility for some tweets on politics news.

1 Introduction

As of May 2013, an average of 58 million tweets are posted per day on Twitter.[1] Currently Twitter not only acts as a social medium, it is also becoming a news media source. Twitter citizens not only share news headlines from newswires, but also report real time events before they reach the press [5]. News on Twitter comes from a wide variety of sources: some from well known news organisations and government departments, while most from members of the public. Consequently twitterers often need to judge the credibility of tweets. Morris et al. (2012) discovered that twitterers have poor judgement on the truth of information on Twitter. Features such as the number of retweets, information on users who post tweets and their relationship network (number of followers and followees) help little in determining the level of information credibility on Twitter.

Spammers exploit the anonymity feature of Twitter to propagate their messages, retweeting them to increase their popularity rating [10]. In a Twitter dataset analysed by Gupta and Kumaraguru (2012), nearly half of the tweets about an event were found to be spam. In the work by Castillo et al. (2011), it was discovered that the credibility of information on Twitter is determined mainly by four types of features: message-based, content-based, user-based, and propagation-based. In most existing work, the features may need to be compiled by crawling the Twitter space and extracting the link relationship between twitterers. The purpose of these features are for automatic prediction and may not necessarily be users' perception of important signals for credibility.

[1] http://www.statisticbrain.com/twitter-statistics/

In this paper, we focus on studying the tweet-based features that the general public mostly use to determine the credibility level of newsworthy tweet messages. The research questions that we will cover in this current work are:

1. What features do users use to judge the credibility of tweets?
2. How do users use tweet features to make their credibility judgment?
3. Does the tweet topic have an effect on a user's credibility judgment?

Related Work: Information credibility on Twitter has attracted significant research recently [1–4, 6, 7], where most focuses on automated approaches predicting the credibility of topics [1] and events [2, 3] by engineering complex features based on data and meta-data for tweets as well as their social structures. Morris et al. [6] study user perception of information credibility on Twitter and show that users rely on tweet contents and other heuristics for credibility judgments. On the other hand, it is established that Twitter posts report real-time news overlapping with the reported news in newswire with the addition of minor and local news not reported by other sources [8]. Despite the real-time news post, Twitter users are more concerned with the credibility of tweets relating to breaking news, politics, and disaster events [6].

2 Methodology

We design a user study based on the CrowdFlower[2] crowd source platform to examine user perception of the credibility of tweets for news events. We select news event topics, and their relevant tweets for our study. We recruit crowd source evaluators to judge the credibility level of tweets, and leave comments on their judgments. Through their comments, we extract the features and apply predictive association rule analysis [9] to establish the associations between features and credibility levels.

In this research, credibility is defined as *"the quality of being believed or accepted as true, real, or honest"*. [3] The criteria to determine credible tweets [1] are that they must affirm a fact, be informative for the public, not be self opinionated, and not be a chat between friends. To ensure that relevant and truthful news is used in our dataset, we selected twenty news event-related topics (judged by the authors) based on major news recently reported on on-line newswire including BBC, Reuters, CNN, Guardian, and The New York Times. The news events occurred between 1 June 2013 and 15 October 2013. Table 1 describes the twenty topics we selected. Tweets were collected based on the search API of Twitter, using the news event topics shown in the left column of Table 1 as query terms. To ensure that we do not include redundant tweets, directly retweeted messages are excluded. In total, 400 credible tweets in English for twenty news events were presented to CrowdFlower evaluators to judge.

[2] https://crowdflower.com
[3] http://www.merriam-webster.com/dictionary/credibility

Table 1. Twenty news event topics

Topic	**News event description**
US government shutdown	US Government heads toward a shutdown
Iran-US relationship	Iranian President takes steps to thaw relations with the West
Sarin attack in Syria confirmed	United Nations confirms use of chemical weapons in Syria
Shipwrecked at Europe	Boat sinks in the Mediterranean, killing dozens
Egypt state of emergency	Egypt declares state of emergency
Train kills dozens in India	Train kills dozens of religious pilgrims in India
Navy Yard shooting	Gunman and 12 victims killed in Washington D.C. Navy Yard shooting
Earthquake in Pakistan	Magnitude 7.7 earthquake kills at least 327 in Pakistan
Terrorist attack mall	Somalian militants terrorize luxury mall
Military ousted president	President Morsi deposed by military after one year in office
NSA whistle blower	Edward Snowden: whistle-blower behind NSA surveillance revelations
UK new prince	The Duchess of Cambridge gives birth to a baby boy
Oil train derails	A train in Quebec derails and explodes
Colorado flood	Colorado flood 2013 tragedy
Australia's new prime minister	Australia's new Prime Minister Tony Abbott
Iraq suicide attacks	Suicide bomb attacks on Iraqi school, Shi'ite pilgrims, kill 29
Mexico storm disaster	Mexico storms death toll rises, crop lands damaged
Cyclone hits India	Many evacuated as Powerful Cyclone Hits India
Protest in Egypt	More than 50 people are killed as pro-Morsi protest
Riot in Moscow	Rioting erupts in Moscow after killing blamed on migrant

