Detecting Rumor Patterns in Streaming Social Media

Shihan Wang
Department of Computational Intelligence and Systems Science
Tokyo Institute of Technology
Yokohama, Japan
ShihanW@trn.dis.titech.ac.jp

Takao Terano
Department of Computational Intelligence and Systems Science
Tokyo Institute of Technology
Yokohama, Japan
terano@dis.titech.ac.jp

***Abstract*—Rumor detection in streaming social media is a significant but challenging problem. In this paper, we present a method to identify rumor patterns in the streaming social media environment. Patterns which combine both structural and behavioral properties of rumor are firstly proposed to distinguish false rumors from valid news. A novel graph-based pattern matching algorithm is also described to detect rumor patterns from streaming social media data. Compared within twitter data of rumors and non-rumors, our selected rumor patterns contain distinct properties of rumors in short-term series.**

***Keywords*-rumor detection; social media; streaming pattern matching; socioeconomic sustainability**

I. Introduction

As Microblog platforms like Twitter and Sina Weibo rapidly grow, social media has become a popular communication tool in our daily life and attracts more and more attention. Thanks to its tremendous reachability, social media provides organizations and individuals wider opportunities of collaboration and is considered as a new driver of sustainability. Nevertheless, social media brings not only effective valid information, but vast false rumors as well. In fact, with the extremely fast and wide spread of information, online rumor causes devastating socioeconomic damage before being effectively corrected. Therefore, rumor detection in online social media is significant for the sustainable development.

Rumor is known as a piece of information or statement that cannot be verified as true or false, but quickly spreading from person to person [1]. Recently, many researchers have focused on automatically detecting rumor and determining its credibility. While they only analyze and evaluate rumor after it has been widely spread, there is still an important gap of rumor detection in the real-time streaming environment. In fact, it is essential to discover the rumor directly from online social media streams before it causes too much damage.

Here rumor detection in streaming social media is very challenging, not only because of the massive and noisy dataset but also the streaming environment. Most of the traditional methods employ classification or clustering techniques to identify rumor, which is limited in streaming scenario. Faced with these challenges, we expect to detect rumors in streaming social media using pattern matching approach. In this paper, we focus on discovering important rumor patterns and detecting them in streaming dataset.

We make two contributions in this work. First, we present a group of rumor patterns combining both structure and behavior features, which has never been done particularly for the streaming detection environment. Second, we propose a novel graph-based pattern matching algorithm, which is designed to identify patterns from real social media streams.

The rest of this paper is organized as follows. In section 2, we review the related work in rumor detection. Section 3 describes our pattern design and its theory base. Section 4 explains the pattern matching algorithm while we present the preliminary experiments and results in section 5. In section 6, we summarize this paper and our future work.

II. Related Work

While rumors have been a hot topic in the psychology field for a long time [2], computer scientists focus on automatic rumor detection of online social media only in recent years. Since the research on rumor detection in streaming environment is quite limited, in this section, we mainly review the related work on traditional offline identification. Regardless of literature focus on either Twitter or Chinese Sina Weibo data source, we group them based on their approaches: classification-based and pattern matching-based.

A. Classification-based Rumor Detection

Much previous literature has considered rumor detection as a binary classification problem. Researchers utilized supervised learning approach to automatically determine whether one trending topic that is spreading is truth or false. As identifying the credibility of information is complex, most existing approaches employ various kinds of features beyond the text of the posts only [3].

Catillo et al [4] firstly grouped and reviewed several features that are widely used in rumor detection, including content-based feature, user-based feature, behavior-based feature and propagation-based feature. Other works extended these features using own specific properties. Sun et al [5]

978-1-4799-9926-2/15/$31.00

and Yang et al [6] extracted multimedia-based and location-based features respectively to distinguish rumors in Sina Weibo from ordinary posts. Kwon et al [7] firstly examined temporal characteristics in rumor spreading.

