2013 IEEE 13th International Conference on Data Mining

Prominent Features of Rumor Propagation in Online Social Media

Sejeong Kwon* Meeyoung Cha* Kyomin Jung† Wei Chen‡ Yajun Wang‡
*Korea Advanced Institue of Science and Technology, Republic of Korea
{gsj1029, meeyoungcha}@kaist.ac.kr
†Seoul National University, Republic of Korea, kjung@snu.ac.kr
‡Microsoft Research Asia, China, {weic, yajunw}@microsoft.com

***Abstract*—The problem of identifying rumors is of practical importance especially in online social networks, since information can diffuse more rapidly and widely than the offline counterpart. In this paper, we identify characteristics of rumors by examining the following three aspects of diffusion: temporal, structural, and linguistic. For the temporal characteristics, we propose a new periodic time series model that considers daily and external shock cycles, where the model demonstrates that rumor likely have fluctuations over time. We also identify key structural and linguistic differences in the spread of rumors and non-rumors. Our selected features classify rumors with high precision and recall in the range of 87% to 92%, that is higher than other states of the arts on rumor classification.**

I. Introduction

Social psychology literature defines a *rumor* as a story or a statement in general circulation without confirmation or certainty to facts [1]. Rumors are known to arise in the context of ambiguity, when the meaning of a situation is not readily apparent or when people feel an acute need for security [6]. Rumors hence are a powerful, pervasive, and persistent force affecting people and groups [5].

The spread of rumors and misinformation has been studied in the context of quantifying the credibility of a given piece of information [3] and in detecting an outbreak of misinformation [10]. With the growing popularity of online social networks and their information propagation potentials, the ability to control the type of information that propagates in the network has become ever more important.

Numerous definitions of a rumor exist. A piece of information can be considered either verified or unverified, based on the judgments made at the time of circulation. The latter, a piece of information that cannot be verified as true or false at the time of circulation (i.e., unverified), is commonly considered as a rumor in social psychology fields. In this paper, we rigorously divide the latter further into three types: true, false, and unknown, based on the judgments made *after* the time of circulation. The first type “true” represents when a piece of information that was unverified during circulation is officially confirmed as true after some time. This could be interpreted as information leakage, marketing, or prediction with enough reliable evidence. The other two types, “false” and “unknown”, which later in time get confirmed as false or remain unverified respectively, are what we define as *rumors*.

We propose a novel approach to identify rumors based on temporal, structural, and linguistic properties of rumor propagation. Our work is real data-driven. We utilize three and a half years worth of near-complete data of Twitter and extracted 104 viral events, each of which involves at least 60 posts. For the study, we employed four human coders to have each of the viral events annotated.

Based on the annotated data and guided by the theoretical studies on rumors, we analyzed the temporal, structural, and linguistic properties of rumors and non-rumors. For the temporal characteristic, we propose a new method called the Periodic External Shocks (PES) model that can describe the periodic bursts unique to rumors due to the daily cycle and the external shock cycle. For the structural characteristic, we extract properties related to the propagation process such as the proportion of isolated rumor spreaders and the proportion of propagation from low- to high-degree users. For the linguistic characteristic, we examine the word-level categories and sentiments particular to rumors such as the use of negation and negative affect words.

We built three classifiers based on decision tree, random forest, and SVM to determine whether a topic is a rumor or a non-rumor given the set of tweets about the topic. Our classifiers based on the temporal, structural, and linguistic features provide high precision and recall in the range of 87% to 92%, that is higher than other states of the arts on rumor classification.

This paper is one of the first to analyze the underlying process of rumor propagation based on annotated dataset drawn from a near-complete social media stream. While most existing work focused on the structural and linguistic features for related problems, we highlight that rumors show bursty fluctuations over time and this temporal feature has the highest predictive power. The newly proposed PES model can effectively capture the bursty temporal pattern, opening the door to new methods for identifying rumors, as temporal features are often more easily accessible than others (e.g., structural feature or linguistic feature). Our findings on structural and linguistic patterns could also stimulate further research on understanding the connection between these features and rumor propagations. Fore the wider community use, we share the annotated rumor and non-rumor datasets at http://mia.kaist.ac.kr/publications/rumor.

