Journal of Theoretical and Applied Information Technology
31st December 2014. Vol.70 No.3

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

MODELING CREDIBILITY ASSESSMENT AND EXPLANATION FOR TWEETS BASED ON SENTIMENT ANALYSIS

DWI H. WIDYANTORO & YUDI WIBISONO
Institut Teknologi Bandung
School of Electrical Engineering & Informatics
E-mail: dwi@stei.itb.ac.id

ABSTRACT

Credibility is one of the main issues when dealing with information obtained from Online Social Networks (OSNs). Although a significant number of prior works have addressed many issues in this topic, only a few that have worked on methods for automatic credibility measurement for OSN messages and almost none who has addressed a specific problem in explaining the credibility information. This paper proposes a new approach for modeling credibility of tweets and explaining it to users. We model tweet credibility based on other independent tweet contents that support and oppose the topic issue in question. We also consider the opinions of tweets' followers who either confirm or deny, along with their reputations. This method is based on assumption that the more community is who agree with the claim, the more credible is the claim. Explanation of the credibility measure is based on the tweet content itself. We provide users with representative tweets that can be progressively zoomed-in for more detail explanation. To achieve this goal, all tweets are hierarchically structured and tweet representatives on each node are selected from the ones that are most similar to the centroid. Our evaluation results indicate the feasibility of the proposed methods.

Keywords: *Credibility Assesment, Credibility Explanation, Online Social Netwok, Tweet, Sentiment Analysis*

1. INTRODUCTION

Online Social Networks (OSNs) such as Facebook and Twitter are among the most innovative and 'killer' Internet applications in the first decade of twenty first century. These applications are not only accessible from various platforms, their users have also immensely grown in the past few years. For examples, the Twitter's users were about 100 millions in 2011 [1] while the users of Facebook have reached 955 millions in 2012 [2]. The widespread availability of OSNs has also changed the way people communicate with one another. People can now easily share their views and opinions to friends as well as disseminate information to community they care for, rapidly generating the most up-to-date information to the virtual world.

Because of the vast amount of new information produced from OSNs, it becomes one of the most preferred information sources. Its role as an information source stands out particularly during crisis situations such as natural disaster and terrorism [3, 4, 5, 6 7]. Breaking news about critical and important events spreads very quickly and is available earlier through OSNs than through conventional news media.

The credibility of information obtained from OSNs, however, is questionable. Unlike information from conventional news media that generally has been verified, information from OSNs is much less reliable since it could come from anyone without undergoing editorial process and facts verification. In addition, OSN users who report or produce information could have their own biases, perceptions and purposes, making it less trusted. Even worse, inaccurate but interesting information that is repeatedly reported and is spread out rapidly could eventually change the readers perception that the information is indeed true [8]. In other cases, OSNs have also been used to spread out spams [9] and rumors [3, 10, 11]. Interview with thirteen organizations of international humanitarian relief suggests that information from

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

OSN cannot be used for decision-making even though there is a strong desire to use it [12].

Measuring the credibility of information can address the above problem. A piece of information is considered credible if it is trusted, reliable, neutral and fair [13]. Although credibility models targeted for OSN users have been developed by [14, 15, 16, 17], most of these models are not truly informative to end-users because it only indicates whether an information is credible or not, or somewhere in between. Without reasonable explanation, it is hard to get user trusts for the given credibility information. In addition, most of these models employed supersived [14, 15, 16] or semi-supervised [17] learning approaches. Despite the success of these learning approaches in many classification tasks, it might not be the best approach for predicting the credibility of a new issue in OSNs because the only way to assess its credibility is through verification process (i.e., check and re-check) from various reliable sourceses such as common practices performed by reputable news media.