B. Pattern Matching-based Rumor Detection

Ennals et al [8] used pattern matching techniques to highlight disputed claims from the web. Their method automatically searched lexical patterns for claims, then filtered claims by a classifier and provided a corpus of disputed claims only. On the other side, Zhao et al [9] identified trending rumors in social media based on inquiry phrases patterns. Considering content features show early in the rumor diffusion process, they presented an approach to cluster only signal pattern contained tweets and address controversial events with high rumor likelihood. While both previous works acquired rumor-related patterns, none of them contained properties beyond the post text.

Although multiple feature-oriented classification methods bring decent detection accuracy, most of these features only become available after rumor has already flourished and been transferred by many users. Therefore, it is not practical to use such approaches in a real-time situation, while rumors have already caused serious socioeconomic damage before they were detected and corrected. We expect the rumor patterns detection method using pattern matching techniques to overcome this drawback. While the previous pattern matching-based research only considered text-related features, which are not enough for the rumor detection task [4], we propose to extend social media rumor patterns from various aspects in this work.

III. Rumor Pattern Design

In order to use patterns to detect rumors from social media stream in the future, there are two important aspects need to be balanced within pattern design.

On the one hand, we expect the pattern to be as complex as possible because the combination of various features can contribute to a higher accuracy for rumor detection. On the other hand, the streaming environment restricts the complexity of patterns, as data stream has the one pass constraint, which makes it difficult to do the iterative calculation and limits the computing and storage capabilities [10].

Therefore, we not only focus on the most influential features within rumor detection task, but also consider properties that are practical in streaming process. In total, two significant properties are extracted: propagation structure and behavior of users' opinion on target posts. We will explain our detailed design and theoretical base of both properties in the following part.

A. Structural Design

In the study of [4], authors analyzed the impact of different features for information credibility. They observed that graph structure pattern of propagation is one of the most relevant to detect non-credible news. This drives us to firstly consider propagation structure in the patterns.

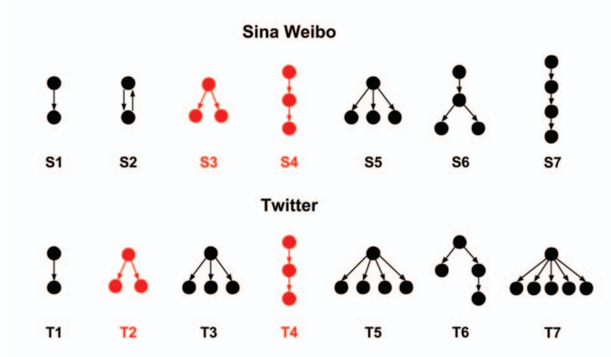

Figure 1. Frequent-ordered Nontrivial Cascades of Trending Topic Propagation in Twitter [12] & Sina Weibo [13]

On social media like Twitter, there is no important community structure [11], also overall properties of the graph are hard to measure in streaming data. So, instead of macro-level measurements, we focus on micro cascade motifs that present representative characteristic in event diffusion network. Zhou et al [12] and Fan et al [13] studied the trace of information propagation in trending topics of Microblog and obtained topological features. Figure 1 shows top seven frequent nontrivial cascade shapes from both Twitter and Sina Weibo data.

According to the figure, we find that, except for the basic shape of two nodes, T2(S3) and T4(S4) are the most important structures among the common cascades. As for one side, they are the top of the most frequent ones. For another side, all other important cascades can be decomposed into a set of them.

In real detecting situation, as social media data stream keeps coming, the propagation graph is growing from the basic structures. Based on this observation, we picked these two subgraphs as the structural features in our pattern.

B. Behavioral Design

Meanwhile, many studies retrieved behavior property of how users feel about the target post and considered it as a significant signal. Therefore, we propose to combine users' behavior feature as well.