In the crowd source evaluation, the date, topic and topic description of each tweet are given to the evaluators to help them distinguish the credibility level of tweets. The credibility definition and criteria are also presented to the evaluators. To trap unreliable evaluators, gold questions are set up, which are credible news tweets mingled with not credible tweets containing opinions or social chats. For each of the twenty topics, two gold questions are randomly inserted into the credible tweets. Only evaluators that judge the gold questions correctly are considered reliable, and their judgments are accepted.

To further elicit the features the public uses to judge the tweet credibility, we also ask CrowdFlower evaluators to leave textual comments to explain their judgements. We manually examine the comments to ensure quality comments are used to analyse user perception. To this end, we remove nonsensical comments, such as those containing the word "none", numbers or words that are out of context for the topic.

Table 2. Distribution of credibility ratings for 400 tweets

Credibility level	**#comments**
Definitely credible	342 (85.5%)
Seems credible	2 (0.5%)
Not credible	35 (8.75%)
Can't decide	0 (0%)
No consensus rating	21 (5.25%)

Table 3. Features derived from user comments for credibility rating

Category	**Feature**	**#cmts**
Topic-based	Topic keyword - *e.g. Prince (UK new prince topic)*	315 (54%)
Message-based	Link in tweet - *URLs, URL shortener, image links*	95 (16.3%)
User-based	Display name - *Twitter ID e.g. BBCNews, Anonymous*	88 (15%)
User-based	User belief of the topic - *e.g. plausible, professional, it actually happened, facts, informative*	44 (7.5%)
Message-based	Credibility keyword - *e.g. Update, Breaking, Liveupdates*	26 (4.5%)
Message-based	Hashtag - *e.g. #Lampedusa, #Egypt*	8 (1.4%)
Message-based	Retweet - *Contains the letters 'RT' in the tweet messages*	6 (1%)
User-based	User mention - *e.g. @OMBPress, @cctvnewsafrica*	2 (0.3%)

3 Deriving and Analysing Features for Credibility

In our user study, evaluators were asked to judge the credibility level for each tweet as "Definitely credible", "Seems credible", "Not credible", or "Can't decide". At the conclusion of our user study, a total of 2,005 judgements by 98 evaluators for 400 tweets were collected, where five out of 400 tweets received six judgments and the rest received five judgments each. The consensus rule was used to assign credibility rating for tweets. If a tweet receives three out of five or four out of six votes for a credibility level, the message is assigned the corresponding credibility rating; otherwise no consensus credibility rating (recall that there are three credibility levels) can be reached for the tweet. Table 2 lists the distribution of credibility ratings for all tweets. Note that none of the tweets received the judgment of "Can't decide". Our results confirm that users generally trust the information disseminated on Twitter, which mirrors the findings in [1].

3.1 Analysing User Comments for Credible Tweets

We analyse user comments for 342 and two tweets received "Definitely credible" and "Seems credible" ratings to derive features users use for their credibility judgments. The comments collected from the user study consist of 558 valid comments from 22 evaluators, which describe features they feel important for their judgment of the truth and falseness for tweets. Following the categorisation in [1] and [3] we manually summarise the comments into three categories of eight features, as shown in Table 3.