1550-4786/13 $31.00 © 2013 IEEE
DOI 10.1109/ICDM.2013.61

II. Related Work

A. Theories on rumor propagation.

The first insight we gain from existing research on rumor propagation is about the temporal properties of rumor spreading. Social psychologists theorized that rumormongers have a short attention, because a rumor can flourish only during a short time window when there is a need for information in the absence of news from institutional channels [12]. As a result, allegations and interrogatory statements based on circumstantial evidence often become rumors, as people seek out information while the true facts are shrouded in ambiguity [2].

The next insight is about the structural properties of rumor spreading. A study on gossip, which is oriented more towards 'inner-circle' content than rumors, reveals that denser network structures are less vulnerable to social fragmentation than sparser networks [7]. This study shows that gossips spread more widely in sparse structures. One study focused on the various roles rumormongers play, for instance, a messenger (i.e., a person who brings pertinent information to the group) and a skeptic (i.e., a person who expresses doubt over authenticity of the information) [12]. Each of these roles are expected to affect the structure of a rumor network.

The final insight is about the linguistic properties of rumor spreading. A study based on laboratory interviews suggests that rumors are expected to be dominated by certain types of sentiments like anxiety, uncertainty, credulity, and outcome-relevant involvement [11].

B. Rumors and other diffusions in OSNs.

Ratkiewicz et al. designed a system, called Truthy, to mine and visualize astroturf political campaigns in Twitter [10]. Their work proposed several key structural features that can efficiently identify astroturf memes, which then can be used for terminating the associated accounts. While not directly related to rumors, Matsubara et al. recently introduced the time series fitting model to describe the temporal pattern of a viral event with a diffusion mechanism and periodicity [8], We extend their model to better describe the temporal fluctuations of rumors and introduce a new model with additional features (e.g., periodic external shocks).

The work most similar to ours is by Castillo et al. [3], which proposed a set of features to assess the credibility of social media content. The authors proposed a way to automatically classify them. Their algorithm tries to judge the credibility of information based on a wide range of features related to the message, users, topics, and propagation pattern within Twitter. In contrast, our proposed features are drawn from extensive theories in social psychology and are almost non-overlapping with the feature set of [3], hence providing a complementary view to the problem.

III. Data

We describe the Twitter dataset and the methodology used for collecting and annotating rumor and non-rumor cases.

A. Collecting rumor and non-rumor cases.

We used the Twitter data reported in [4], which comprise the profile information of 54 million users, the 1.9 billion follow links between them, and all of the 1.7 billion public tweets posted over the course of three and a half years since the launch of Twitter in March 2006. The link information is based on a snapshot of the network at August 2009.

To collect rumor and non-rumor cases, we investigated websites like snopes.com, urbanlegends.about.com, pcmang.com and times.com. We chose a total of 130 topics (70 rumors and 60 non-rumors) that circulated during the time period covered by the Twitter dataset. For each topic, we defined regular expression and extracted every relevant tweet, based on regular expressions that were defined by consulting these websites and 4 annotators agreement. We focused on a period of 60 days starting from one day prior to a key date, which corresponds to the date when the rumor was first reported.

B. Annotation and agreement.

In order to ensure that all rumors and non-rumors are valid, we asked the 4 participants to classify each topic as either rumor or non-rumor after providing them four randomly-chosen relevant tweets and a list of URLs on the topic identified by the annotators. We then selected topics that were evaluated by at least four participants and had the majority agreement (three or four classified the same) for this study. This process yield a total of 130 topics, out of which 68 were rumors and 57 were non-rumors. The intra-class correlation coefficient (ICC), which measures the amount of agreement among participants, was 0.992 and the p-value was zero. Table I lists examples of rumors and non-rumors, respectively. The final 130 topics that are annotated had varying sizes. In this study, we further limit to only those topics that contain at least 60 tweets and as a result retained 102 topics (47 rumors and 55 non-rumors).