A study about how information propagated through the Twitter network after 2010 Chile earthquake provides a promising support about user opinion analysis. They exhibited that, user attitude is one obvious difference in the propagation of tweets between rumors and valid news. In fact, more negative and doubted users tend to be involved into false rumor, while tweets exhibit an active attitude are more related to credible information [14][4]. At the same time, other research indicated the importance of question-asking behavior in social media in further analysis [9].

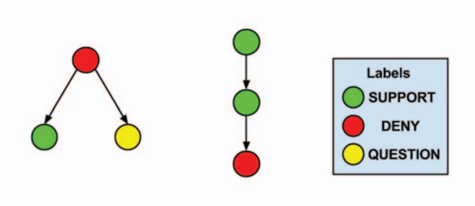

Figure 2. Two Examples of The Rumor Pattern

Overall, considering two parts of the design, our rumor pattern is the labeled graph. Two examples are shown in Figure 2. While two essential subgraphs is employed as structural base, three different labels *SUPPORT*, *DENY* and *QUESTION* are used to present user opinion.

In total, 45 possible patterns are generated. For each node in graph pattern, three possible labels are enumerated in various positions. However, because two sons in 'Star' patterns are symmetric, which means they represent the same propagation information. So, we consider patterns like $\{SUPPORT \leftarrow DENY \rightarrow QUESTION\}$ and $\{QUESTION \leftarrow DENY \rightarrow SUPPORT\}$ as the same one.

IV. Pattern Matching Algorithm

In this section, we present an algorithm that tracks matches of above graph-based rumor patterns from streaming social media data.

Overall, according to the pattern design, a labeled and directed graph is first extracted from a stream of posts, which is the original social media data. Here, each post is preprocessed by semantic analysis to address both user attitude and information propagated relationships (like retweet or mention). If this post contains propagated relationships, it is transformed into an edge. Then, this stream of edges is provided as the input of a pattern matching algorithm. The direction of an edge is defined by information spreading direction, and label attributes of nodes on this edge are defined by an opinion feature of its poster. Our algorithm processes data stream and provides a list of matched patterns and their appearing time.

We begin with introducing the indexical data structure for dynamically labeled graph pattern search, then proceed to present the detailed algorithm.

A. Relational Index Structure

We firstly introduce a data structure called *Relational Index* (R-index). R-index is responsible for storing attitude (label) information related to each node. It contains label information of the current node, as well as that of all nodes link to this one. To save the storage space, total numbers of indegree and outdegree for each kind of label are counted and collected, instead of every individual node ids. In our pattern graph, there are three kinds of label: *SUPPORT*, *DENY* and *QUESTION*. This information supports us adequate information to discover incremental patterns of each step as edges are updating in streaming. An example of our current basic structure of R-index is shown in Figure 3.

B. Graph-based pattern matching algorithm

With this R-Index structure defined, we describe the graph-based pattern matching algorithm. Here are some basic definitions we used in the algorithm.

Definition 1. Given a set of labeled nodes $N_T = \{n_1, n_2, n_3...\}$, each edge contains two nodes and time when it is shown, defined as $e =< n_{start}, n_{end}, time >$. **Edge Stream** is the continual sequence of edges, defined as $ES = \{e_1, e_2, e_3...\}$.

Definition 2. Since there are two kinds of pattern structure, **Pattern** is defined in the following two types: $p = \{'Star', n_{root}.label, n_{left}.label, n_{right}.label\}$ and $p = \{'Path', n_{root}.label, n_{up}.label, n_{down}.label\}$. A set of patterns is defined as $P_T = \{p_1, p_2, p_3...\}$. For example, two patterns in Figure 2 are defined as $\{'Star', n_{root}.label = DENY, n_{left}.label = SUPPORT, n_{right}.label = QUESTION\}$ and $\{'Path', n_{root}.label = SUPPORT, n_{up}.label = SUPPORT, n_{right}.label = DENY\}$ respectively.