We propose a model for credibility assessment and explanation that capitalizes online social networks. Given a particular topic issue, a system with this model will retrieve user sentiments and responses on the issue from OSN (e.g., Twitter's micro-blogging) as the main information sources to evaluate its credibility. We then use representative user opinions that can be progressively zoomed as a means to explain the credibility value. This paper offers two main contributions. First, we introduce two special types of sentiment, which are *supporting* and *opposing* opinions, for credibility level measurement. It assumes that information is credible if more people are in agreement with its content, and vice versa. Second, we develop a model for credibility explanation based on representative user opinions that are hierarchically organized, which allows dynamic zooming of supporting and dissenting opinions. Agglomerative clustering is employed to hierarchically cluster the user opinions and the representative opinions at every level in the cluster are selected based on the similarity with cluster center .

The rest of this paper is organized as follows. Section 2 describes related work on credibility models. Sections 3 & 4 provide description of our approaches for modeling credibility and for generating explanation, respectively. Section 5 discusses the evaluation of our approaches, followed by conclusion in Section 6.

2. RELATED WORK

Credibility is a multi-dimensional concept [18] involving factors such as source (i.e., trustworthiness, expertise, credential), receiver (i.e., prior knowledge of the issue, issue involvement, issue relevance), message (i.e., topic, supporting data, familiarity), context (i.e., distraction, timeline) and media characteristics (i.e., organization, usability).These factors may interact with one another. For example, the sources can have effects on the message factor and the receiver factor. In computer-based media, credibility factors also include interface design, loading speed, accessibility and interactivity.

Automatic assessment of information credibility can be broadly categorized into two approaches. The first approach is to use hand-crafted patterns to identify disputed or confirmed claims. For examples, patterns such as 'it is not true that S' or 'false claims that S' are indicative of disputed sentences. Ennals [19] employed bootstrapping-like algorithm to semi-automatically identify the pattern. They started with a set manually crafted pattern. Additional patterns are then manually identified from text with known disputed claims by observing common prefix on those texts.

Another approach is to employ supervised learning such as Decision Trees [14, 20] and Bayesian Classifier [10, 16] for classifying if a claim is credible or not as well as SVM-Rank [15, 17] for providing crediblity ranking of tweets. Castillo [14] explored supervised learning techniques and tweet features that affect its credibility. In their experiment, they found that J48 decision tree provides the best results compared to SVM, Decision Rules and Bayes Networks.

Castilo [14] found that the main tweet features that affect credibility include how long have been twitter users, the number posting, the number of friends/followers and the number of re-tweets. Donovan et al. [21] showed that on eight separate event tweets, the best indicators of credibility were URLs, mentions, retweets and tweet length. Meanwhile, Morris et al. [22] reported that users tend to be biased to information that are visible at a glance such as username and picture of a user. An analysis by Yang et al. [23] on two mirco-blogging websites revealed that network overlap features and location had the most influence on determining the credibility perceptions of users. They also found that different culture has different sensitivity to the context of event. Gosh et al. [24] had been able to

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

use features based on user-created list to predict topic-based Experts on Twitter.

Rumor detection is a problem similar to identifying information whose credibility is questionable. Related problem is believe classification, which identifies users who confirm or deny the misinformation. Qazvinian [10] investigated these two problems using statistical models and maximize a linear function of log-likely hood ratio. They experimented with content-based features, network-based features and twitter specific memes and concluded that content-based features gave the best results.

To our knowledge, little prior work (if any) has investigated the explanation in information credibility. Explanation facility can be designed for various purposes such as transparency, scrutability, trustworthiness, effectiveness, persuasiveness, efficiency, or satisfaction [25].

3. CREDIBILITY ASSESSMENT MODEL

Our model of credibility measure is based on the following two observations from prior works: (1) an information is more likely to be true if more users confirm it [26], and (2) rumor is likely disputed by other users [3]. Therefore, a piece of information is considered more credible if there are more users who agree with it than those who deny it. In this approach, the credibility of a tweet mentioning a topic issue in its simplest form is determined by the difference between the number of users who confirm and dispute it. In this setting, a user is defined by followers who respond directly or indirectly to the tweet. Because each user can have different reputation, each user can be weighted accordingly. This model is then extended by also considering other tweet postings that agree and disagree with the topic issue.