IV. Feature indentification

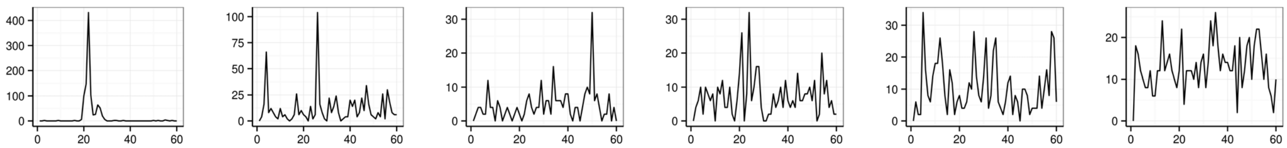
Based on the theories of rumor propagation in the existing literature, we investigate the three aspects of rumor propagation, namely temporal, structural, and linguistic properties.

A. Temporal properties

The first set of properties we investigate are temporal properties in rumor spreading. Figure 1 shows several sample time series of rumors and non-rumors. A distinct feature observed from these time series is that rumors tend to have multiple and periodic spikes, whereas non-rumors typically have a single prominent spike.

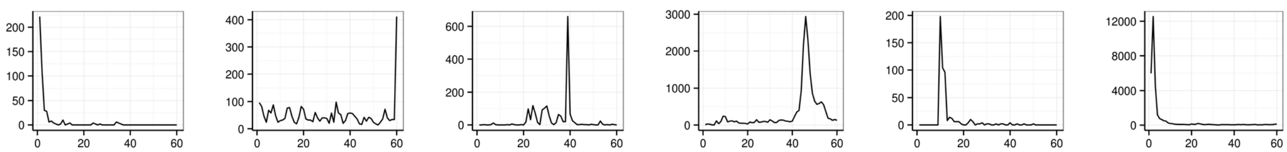
Table I: Representative rumor (Bigfoot) and non-rumor (Summize) cases and their tweet data summary

Topic	**Spreaders (Audience)**	**Tweets (Mentions)**	**Description (Regular Expression) Example tweets**
Bigfoot	462 (1731926)	1006 (40)	The dead body of bigfoot is found (bigfoot & (corpse — (dead body)) “Bigfoot Trackers Say They’ve Got a Body, I Say They Don’t”
Summize	2054 (4367672)	969 (285)	Twitter buy an IT company (twitter & buy & summize) “Twitter buying summize is BRILLIANT. I bet it powers the home screen with all the updates.”



(a) Bigfoot (b) AdCall (c) Catfish (d) HIVKetchup (e) FanDeath (f) Dork

Rumor



(g) Havard (h) Summize (i) Rockfeller (j) District9 (k) BreastIceCream (l) ObamaFly

Non-Rumor

Figure 1: Examples of extracted time series, with x-axis as days and y-axis as the number of tweets on the topic.

Our model is an adaptation of the SpikeM time-series fitting model that was proposed by Matsubara et al. [8]. The SpikeM model extends the Susceptible-Infected (SI) model in epidemiology to cover periodic spiky behavior and power-law decays observed in many time series. The full SpikeM model is given by the equations below.

SpikeM with parameters $\theta = \{N, \beta, n_b, S_b, \epsilon, p_p, p_a, p_s\}$:

$$U(n+1) = U(n) - \Delta B(n+1),$$
$$\Delta B(n+1) = p(n+1) \cdot [\frac{\beta}{N} \cdot U(n) \cdot \sum_{t=n_b}^{n} (\Delta B(t) + S(t)) \cdot (n+1-t)^{-1.5} + \epsilon] \quad (1)$$
$$p(n) = 1 - \frac{1}{2} p_a \left[1 + \sin\left(\frac{2\pi}{p_p}(n+p_s)\right)\right],$$
$$S(t) = S_b \text{ when } t = n_b; \text{ otherwise } 0.$$
$$U(0) = N, \Delta B(0) = 0.$$