Algorithm 1 matchGraphPattern(ES, P_T)

1: graph $G \leftarrow \emptyset$
2: **for** each $e =< n_{start}, n_{end}, time >\in ES$ **do**
3:   **for all** $n_i \in \{n_{start}, n_{end}\}$ **do**
4:     createNodeIfNew(n_i) in G
5:     **for all** $p_i \in P_T$ **do**
6:       **if** $n_i.label$ matches $p_i.n_{root}.label$ **then**
7:         $n_{root} \leftarrow n_i$
8:         **if** e is subgraph of p_i **then**
9:           $num \leftarrow$ getNumOfNewPattern(n_{root}, e, p_i)
10:           updateResult(p_i, num, $e.time$)
11:         **end if**
12:       **end if**
13:     **end for**
14:     updateIndex(n_i)
15:   **end for**
16: **end for**

The input to **matchGraphPattern** algorithm is an edge stream ES and a set of query patterns P_T. For every coming edge e, all of its nodes that are new for graph G are added into the graph at first (line 4). We iteratively go through every query pattern (p_i) to identify matches (line 5). Then, every node of e that shares the same label with a root node in the given pattern is selected and recorded as the root node of possible matches (line 6-7). Next, we utilize basic subgraph isomorphism to check whether this new edge is a subgraph of p_i, which is the necessary condition for further identification (line 8). As R-index maintains all previous label-related information of root node n_{root}, it is efficient to acquire the total amount of nodes that have been linked

$NodeId$	$Support\_in$	$Support\_out$	$Deny\_in$	$Deny\_out$	$Question\_in$	$Question\_out$

Figure 3. The Format of R-index Structure

to n_{root} and matches another label of p_i (line 9). After that, the algorithm provides real-time updating matches with this new edge e in the format of $< p_i, num, e.time >$ (number of new matched query patterns and timestamp) (line 10). In the end, the R-index of both nodes are updated for future calculation(line 14).

An example is given to explain the main matching procedure. Given the star pattern in Figure 2, $p = \{'Star'$, $n_{root}.label = DENY$, $n_{left}.label = SUPPORT$, $n_{right}.label = QUESTION\}$ and a new edge $e = < n_{start}, n_{end}, time >$ ($n_{start}.label = DENY, n_{end}.label = SUPPORT$), we firstly find that n_{start} is root node n_{root} and e is a subgraph of p. In the next step, we process into **getNumOfNewPattern**. As p is 'Star' type, we continue to find matches of another part in p, which is an edge with $n_{start}.label = DENY$ and $n_{end}.label = QUESTION$. Therefore, we check whether outdegree of label *QUESTION* ($Question\_{out}$) in R-index of n_{start} is zero. If not, it means we successfully discover new matched patterns of p that are contributed by this new coming edge. In this way, we capture the amount of new patterns and their discovered time ($e.time$).

V. Preliminary Experiment

In this section, we present the preliminary experiment to extract a set of rumor patterns from streaming social media data and distinguish false rumor and new events based on them.

A. Data Set

We used the dataset that was published in the work of Kwon et al [7]. It collected Twitter datasets of the trending topics, which are separated into false rumor and credible news. The validation of rumor and non-rumor label has been well annotated and evaluated by previous researchers based on both investigation websites and human participants. As the size of total 109 topics is various (from 10 to 33401 tweets), we selected 5 rumors with a larger amount of tweets, as well as 5 non-rumors that have similar size with the picked rumors. In summary, the average tweets of each topic are around 5000 and the least one has more than 2000 tweets.

B. Data Pre-process and Visualization

After ranking tweets of each topic by the timestamp, we firstly processed every group of tweets into a stream of posts, which fits the data in real-time. Then, the tweet frequency per hour of all topics is counted and collected one by one. In the previous work, they investigated tweet frequency in each day and presented bursty fluctuations over