Let t={*content-based features, non-content-based features*} be the feature vector of tweet t whose claim is to be assessed. To formally define the credibility of t, we introduce *Support* (t) and *Oppose*(t), which are values associated with on-line community votes who support and oppose the claim t, respectively.

Let also *Agree*(t) be a set of other independent tweets with the same claim, *Confirm*(t) & *Deny*(t) be the set of followers of m that confirm and deny the m's claims respectively where $m \in \{t \cup Agree(t)\}$ and $R_i = (0, 1]$ be the reputation of follower i. The support provided by on-line community is defined as follows.

$$Support(t) = \sum_{m \in Agree(t)} \left(1 + \left(\sum_{i \in Confirm(m)} R_i - \sum_{j \in Deny(m)} R_j \right)\right) \quad (1)$$

Likewise, let *Disagree*(t) be a set of other independent tweets with the opposite claim on the same topic, *Confirm*(n) & *Deny*(n) be the set of followers of n that confirm and deny the n's claims respectively where $n \in Disagree(t)$. The weighted votes from on-line community that oppose the claim of tweet t can be defined similarly as follows.

$$Oppose(t) = \sum_{m \in Disagree(t)} \left(1 + \left(\sum_{i \in Confirm(n)} R_i - \sum_{j \in Deny(n)} R_j \right)\right) \quad (2)$$

It is obvious from Equations 1 & 2 above that followers who deny the opposing tweets are basically the same as those who confirm the supporting tweets. Similarly, followers who deny the supporting tweets will contribute to the vote for opposing tweets.

The credibility of claim of tweet t is then defined by:

$$Credibility(t) = \frac{Support(t) - Oppose(t)}{Support(t) + Oppose(t)} \quad (3)$$

The denominator performs normalization and hence, the range of credibility value is [-1,1]. Positive values indicate that the claim is credible to some degree and highest level of credibility is obtained when there is no other independent tweets that oppose the claim. A claim tends to be rumor (not credible) if its credibility value is negative, i.e., more independent tweets that oppose the claim than those that support it. In the case of negative value, the degree that the claim is not credible is given by $|Credibility(t)|$.

In order for a tweet's content to be highly credible, it must have a lot more supporting evidences than opposing ones. In the case that the reputation of all followers are assumed to be equal, the credibility level is determined mainly by the difference of followers that confirm from those that deny the tweet content. If the follower reputation can be correctly modeled, it could be possible to obtain high credibility level by requiring only a few numbers of highly reputable followers that confirm the tweet's content.

Our approach to model tweet's credibility based on sentiment analysis as described above is similar to the one developed by Ikegami et al. [27] that is based on the majority decision of two contrastive opinions. Unlike their majority decision approach

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

that results in binary conclusions (either credible or not credible), our credibility model also provides an information that can be interpreted as to what degree the claim is credible or not credible, that can be directly used as the membership function for linguistics variable in fuzzy model (whenever needed). In addition, our model has incorporated the factor of user's reputation, which is ignored in Ikegami et al.'s credibility model.

The process of identifying tweets returned by *Agree* and *Disagree* functions is as follows:

1. Retrieve tweets based on text with the topic issue t: $T \leftarrow Retrieve(t, Twitter)$
2. Select only the most similar tweets: $ST = \{st \mid st \in T \text{ and } Similar(st, T) \geq Thresold\}$
3. Label each $st \in ST$ whether it agrees or disagrees with the claim of topic issue t, i.e., $Agree(t) = \{t_a \mid t_a \in ST$ and t_a claim is the same as t claim, and $Disagree(t) = \{t_d \mid t_d \in ST$ and t_d claim is the same as the **negation** of t claim.

The range of threshold in Step 2 is (0,1) and can be empirically determined. The threshold should be set high enough to guarantee that the two tweets are of at least very similar in topics. Given two tweets with the same/similar topics, the main issue in Step 3 is to identify if the two tweets have contradicting (*Disagree*) or the same (*Agree*) claim.

Assuming that two claims discuss the same/similar topics, consider the following four possible conditions than can arise between both claims.

a) One of the claims has contradicting word/phrase that is not found on the other.
b) One of the claims has word/phrase with the meaning opposite to the meaning of word/phrase on another claim.
c) One of the claims satisfies both conditions.
d) None of the claims satisfies above conditions.