In the model, $U(n)$ is the number of uninfected (or uninformed) nodes in the network at time step n; $\Delta B(n)$ is the newly infected (or informed) nodes at time n; N is the total number of nodes involved in the diffusion process; n_b is the time when the first external shock on the event occurs, and S_b is the scale of this first external shock, i.e., the number of nodes that are infected at the beginning of the event at time n_b. Thus the term$\sum_{t=n_b}^{n}(\Delta B(t) + S(t))$ represents the total number of infected nodes at time n. Then $\Delta B(n+1) = \frac{\beta}{N} \cdot U(n) \cdot \sum_{t=n_b}^{n}(\Delta B(t) + S(t))$ would be the standard SI model describing that at time $n+1$ each uninfected node u randomly picks one node v out of all the nodes and if v is already infected, u becomes infected at time $n+1$ with probability β, the parameter for infective strength.

The SpikeM model extends the SI model by introducing (a) *a power-law decay* term $(n+1-t)^{-1.5}$ in equation 1 so that the strength of infection of the earlier infected nodes becomes weaker in a power-law decay pattern, and (b) a *periodic interaction function* $p(n)$ to reflect people’s periodic interaction patterns (e.g. people may have more time to interact on Twitter in the evening than during the day when they are at work or school). Parameters p_p, p_a, and p_s correspond to the period, amplitude, and phase shift of the periodic interaction function, respectively. Finally parameter ϵ is a background noise term not interpretable by infection.

However, the SpikeM model is not appropriated for rumor analysis, because conceptually there is no parameter in the model that could explain the multiple spiky pattern of the rumors versus the single-peak pattern of non-rumors as seen in Figure 1. The periodic interaction function $p(n)$ models the cyclic interaction patterns of users caused by their daily or weekly routines, yet it is not likely to be different between rumors and non-rumors.

External shock may incur not once but *multiple* impacts over time.. For simplicity, we assume that external shocks have a short periodic cycle. Based on this assumption, we extend the SpikeM model and introduce the *Periodic External Shocks (PES)* model:

PES model with parameters $\theta' = \{N, \beta, n_b, S_b, \epsilon, p_p, p_a, p_s, q_p, q_a, q_s\}$:

$$\Delta B(n+1) = p(n+1) \cdot [\frac{\beta}{N} \cdot U(n) \cdot \sum_{t=n_b}^{n} (\Delta B(t) + \bar{S}(t)) \cdot (n+1-t)^{-1.5} + \epsilon], \quad (2)$$

$$\bar{S}(t) = S(t) + q(t),$$

$$q(t) = q_a \left[1 + \left(\sin\left(\frac{2\pi}{q_p}(t + q_s)\right)\right)\right],$$

All other terms are the same as in SpikeM.

In the PES model, we add *periodic external shock function* $q(t)$ to the initial shock function $S(t)$, and $q(t)$ has parameters q_p, q_a and q_s representing the period, amplitude, and the shift of the periodic external shock function, respectively. When $q_a = 0$, the PES model is reduced to SpikeM, hence working as a generalization of the SpikeM model.

For parameter learning, we use Levenberg-Marquard methods to minimize the sum of the squared errors: $D(X, \theta) = \sum_n (X(n) - \Delta B(n))^2$ with a given timseries $X(n)$. We will show in Section V that these new features introduced by the PES model turn out to be the most effective in classifying rumors. Table II displays the temporal features in the PES model.