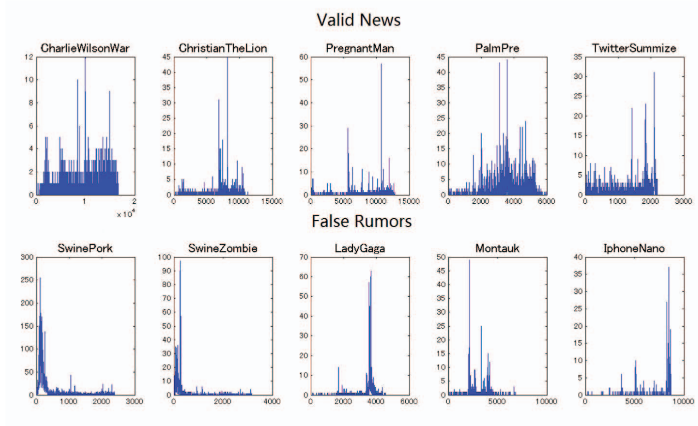

Figure 4. Tweet Frequency of Valid News and False Rumors in Short-term Series

60 days [7]. As their extracted temporal features usually last for days or even weeks, it is limited in real-time streaming detection. Therefore, we focus on hour-based frequency because such short-term temporal property can be captured even in streaming analysis. Figure 4 shows such frequency of tweets in time series for both non-rumors and rumors.

In each image of Figure 4, the x-axis represents the time where one hour is a unit, and the y-axis represents how many tweets are posted in each unit time. We observed that valid news generally shows dramatic fluctuations, while rumors usually have one sharp peak. It indicates that even in the short-term time series, rumor and valid information commonly differ from each other. Based on this difference, it is possible to identify a kind of pattern to distinguish them in streaming.

C. Tweet Semantical Analysis

Given the stream of posts, we analyzed the semantic information for each post in the next step. In this step, in order to further process data for graph-based pattern detection, we began with extract propagating relationships within tweets, then proceed to analyze user behavior feature of each tweet.

According to the official Twitter APIs[1], the *retweet*, *mention* and *reply* information is provided for each individual tweet. For example, given a tweet t_i, a set of its mentioned tweets can be acquired $T_4 = \{t_m, \|t_n, t_j, \|t_k\}$. Among them, we can identify that t_i retweets t_m and replies t_k.

[1]Source: https://dev.twitter.com/rest/public

Table I
USER OPINION MINING RESULTS OF TRENDING TOPICS

Valid News	CharlieWilsonWar	ChristianTheLion	PalmPre	PregnantMan	TwitterSummize	*Aver_Percentage* of All Users
SUPPORT Users	1093	351	863	667	862	32.74%
DENY Users	315	367	322	216	164	11.81%
QUESTION Users	176	342	303	301	183	11.14%
False Rumors	SwinePork	SwineZombie	LadyGaga	Montauk	IphoneNano	*Aver_Percentage* of All Users
SUPPORT Users	2469	433	337	177	239	13.86%
DENY Users	4309	1148	474	366	137	24.40%
QUESTION Users	3641	379	985	542	507	22.96%

Then, we captured linkages within the propagating information. In this example, the *retweet* and *reply* implies that information is transferred from t_m to t_i and from t_i to t_k respectively. For the rest of mentions, the direction of transferring is from t_i to mentioned nodes (t_n and t_j).

In our dataset, because some historical tweets have been deleted or shielded, some retweet information is missing. So, we combined the signal 'RT' of text into consideration to identify retweets. In the real streaming tweets, such information is fully provided by Twitter Streaming APIs[2].

On the other hand, we employed sentiment analysis [15] techniques to identify user opinion from tweet content. We analyzed and collected the positive (*SUPPORT*) and negative (*DENY*) attitudes through the free version of Semantria[3]. At the same time, we identified question asking tweets using simple lexical patterns based on previous research. We utilized question mark and 5W1H question words (What, Why, Who, When, Where and How) as basic patterns, but restricted 5W1H only appear at the beginning of one sentence [16]. Another pattern regular expression '*is (that| this| it) true)*' [9] is also combined to improve the precision. Besides three types of opinion, there is still a group of users who do not show any attitude. We do not consider them in our behavioral patterns. Overall, the identification results (*SUPPORT*, *DENY* and *QUESTION*) of ten topics are summarized in Table I.