Two claims are considered *Disagree* (contradictive) if the claims satisfy exactly one condition (a) or (b) and not both. Furthermore, both claims are considered in agreement (*Agree*) if it satisfies either condition (c) or (d). Note the conditions (c) and (d) are mutually exclusive. Condition (c) is basically double negations (i.e., negation of the opposite meaning).

In this paper, we intentionally maintain a generic model of credibility assessment because the performance of the model depends greatly on the model instantiation, i.e, the specific method that is implemented in the model's component. In particular, the model's component that affect the model effectiveness include (1) features that are incorporated to represent a tweet, (2) methods employed to determine tweet's topic similarity, (3) specific algorithms for identifying contradictive claims, and (4) parameters involved for calculating the user reputations. In the following we will discuss various altenatives of existing works that can address these issues.

3.1 Tweets Features

Although various features for tweet represenation have been introduced in the past, they basically can be broadly divided into two categories: *conten* and *non-content* based features. The content-based features can consist of unigram, bigram and/of trigram of terms/tokens that occur in the tweets. Because sentences in tweets are mostly informal, pre-processing is usually applied to normalize its content such as resolving various abbreviation, identifying emoticons, etc. The non-content based features can be #follower, #retweet, the network structure, etc.

3.2 Topic Similarity Measure

One of the most widely studied methods for topic similariy measure is the one based on the vector space model. In this model, a tweet is represented as a bag-of-words where TF-IDF (Term Frequency-Inverse Document Frequency) as well as Cosine function is the most common and effective methods for term weighting and topic similarity measure, respectively.

3.3 Contradiction Identification

One of approaches to test the presence of contradicting word as in the condition (a) above is to use various patterns with lexical cues that can indicate a contradictive claim. For example, a claim with pattern "*X Y*" contradicts with claim "*X S Y*" where S ∈ {"did not", "is not", "don't", "haven't", …}. In condition (b) , WordNet's antonym can be employed to check word with opposite meaning. For example, "OB wins the 2013 election" has the opposite meaning with "OB loses the 2013 election". In the above example, "win" has the same meaning with "does not lose", which satisfies both conditions. Table 1 provides examples of cue phrases for confirmation and denial. The problem contradiction identification can also be approached using textual entailment. Two tweets of the same topic is considered contradictive if one of the tweets is not an entailment of the other. Various textual entailment algorithms have been developed such as those based on syntactice similarity,

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

symbolic meaning, logic-based approach, surface string, vector space model, rule extraction and combination of these approaches.

Table 1: *Cue Phrases for Confirmation and Denial*

Confirmation	Denial
• confirmed • its true • so true • believe that • truth that	• it is apparently a fake • still rumors floating around • Who believes stuff like that? • its not true that • be careful what you read • misinformation about

3.4 Follower Reputation

The last issue to address is modeling the reputation of information sources (e.g., follower). A lot of attributes can be considered to estimate the reputation of information sources such as maturity (active period of information source), authority, influence, social network structure, history of issuing/forwarding correct information, bias to political situation, the number of posting, etc. In the case of no other information is available, all users' reputations are set the same values, e.g., to value one. Interestingly, although estimating the reputation of information sources is an interesting and important problem, there is still a handful literatur that discussed this problem, opening up many research problems to address.

4. CREDIBILITY EXPLANATION

The role of explanation in information credibility is similar to that of in Expert Systems and Recommender Systems [28] in that it serves as a tool to help decision-making. When a user receives a news from OSNs, the explanation system will provide credibility assessment along with its justification. It offers at least two main benefits: (1) users will understand the reasoning behind the system's credibility assessment so they can asses their degree of confidence to the system's output and (2) users will be knowledgeable with the system's strengths and limitations, which help earn users' acceptance even in the situation where they do not agree with the system.