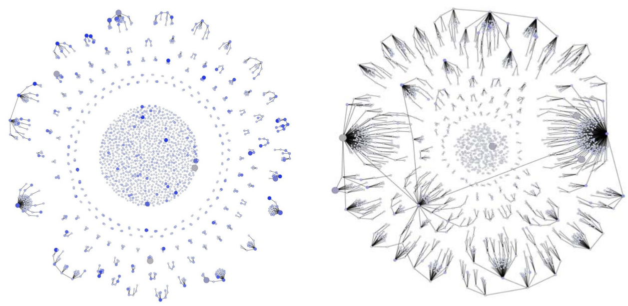
Table II: Temporal features

Symbols	Definition
N	Total population of available users
β	Probability of infection
n_b	Starting time of breaking news
S_c	Strength of external shock at birth (time n_b)
ϵ	Background noise
p_a	Strength of interaction periodicity
p_s	Interaction periodicity offset
q_a	Strength of external shock
q_p	Periodicity of external shock
q_s	External shock periodicity offset

B. Structural properties

We define the *friendship network* as the induced subgraph of the original follower-followee graph induced by those users who posted at least one related tweets and follow links among them.

Diffusion is defined as follows a topic is diffused from user A to user B, if and only if (1) B follows A on Twitter and (2) B posted about a given topic by mentioning the appropriate keywords only after A did so. In case a user has multiple possible sources, we pick the user who posted about the topic the latest as the source. Removing all the links over which a diffusion did not occur from the friendship network, then yields the *diffusion network*.



(a) Bigfoot (rumor) (b) Summize (non-rumor)

Figure 2: Diffusion network examples

Figure 2 depicts the diffusion networks of two topics: Bigfoot and Summize. Nodes represent rumor spreaders, and edges represent incidents of information diffusion. The figure shows that Bigfoot (rumor) involved a larger fraction of singletons than Summmize (non-rumor). We could find similar trends in other rumor and non-rumor cases.

From a friendship network, we also extract the largest connected component (LCC). We test the structural properties based on three network structure: the friendship network, the LCC of the friendship network, and the diffusion network. Table III shows a total of 15 structural features.

Table III: Structural features.

Symbols	Definition
V_g	Number of Nodes in the friendship network
E_g	Number of Links in the friendship network
D_g	Density of the friendship network
C_g	Clustering Coefficient of the friendship network
I_g	Median in-degree of the friendship network
O_g	Median out-degree of the friendship network
F_l	Fraction of nodes in the LCC
V_l	Number of nodes in the LCC
E_l	Number of links in the LCC
D_l	Density of nodes in the LCC
C_l	Clustering Coefficient in the LCC
I_l	Median in-degree in the LCC
O_l	Median out-degree in the LCC
S_d	Fraction of singletons in the diffusion network
F_d	Fraction of diffusion from low- to high-degree nodes

C. Linguistic properties.

We utilize a sentiment tool called the Linguistic Inquiry and Word Count (LIWC), which is a transparent text analysis program that counts words in psychologically meaningful categories [9]. Its dictionary includes around 4,500 words and word stems and the program shows the proportion of words that are related to a given sentiment in an input file. The LIWC provides five major categories and a number of subcategories in its psychological processes (e.g., social, affective, cognitive, perceptual, and biological processes). Full list available at http://www.liwc.net/descriptiontable1.php. We tested whether any of the major or subcategories appear dominantly in rumors and non-rumors.

Before applying the sentiment tool, we cleaned up the tweet data by removing usernames, short URLs, as well as emoticons that the tool could not parse. Because sentiment tools require some minimum amount of text as input (e.g., 50 words), we consolidated all the cleaned up tweets into a single file per topic for analysis.

V. Rumor Classification in Twitter

We consider the features on the rumor diffusion patterns of over time (11 *temporal* features), the shape of the diffusion network and the friendship network (15 *structural* features), and the language used in the content (65 *linguistic* linguistic features) to classify rumors in Twitter.

A. Feature selection

We used random forest and logistic model to quantify which features are most informative. For the variable selection in random forest, we used 2-fold cross-validated prediction with sequentially reduced number of predictors [15] to find the best number of variables. By this strategy, we selected 11 variables with the highest importance.Table IV displays the set of all such significant features determined from the t-test at 0.05-level (p<0.05).

Table IV: The top singificant features. In the"Type" column, "N" and "R" mean the feature had higher value for non-rumors and rumors, respectively. "RF" and "LR" columns show features selected by random forest and logistic regression, respectively.