Table I exhibits the total number of individual users (their posts) are identified into three attitudes. We summarized the average amount of tweets in rumor and non-rumor topics separately. Overall, one-third of total users have the positive opinion on credible information, which is three times as much as negative or questioning people. In contrast, more users tend to deny and question the non-credible rumors. This result is consistent with previous studies [14] and ready for the following process.

D. Rumor Pattern Detection

Based on information of propagating relationship and user opinion, we detected rumor patterns in streaming trending topic data using proposed pattern matching algorithm.

In the first step, we iteratively processed data stream of every topic to generate the number of matches and

[2]https://dev.twitter.com/streaming/overview
[3]https://semantria.com/

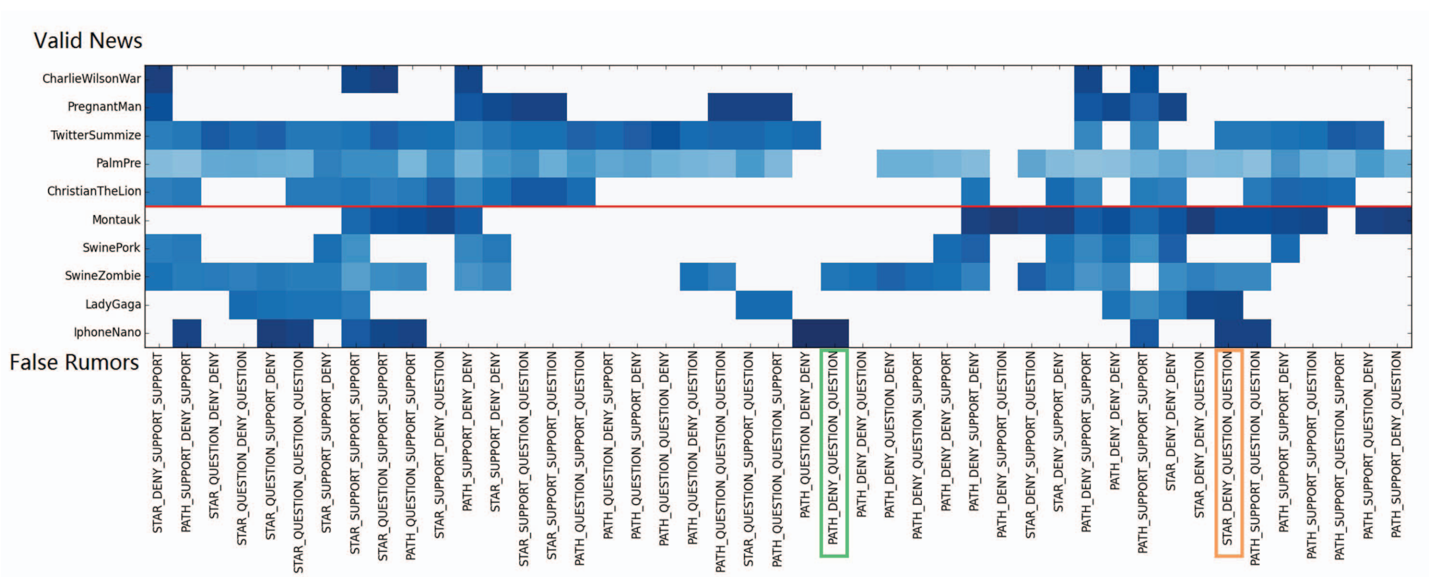

Figure 5. The Correlation Matrix Between Trending Topics and Patterns

matched time. Then, we analyzed the matched patterns from both rumors and non-rumors. In order to discover distinct patterns, especially relevant and important in rumor events, we evaluated them through term frequency-inverse document frequency (TF-IDF). We expect it to adjust patterns that appear frequently in general and distinguish rumor patterns from non-rumors.