How credibility can be explained depends on how it is measured and what factors are involved. Accordingly, our credibility explanation method takes benefit a lot from all information obtained during computing the credibility measure. In particular, we capitalize data comprising of sets of tweets whose topical content support or opposed the claim in quesion along with statistical informations about their followers who confirm and deny the mentioned topics.

Rather than presenting a set of keywords deemed representative to the topic issue (which tends to confuse users), we provide explanation by presenting complete tweets' contents that are representative enough for each tweets category (i.e., either supporting or opposing claim). To avoid information loss, we arrange each tweets group category into a hierarchical structure. Each node in the hiearchy contains a tweet that is representative for all tweets underneath. Therefore, all leaves in this hierarchical structure are individual tweets invloved in calculating the credibility measure.

Any hierarchical clustering algorithm (either divisive, agglomerative or their variants) can be employed to hierarchically structure the tweets for each category. During the process, whenever a new cluster at a hierarchy level is formed, a tweet is selected and the statistical information about its followers is calculated. We select representative tweets from among those that maximizes its similarity with centroid. The selected tweet along with its statistical information will represent the newly formed cluster.

Credibility explanation is provided by initially presenting the representative tweets from the root node (zero-level hierarchy). More detail explanation can be obtained by zooming-in the representative tweets from nodes at the next levels. For each representative tweet, users also can view its statistical information. The fact that tweets are short in length and mostly contain concise text are of great advantage in the clustering task. Unlike a long, fulltext message, no post text processing is needed in order to generate concise explanation texts.

5. EVALUATION

In this section we provide an evaluation of our proposed credibility model and explanation with two main objectives. The first objective of evaluation is to assess the effectiveness of the proposed credibility model. Because of the lack of common data for evaluating credibility model and it it is very difficulut (if not impossible) to consider all cases for the evaluation, We want to show that at least there exists a real event on tweet data in which the credibility model can fit. To achieve this, we use a test case of real tweets in a popular case (whose ground truth has already been well known by public) and measure how the credibility level measured by the main component of the

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

proposed model agrees with the known truth. The second objective is to evaluate the effectiveness of our proposed credibility explanation approach in clarifying the reasons behind various opinions.

5.1 Credibility Model Evaluation

To evaluate the model's effectiveness, we collected tweets on a topic related to "Sandy Hook School Shooting". This topic was triggered by a video on YouTube stating that incidence in Sandy Hook shooting is a hoax created by the government as the pretext for gun control. The ground truth was that the shooting did happen based on reliable sources such as CNN, NYTimes and Reuters. Twitter's API search was used to retrieve tweets during 14-17 January 2013 period. A total of 2150 tweets were identified to contain the "Sandy Hook School Shooting" out of a total of 235737 retrieved tweets. Of this number, we identified around 116 disputes, consisting of 66 independent tweets that agree with the video statement and 50 independent tweets that disagree. The rest of the tweets contain mostly neutral statements. For each independent tweet, we trace its followers and identify its response whether it confirms or denies the original tweet statements.

The following are examples of tweets that agree and disagree with the statements on video.

Agree: "So the Sandy Hook shooting was a huge hoax. Actors, one big movie set. Made by OB. I am sickened right now. I'm about to share the video".

Disagree: "If you think Sandy Hook was a hoax so OB could bust into your basement arms depot and steal your cannons, you disgust me. Unfollow me NOW"

Table 2: Statistics of samples based on "Sandy hook shooting was a hoax" tweet.

	#Independent Tweets	#Confirming Followers	#Denying Followers
Agree	66	131	9
Disagree	50	85	4

Table 2 provides the statistics of the samples of tweets. Independent tweets represent originating tweets. Confirming (Denying) followers columns contain the total numbers of followers of independent tweets that confirm (deny) the content of independent tweets. In this samples we do not have any information about the reputation of followers, so we can assign all the reputation values to one (i.e, $R_i = 1$). Hence, based on Equations (1) & (2),.