Symbol	Definition	Type	RF	LR
Temporal features				
q_p	Periodicity of external shock	N	✓	✓
q_s	External shock periodicity offset	N	✓	✓
p_s	Interaction periodicity offset	N	✓	
Structural features				
C_g	Clustering of the friendship network	R		
D_l	Density of the LCC	R		
C_l	Clustering of the LCC	R		
S_d	Fraction of isloated nodes	R		✓
F_d	Fraction of low-to-high diffusion	R	✓	
Linguistic features				
posemo	love, nice, sweet	N	✓	✓
negate	no, not never	R	✓	✓
social	mate, talk, they, child	R	✓	
cogmech	cause, know, ought	N	✓	✓
excl	but, without, exclude	R	✓	✓
insight	think, know, consider	R		
tentat	may be, perhaps, guess	R	✓	✓
see	view, saw, seen	N	✓	✓
hear	listen, hearing	R		

We make several observations. First, the periodicity of external shock (q_p) had the highest predictive power among all features. The mechanism behind short cycle of rumors may have to do with the credibility of information. Solove [13] said that reputation gives people a strong incentive to conform to social norms and to avoid breaching people's trust. Because rumors spread without strong evidence, rumor audience may simply neglect the message, incurring low infection rate and often terminating the propagation process. As a result, rumors may rely more on the birth of new seeds (i.e., new root nodes), who are influenced by external sources (i.e., external shock). The fact that this temporal pattern is the most effective predictor for classifying rumors is novel.

Second, the structural features also had high predictive power. Among them, the fraction of information flow from low to high-degree nodes (i.e., diffusion from less influential to more influential people) was the most effective. The fraction of singletons (i.e., users whose tweet were not mentioned or retweeted by others) is also one of the recommend classifier, indicating that most of the times the followers neglect the message. These findings are in tune with the attention seeking behavior of rumor initiators [14].

Third, the linguistic features indicated reaction of people toward rumors. Rumors were significantly less likely to contain positive affect words (i.e., 'posemo' in Table IV than non-rumors (p-$= 1.7e^{-5}$). Also, users are far more likely to mention negating words (e.g., no, not, never) in sentence and take a cognitive action (e.g., cause, know) to the rumor related content ($p = 3.2e^{-6}$). Also users are more likely to take an inferring action (i.e., tentatative) for rumors. This can be an indicator that many users try to examine credibility of rumor when they faced it.

B. Classification Results

So far we have demonstrated that periodicity of the external shock is the most important feature for distinguishing rumors by the two variable selection methods. This result supports that the PES model and its application is a good method to extract reliable features for rumor spreading. In order to test whether the selected features are effective classifiers, we adopted 15 features that were used in [3] as described in Table V. We took these features as a baseline feature set (B) for the classification task.

Table V: Features for determining credibility of information described in [3], used as baseline.

Feature definition
Fraction of tweets containing a URL
Fraction of tweets containing negative sentiment
Fraction of tweets containing positive sentiment
Fraction of tweets containing a question mark
Fraction of tweets containing a mention
Fraction of tweets containing a smiley emoticon
Fraction of tweets containing the first person pronoun
Spreader's average number of posts
Spreader's average number of friends
Spreader's average number of followers
Spreader's average number of days since registration
Average sentiment score in tweets
Number of distinct short URLs in tweets
Maximum level of the diffusion tree
Fraction of tweets by the most prolific spreader

To compare the discriminative power of the newly proposed features in this paper, we consider two different sets

Table VI: The average performance for each classification method: B (baseline with 15 features from [3]), $S1$ (proposed in this paper with 11 features), and C (combinaton of baseline and our proposed method with 27 features)