In Figure 5, a large matrix shows the correlations between 10 topics and 45 patterns, where 5 valid news are located on the upper side and rumors are located on the lower side. In addition, the larger TF-IDF is corresponding to a darker gray in each grid. Interesting patterns can be observed from Figure 5. For example, several patterns like *'PATH:DENY_QUESTION_QUESTION'* (marked with green) only appeared in rumors. And pattern *'STAR:DENY_QUESTION_QUESTION'* (marked with orange) appears in the majority of rumors and also shows higher TF-IDF in rumors. Such phenomenon indicates that it is possible to identify patterns are either unique or more relevant in rumor events.

Next, we further evaluated the TF-IDF values of patterns and selected a set of important rumor patterns, whose average TF/IDF in rumors is over 10 times larger than that in non-rumors. In Table II, a list of selected important rumor patterns is given. Among them, the top three patterns appeared only in the false rumors, while others show closer correlations with rumors.

Using the set of selected rumor patterns, we calculated the pattern frequency in time series with the same interval one hour as previous tweet frequency analysis. The comparison images of both valid news and false rumors are exhibited in Figure 6 respectively.

Comparing the left and right part of Figure 6, we observed an obvious difference between rumors and non-rumors. In general, the temporal frequency of selected patterns matches the trend of tweets bursting in rumors very well. However, patterns do not often appear in the credible news: they do not appear in two events and do not consist with the shape

Table II
A LIST OF SELECTED IMPORTANT RUMOR PATTERNS

PATH:DENY_DENY_QUESTION
PATH:DENY_QUESTION_QUESTION
PATH:DENY_SUPPORT_QUESTION
PATH:DENY_DENY_SUPPORT
PATH:DENY_QUESTION_DENY
PATH:DENY_QUESTION_SUPPORT
PATH:DENY_SUPPORT_DENY
PATH:QUESTION_DENY_DENY
PATH:QUESTION_SUPPORT_SUPPORT
PATH:SUPPORT_DENY_QUESTION
PATH:SUPPORT_QUESTION_QUESTION
STAR:QUESTION_QUESTION_QUESTION
STAR:DENY_DENY_QUESTION
STAR:DENY_QUESTION_QUESTION
STAR:DENY_SUPPORT_DENY
STAR:DENY_SUPPORT_QUESTION

of the tweet frequency in other events. Such result indicates that the patterns acquired represent the significant properties of rumor events and are capable of distinguishing rumors from non-rumors. It provides a good potential to utilize our proposed patterns to detect rumors in streaming social media.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have described the streaming rumor detection problem by detecting rumor patterns in social media data streams. First, we extended previous work to combine properties of propagation structure and user behavior into the rumor pattern design. Second, our proposed algorithm directly explored the streaming datasets of both valid news and false rumors. We addressed a set of distinct rumor patterns that differentiate rumors from non-rumors. The short-term temporal frequency of selected patterns matched the trend of rumor-related tweets very well, which indicates a good potential to use this approach to detecting rumors in the real-time social media streams.