Support("Sandy Hook shooting is a hoax")

$$= \sum_{1}^{\|Agree()\|}\left(1+\left(\sum_{1}^{\|Confirm()\|}1-\sum_{1}^{\|Deny()\|}1\right)\right)$$

= 66 + 131 – 9 = 188, and

Oppose("Sandy Hook shooting is a hoax")

$$= \sum_{1}^{\|Disagree()\|}\left(1+\left(\sum_{1}^{\|Confirm()\|}1-\sum_{1}^{\|Deny()\|}1\right)\right)$$

= 50 + 85 – 4 = 131, therefore

Credibility ("Sandy Hook shooting is a hoax")

$$= \frac{188-131}{188+131} = 0.18$$

which is a very small value (i.e, its credibility is questionable). The fact that the incidence is not a hoax confirms the agreement of calculated result based on the proposed credibility model with the ground truth. At least it shows that there exists real tweet data that the proposed model can be used to provide approximately correct assesment of the tweet's credibility level.

It is important to note that the credibility level calculated by the proposed model and its interpretation are limited only to the tweet in question. Specifically, a closely-related hyphothetical statement cannot be inferred from the credibility level of tweets being discussed. For example, the conclusion that the credibility of "Sandy hook was a hoax" is low (calculated based on tweet's responses) does not imply that a hypothetical statement "Sandy hook was not a hoax" has high credibility level. The only way to assess the credibilty level of a statement is by examining the opinions given by its responses.

Observation of tweets data also reveals that followers mostly agree with the originating tweets. It explains why the portion of the number of confirming followers is much larger (encompassing about 94%) than that of the number of denying followers for a given opinion.

Journal of Theoretical and Applied Information Technology
31st December 2014. Vol.70 No.3

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

5.2 Evaluation of Credibility Explanation

To evaluate the effectiveness of credibility explanation, we use the same data set as the one for credibility model evaluation. For each tweet groups, we apply agglomerative hierarchical clustering. It is a bottom-up strategy that iteratively merges clusters into larger clusters from each object in its own cluster until all objects in a single cluster (Han & Kamber, 2001). The following outlines the agglomerative clustering algorithm.

1. Represent each tweet as *tf-idf* vector based on its content-based features.
2. Create initial singleton cluters from individual tweets.
3. Repeat until number of clusters=1:
   a. Calculate centroid of each cluster
   b. Calculate similarity between clusters by using cosine similarity
   c. Merge each two closest clusters into one cluster, and determine its representative tweet.
4. For each cluster at each level, select a set of top *k*-representative tweets based on tweet's similarity with the centroid.

This stage returns a hierarchical tree of representative opinions. The following are examples of explanation tree for each opinion group.

Explanation tree of supporting group:

Sandy hook was a hoax!

-- Sandy hook was a hoax! OB has some tricks up his sleeves (14)
  -- 9/11 was a hoax. **The BU administration blew up the buildings. Sandy Hook is OB's 9/11** (7)

-- I'm hearing things that **Sandy Hook was fake and Obama did it to enforce better gun laws**? (12)
  -- did you hear there were like fake actors there trying to support OB its weird and sandy hook is in the dark knight (6)

Explanation tree of opposing group:

Everyone so concerned with **"is OB really Muslim"** and **"was sandy hook a hoax?"** (8)

-- WAYMENT! What's this about Sandy Hook being a "hoax" and an OB set up on my TL?! **What's going on here?!** (5)
  -- So people think OB set up sandy hook to make gun laws and espn is showing a press conference about a fake girlfriend I'm confused (5)

-- **Sandy Hook 'truthers' warn about OB gun plan. Have we entered an alternate ..**(6)
  -- The last 3 posts I've seen from one guy on Facebook are two Sandy Hook conspiracy posts and one "proving" OB's birth certificate is fake. (5)

The top-level tree in the supporting group often mentions "Sandy hook was a hoax" confirming that they agree with it. The second level of tree starts providing more information that can be considered as an explanation of why they believe it was a hoax. Specifically, they suspect that it is incumbent tricks to earn people support in the similar way as the 9/11 on WTC incidence, as well as an effort to push legistative proposal (i.e., better gun law). As for the tree of opposing group, the disagreement is expressed as satire such as why sandy hook was a hoax if it was contradictory with their (those who agree) own concerns.