Set	Method	Accuracy	Precision	Recall	F_1
B	**Decision Tree**	0.669	0.821	0.717	0.685
	Random Forest	0.747	0.771	0.813	0.762
	SVM	**0.811**	**0.891**	**0.753**	**0.788**
$S1$	**Decision Tree**	0.791	0.854	0.772	0.780
	Random Forest	**0.900**	**0.935**	**0.892**	**0.893**
	SVM	0.875	0.934	0.838	0.859
C	**Decision Tree**	0.821	0.853	0.843	0.822
	Random Forest	**0.897**	**0.923**	**0.883**	**0.878**
	SVM	0.873	0.908	0.873	0.867

of recommended features based on random forest (which we call $S1$ and is composed of 11 significant features as listed in Table V and logistic regression (which we call $S2$, composed of 9 features). We also constructed a combined set, $C = B \cup (S1 \cup S2)$, with a total of 27 features. We compared the baseline (B) with one of our features ($S1$) and the combined set (C), by adopting standard three classification algorithms: decision tree, random forest, and SVM.Random forest gave marginally better result than logistic regression in the comparison of $S1$ and $S2$. Hence, we use $S1$ as the representative classification method of our features. For each feature set, we applied 1000 times of 2-fold cross validation and calculated the average recall, precision, and F_1 measure.

Table VI displays the results. For each feature set, we highlight the best performing classification method in bold text. For instance, SVM gave the best result for baseline and random forest gave the best result for our features. Overall, the table clearly demonstrates that our features ($S1$) retain a much higher value for every measure compared to baseline. The difference between based on the F_1 measure is 0.788 for baseline and 0.893 for our features. Furthermore, the combined set (C) does not always give improvement and sometimes even lead to overfitting, as indicated by the lower value of the F_1 measure of C (0.878) compared to that of $S1$ (0.893). These results indicate that our original features play an important role in rumor classification compared to the best known features that have been studied before.

VI. Conclusion

We studied the rumor spreading pattern on Twitter and tried to classify rumors from non-rumors. Here, three sets of features are explored: temporal, structural and linguistic. To extract temporal and structural feature sets, we addressed new time series fitting model and network structure. Combined with a set of linguistic features, our integrated feature set shows a more accurate result of identifying rumors from other type of information than baseline features.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0012988) and the ICT & Future Planning (2012R1A1A1014965).

References

[1] G. Allport and L. Postman. *The psychology of rumor*. Rinehart & Winston, 1947.

[2] P. Bordia and R. Rosnow. Rumor rest stops on the information highway: Transmission in a computer-mediated rumor chain. *Human Communication Research*, 25:163–179, 1998.

[3] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. In *WWW*, 2011.

[4] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *ICWSM*, 2010.

[5] N. DiFonzo and P. Bordia. *Rumor psychology: Social and organizational approaches*. American Psychological Association, 2007.

[6] N. DiFonzo, P. Bordia, and R. Rosnow. Reining in rumors. *Organizational Dynamics*, 23:47–62, 1994.

[7] E. Foster and R. Rosnow. *Gossip and network relationships*. D.C. Kirkpatrick, S. Duck and Foley, M. K. (eds) *Relating Difficulty: the Processes of Constructing and Managing Difficult Interaction*, 2006.

[8] Y. Matsubara, Y. Sakurai, B. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: Model and implications. In *KDD*, 2012.

[9] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer. Psychological Aspects of Natural Language Use: Our Words, Ourselves. *Annual Review of Psychology*, 54:547–577, 2003.

[10] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer. Detecting and Tracking Political Abuse in Social Media. In *ICWSM*, 2011.

[11] R. Rosnow. Inside rumor: A personal journey. *American Psychologist*, 46:484–496, 1991.

[12] T. Shibutani. *Improvised news: A sociological study of rumor*. Bobbs-Merrill, 1996.

[13] D. Solove. *The future of reputation: Gossip, rumor, and privacy on the Internet*. Yale Univ Pr, 2007.

[14] C. Sunstein. *On Rumours: How Falsehoods Spread, Why We Believe Them, What Can Be Done*. Penguin, 2011.

[15] V. Svetnik, A. Liaw, C. Tong, and T. Wang. Application of breimans random forest to modeling structure-activity relationships of pharmaceutical molecules. *Multiple Classifier Systems*, pages 334–343, 2004.