Table 4. Top association rules

Association Rules	**Accuracy**
Link in Tweet=available 74 => Credible 72	97.7%
Hashtag=yes 8 => Credible 8	97.6%
Retweets=yes 6 => Credible 6	97.2%
Twitter display name=yes, User belief=yes 3 => Credible 3	96.2%
Twitter display name=yes 88 => Credible 81	91.0%
User belief=yes, Topic keyword=yes 36 => Credible 27	77.4%
User belief=yes 44 => Credible 33	76.7%

Note that "User belief of the topic" refers to user's prior belief on the relevant topic and is external to Twitter, while in [1] all features are derived based on Twitter. Table 3 shows that users perceive these features in general with significantly different weights, where Topic keyword is commonly used and User mention is rarely used. In contrast the carefully engineered tens of features in [1] are used collectively by machine learning models for predicting topic credibility.

3.2 Analysing Misjudged and Difficult-to-judge Tweets

We analyse the 35 tweets with the "Not credible" rating in Table 2. These tweets are misjudged by evaluators, as all tweets in our study have been manually verified as credible. The politics news topics 'Iran and US relationship' and 'US Government shutdown' have the largest number of misjudged tweets. We observe that these tweets are often questions, which may be why users have misperception of their credibility; indeed they are titles for news articles from reliable news agencies with short url links. Although Link in tweet is an important feature for users to judge credible tweets (See Table 4), the language features of tweets also play important roles for user perception of credibility.

We also analyse the 21 difficult-to-judge tweets where users could not reach consensus ratings. 95.6% of these difficult tweets are breaking news (42.8%) and politics news (42.8%). We observe that these tweets mostly lack links to external sources, which result in that users can not consistently judge their credibility. This Link in tweet and tweet credibility association is also shown in Table 4.

3.3 Feature and Credibility Association Analysis

To uncover relationships between features and tweet credibility, we apply association rule mining to the 379 tweets in Table 2 with consensus ratings of Definitely credible, Seems credible and Not credible based on the features in Table 3. We use the WEKA Predictive Apriori package [9][4] to mine for the best 100 association rules of the form "feature set => credibility" with an accuracy threshold of 70%. Table 4 lists the top association rules, where numbers of comments supporting the left and right hand sides are shown. According to the table,

[4] http://www.cs.waikato.ac.nz/ml/weka/

for all top rules the right hand side is always the Credible rating – users tend to believe in the information conveyed in tweets yet can not reach the Not credible rating consistently. Moreover, Link in tweet, Display name and User belief are important features often leading users to the Credible rating for tweets. From Tables 3 and 4 it can be seen that the Topic keyword feature, the most important feature commented by evaluators, does not form a strong association rule; only when combined with User belief it gives high accuracy for predicting credible tweets. The Link in tweet feature is used as an indicator for credibility.

4 Conclusions

We have studied the user perception of credibility for news tweets on Twitter via a user study on the CrowdFlower platform. Through analysing user credibility judgements and comments, eight features have been identified, where display name, link in tweet and user belief in the tweet topic are most important. By feature and credibility association analysis, we find strong associations between features and tweet credibility. We further find that politics and breaking news are more difficult for users to consistently reach credibility rating.

Acknowledgments

This research is partially supported by Universiti Kuala Lumpur (UniKL), Majlis Amanah Rakyat (MARA), and by the QNRF project "Answering Real-time Questions from Arabic Social Media" NPRP grant number 6-1377-1-257.

References

1. Castillo, C. and Mendoza, M. and Poblete, B.: Information credibility on Twitter. In: Proc. WWW, 675-684 (2011)
2. Gupta, A., Kumaraguru, P.: Credibility ranking of tweets during high impact events. In: Proc. PSOSM, 2-9 (2012)
3. Gupta, M., Zhao, P., Han, J.: Evaluating event credibility on twitter. In: Proc. SDM, 153-164 (2012)
4. Kang, B., O'Donovan, J., Höllerer, T.: Modeling topic specific credibility on twitter. In: Proc. IUI, 179-188 (2012)
5. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proc. WWW, 591-600 (2010)
6. Morris, M., et al.: Tweeting is believing?: understanding microblog credibility perceptions. In: Proc. ACM CSCW, 441-450 (2012)
7. O'Donovan, J., et al.: Credibility in context: An analysis of feature distributions in twitter. In: PASSAT and SocialCom, 293-301 (2012)
8. Petrovic, S., et al.: Can twitter replace newswire for breaking news? In: Proc. AAAI (2013)
9. Scheffer, T.: Finding association rules that trade support optimally against confidence. In: Proc. PKDD, 424-435 (2001)
10. Wang, A.: Don't follow me: Spam detection in twitter. In: Proc. SECRYPT, 1-10 (2010)