The above examples of cluster results with representative tweets qualitatively show that the proposed explanation approach can provide a snap shot for clarifying the reasons for or against a tweet's content in question. Although not perfect, it is a lot improvement over manually scanning hundreds of tweets' contents.

6. CONCLUDING REMARKS

This paper has described our methods for measuring and explaining credibility. We use sentiment analysis of tweets and the opinions of followers for modeling the degree to which the orignial tweet in question is credible or not credible. We also suggest the use of hierarchical agglomerative clustering, in conjunction with the use of representative tweets in a cluster as a means to provide explanation of why people agree or disagree with a certain topic. Our evaluation of the proposed method on a real, well known controversial case show the agreement of our approach with the case ground truth. The tweets contents selected by the proposed method can provide sensible reasoning that explain the position of each opinion.

Methods that capitalize community opinios about a topic issue has the advantage that it also provides richfull information that can be used to explain the credibility of a piece of information. The main drawback of this approach is in a situation where there is a shortage of users who respond or comments about the issue, particularly for a new posted issue on Online Social Network.

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

Information about the degree to which a tweet is credible is not enough for convicing users. Explanation system as described in this paper helps users understand the reasoning behind the system's credibility assessment and can earn system's trust from users eventhough the users might not agree with the system's assesment. Although has not been fully tested, our aproach that provides users with hierarchical representive tweets as a means for explaining the degree of credibility seems promising.

One of the crucial components in our credibility model that has not been fully addressed is modeling the tweet users reputation. Being able to identify highly reputable, authoriative tweet for a particular topic will save a lot of time and effort as well as significantly improve the model prediction accuracy. This will be the subject of our future work.

AKNOWLEDGEMENT

This work is in part supported by the Decentralization Research Grant #603/AL-J/DIPA/PN/SPK/2014 provided by the Directorate General of Higher Education, Republic of Indonesia. Views expressed in this paper are those of authors and do not necessarily reflect those of the funding agency.

REFERENCES

[1] Twitter Blog, One hundred million voices. twitter.com/2009/06/down-time-rescheduled.html http://blog.twitter.com/2011/09/one-hundred-million-voices.html

[2] U.S. Securities and Exchange Commission, *Facebook Current Report*, Form 8-K, Filing Date July 26, 2012

[3] Mendoza, M.;Poblete, B. & Castillo, C. Twitter Under Crisis: Can we trust what we RT? *Proceedings of the First Workshop on Social Media Analytics,* 2010, 71-79

[4] Sakaki, T., Okazaki, M. and Matsuo, Y. "Earthquake shakes Twitter users: real-time event detection by social sensors". *Proceedings of the 19th international conference on World wide web*, ACM, 2010, 851-860.

[5] Vieweg, S., Hughes, A., Starbird, K. and Palen, L. "Microblogging during two natural hazards events: what twitter may contribute to situational awareness", *Proceedings of the 28th international conference on Human factors in computing systems*, ACM, 2010, 1079-1088.

[6] Starbird, K., Palen, L., Hughes, A. and Vieweg, S. "Chatter on the red: what hazards threat reveals about the social life of microbloggedinformation", *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, ACM, 2010, 241-250.

[7] Hughes, A. and Palen, L. "Twitter adoption and use in mass convergence and emergency events", *International Journal of Emergency Management,* (6:3), 2009, 248-260.

[8] Ratkiewicz, J.; Conover, M.; Meiss, M.; Goncalves, B.; Flammini, A. &Menczer, F. "Detecting and tracking political abuse in social media", *Fifth International AAAI Conference on Weblogs and Social Media,* 2011.

[9] Grier, C.; Thomas, K.; Paxson, V. & Zhang, M. "@ spam: the underground on 140 characters or less", *Proceedings of the 17th ACM conference on Computer and communications security,* 2010, 27-37.