As for our future work, further evaluations are first

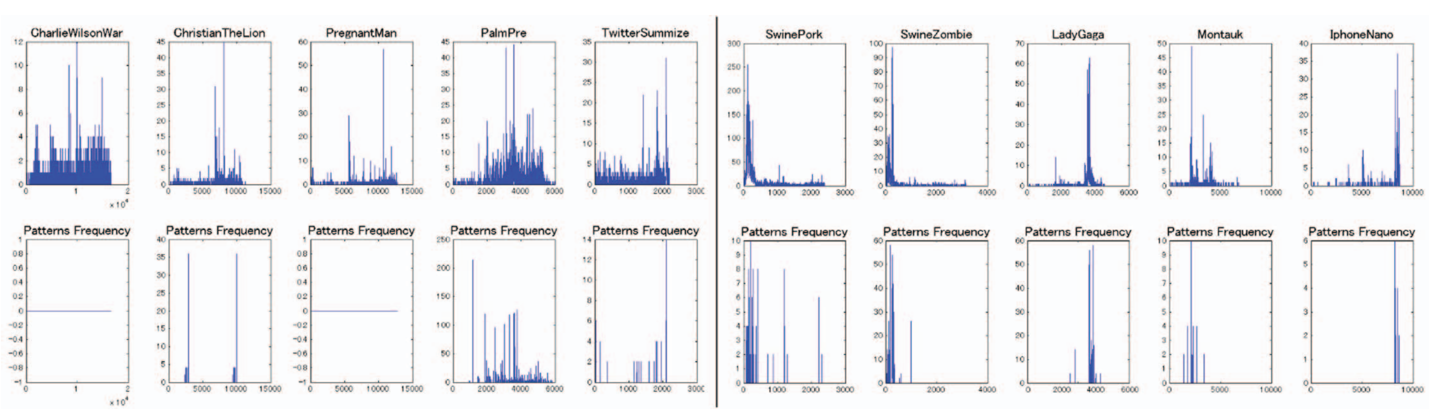

Figure 6. Frequency Comparison between Tweet and Pattern of Valid News (left) and False Rumors (right) in Short-term Series

planned to specify more correlations within the Tweets. In addition, we would like to focus on extending this rumor pattern matching approach to detect rumors in real-time social media streams. The topic-based filtering and monitoring tool will be explored and combined into our method, so that it can be evaluated in real-time streaming social media datasets.

Acknowledgment

We thank the research team from KAIST to provide the Twitter Dataset, as well as Ji Qi for supporting us with the matrix visualization tool.

References

[1] G. W. Allport and L. Postman, "The psychology of rumor." 1947.

[2] R. H. Knapp, "A psychology of rumor," *Public Opinion Quarterly*, vol. 8, no. 1, pp. 22–37, 1944.

[3] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei, "Rumor has it: Identifying misinformation in microblogs," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011, pp. 1589–1599.

[4] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011, pp. 675–684.

[5] S. Sun, H. Liu, J. He, and X. Du, "Detecting event rumors on sina weibo automatically," in *Web Technologies and Applications*. Springer, 2013, pp. 120–131.

[6] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on sina weibo," in *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*. ACM, 2012, p. 13.

[7] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent features of rumor propagation in online social media," in *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE, 2013, pp. 1103–1108.

[8] R. Ennals, D. Byler, J. M. Agosta, and B. Rosario, "What is disputed on the web?" in *Proceedings of the 4th workshop on Information credibility*. ACM, 2010, pp. 67–74.

[9] Z. Zhao, P. Resnick, and Q. Mei, "Enquiring minds: Early detection of rumors in social media from enquiry posts," in *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2015, pp. 1395–1405.

[10] J. Zhang, "A survey on streaming algorithms for massive graphs," in *Managing and Mining Graph Data*. Springer, 2010, pp. 393–420.

[11] O. Aarts, P.-P. van Maanen, T. Ouboter, and J. M. Schraagen, "Online social behavior in twitter: A literature review," in *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*. IEEE, 2012, pp. 739–746.

[12] Z. Zhou, R. Bandari, J. Kong, H. Qian, and V. Roychowdhury, "Information resonance on twitter: watching iran," in *Proceedings of the first workshop on social media analytics*. ACM, 2010, pp. 123–131.

[13] P. Fan, P. Li, Z. Jiang, W. Li, and H. Wang, "Measurement and analysis of topology and information propagation on sina-microblog," in *Intelligence and Security Informatics (ISI), 2011 IEEE International Conference on*. IEEE, 2011, pp. 396–401.

[14] M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis: Can we trust what we rt?" in *Proceedings of the first workshop on social media analytics*. ACM, 2010, pp. 71–79.

[15] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008.

[16] B. Li, X. Si, M. R. Lyu, I. King, and E. Y. Chang, "Question identification on twitter," in *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 2011, pp. 2477–2480.