[10] Qazvinian, V.; Rosengren, E.; Radev, D. & Mei, Q. "Rumor has it: identifying misinformation in microblogs", *Proceedings of the Conference on Empirical Methods in Natural Language Processing,* 2011, 1589-1599

[11] Ratkiewicz, J.; Conover, M.; Meiss, M.; Gonçalves, B.; Patil, S.; Flammini, A. &Menczer, F. "Truthy: Mapping the spread of astroturf in microblog streams", *Proceedings of the 20th international conference companion on World wide web*, 2011, 249-252

[12] Tapia, A., Bajpai, K., Jansen, B., Yen, J. and Giles, L. "Seeking the trustworthy tweet: Can microblogged data fit the information needs of disaster response and humanitarian relief organizations", *Proceedings of the 8th International ISCRAM Conference*, 2011, 1-10.

[13] Hilligoss, B. &Rieh, S. "Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context", *Information Processing & Management*, Elsevier, 2008, 44, 1467-1484.

[14] Castillo, C.; Mendoza, M. &Poblete, B. "Information credibility on twitter", *Proceedings of the 20th international conference on World wide web,* 2011, 675-68.

[15] Aditi Gupta and Ponnurangam Kumaraguru. Credibility ranking of tweets during high

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

impact events. In *Proc. 1st Workshop on Privacy and Security in Online Social Media*, PSOSM '12, ACM, 2012.

[16] Xin Xia, Xiaohu Yang, Chao Wu, Shanping Li, and Linfeng Bao. Information credibility on twitter in emergency situation. In *Proc. Pacific Asia conference on Intelligence and Security Informatics*, PAISI'12, 2012.

[17] Gupta, A., Kumaraguru, P., Castillo, C., & Meier, P. (2014). TweetCred: A Real-time Web-based System for Assessing Credibility of Content on Twitter. *arXiv preprint arXiv:1405.5490.*

[18] Wathen, C. & Burkell, J. "Believe it or not: Factors influencing credibility on the Web", *Journal of the American society for information science and technology, Wiley Online Library,* 2002*,* 53, 134-144.

[19] Ennals, R.; Byler, D.; Agosta, J. & Rosario, B. What is Disputed on the Web? *Proceedings of the 4th workshop on Information credibility,* 2010, 67-74.

[20] Yang, C.; Harkreader, R. & Gu, G. "Die free or live hard?empirical evaluation and new design for fighting evolving twitter spammers", *Recent Advances in Intrusion Detection*, 2011, 318-33.

[21] J. O'Donovan, B. Kang, G. Meyer, T. Hllerer, and S. Adali. Credibility in context: An analysis of feature distributions in twitter. *ASE/IEEE International Conference on Social Computing*, SocialCom, 2012.

[22] Morris, M.R., Counts, S., Roseway, A., Hoff, A., and Schwarz, J. Tweeting is believing?: Understanding microblog credibility perceptions. In Proc. CSCW. ACM, 2012.

[23] Yang, J., Counts, S., Morris, M.R., and Hoff, A. Microblog credibility perceptions: Comparing the usa and china. In *Proceedings of Computer Supported of Collaborative Work*, 2013, 575–586.

[24] Ghosh, S., Sharma, N., Benevenuto, F., Ganguly, N. and Gummadi, K. Cognos: crowdsourcing search for topic experts in microblogs. In *Proceedings of SIGIR*, 2012.

[25] Tintarev, Nava, and Judith Masthoff. "Effective explanations of recommendations: user-centered design", *Proceedings of the 2007 ACM conference on Recommender systems*. ACM, 2007.

[26] Kaczmarek, A. "Web Services Integration with Regard to the Metrics of Data Believability", *The Fourth International Conference on Information, Process, and Knowledge Management*, 2012, 28-32.

[27] Ikegami, Y., Kawai, K., Namihira, Y., & Tsuruta, S. Topic and Opinion Classification based Information Credibility Analysis on Twitter. In *Proc. of IEEE International Conf. on Systems, Man, and Cybernetics (SMC),* 2013*,* 4676-4681.

[28] Herlocker, J.; Konstan, J. &Riedl, J. "Explaining collaborative filtering recommendations", *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, 2000, 241-250.