Rumor: Detecting Misinformation in Twitter

Raveena Dayani	Nikita Chhabra	Taruna Kadian	Rishabh Kaushal
B. Tech. (I.T.) 4th Year	B. Tech. (I.T.) 4th Year	B. Tech. (I.T.) 4th Year	Assistant Professor
IGDTUW	IGDTUW	IGDTUW	Dept. of I.T., IGDTUW
Kashmere Gate, Delhi	Kashmere Gate, Delhi	Kashmere Gate, Delhi	Kashmere Gate, Delhi
09971815867	09654356864	08860859686	09810774114
raveenadayani@gmail.com	nikitachhabra93@gmail.com	kadiantaruna@gmail.com	rishabhkaushal@igit.ac.in

ABSTRACT

The micro-blogging social networking service, Twitter, is currently being used by around 271 million monthly active users. It is responsible for real-time propagation of information to its users. This also makes Twitter an ideal platform for dissemination of information, which many a times is not true (misinformation). This misinformation (referred as "rumor"), percolates through the online Twitter users intentionally or unintentionally. There are various issues to be addressed at different levels. At *user-level*, we need to know, 'Who is the originator of rumor?', 'How to distinguish a rumor-spreader (user with intent to spread rumors) from those who merely forward a rumor unintentionally?' At *content-level,* we need to know, 'Which tweet is talking about topics (concepts) that are potentially related to the rumor?', 'Is such a tweet really spreading rumor or refuting it?' At *network-level*, we need to know, 'What are the propagation paths traversed by rumor?', 'What is the impact (extent of damage) caused by a rumor?', 'What advisories (if any) or preventive measures can be undertaken to contain the spread of rumor?' Our work is a step in the direction of detecting the misinformation in the tweets in Twitter. Our focus is to work at *content-level* and make attempts to answer relevant questions at this level. Our detection approach is based on Rumor-Knowledge-Base (RKB) which is a repository of tweets related to different rumor topics, manually pre-detected and pre-verified, along with sentiment polarities to suggest whether this tweet spreads this rumor or refutes it. Our mechanism follows three key steps. In the *first step,* the aim is to apply text preprocessing methods to determine the topic (or combination of topics) about which the given input tweet is posted. In the *second step*, these topics or their combination are used as keywords to perform a lookup into the RKB to find whether this tweet could potentially be related to any of the rumor topics. If a match is found, then this tweet's sentiment polarity is determined by comparing it with similar tweets already in RKB. In the *last step,* a tweet which is found similar is added to the RKB for future reference. In this way, we aim to cover all kinds of tweets i.e., tweets not related to rumor, tweets in favor or against the rumor. To address a key limitation of above approach, which is manual maintenance of RKB, we propose to extract tweets from the Twitter user accounts of popular news agencies for ascertaining the trustworthiness of tweet contents and automatically maintaining the RKB. This is an ongoing work and we hope to contribute in the field of rumor detection in Twitter space.

Keywords

Rumor Detection, Rumor Analysis, News Websites Scraping, POS Tagging, Rumor Knowledge Base, Rumor Propagation.

1. MOTIVATION

Twitter makes an ideal environment for the dissemination of misinformation or deliberately false information with the huge difficulty of analyzing the content of the compressed 140-character tweet message. To find out if a topic is trending on Twitter and is claimed to be rumored we look through the websites *emergent.*info, *topsy.com* and t*rends24.in*. All these websites give timeline-based information about the topic which is allegedly controversial to be a rumor. Most often rumors subside on their own because they originate from either an unreliable source or an unauthentic (recently logged in) user who sign up to OSNs just to spread incredible information. At the same time, people feel the need to follow every tweet by their ideal person. Thus, if genuine users tweet their opinion about a rumor, then it spreads very quickly from one follower to another.

Though rumor classification is closely related to opinion mining and sentiment analysis, it presents a different class of problem because we are concerned not just with the opinion of the person posting a tweet, but with whether the statements they post appear controversial. Therefore, rumors can be classified in several types based on the intention of the tweet content about the rumor, viz., deaths of celebrity, chain mails, presidential rumors (or other highbrow people), falseness about social networking websites and mobile applications, etc.

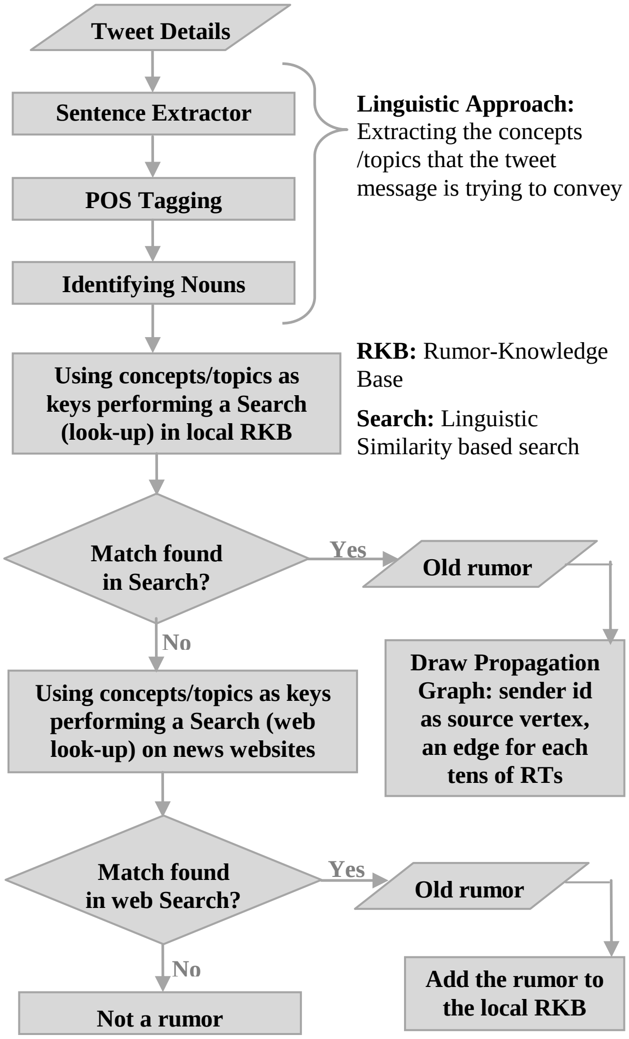
As a background study, we went through some research papers on: (1) sentiment analysis on microblogs to get a hold of linguistic methodology to help with our work, and (2) subjectivity detection to understand the meaning (intention) of natural language of the tweet message.

We reviewed the work on Sentiment Analysis and summarized the work that has been done till now starting from the time when research works in the field of sentiment analysis received global acknowledgement. We deduced all possible *sub-areas* on which research has been conducted in the past. We also concluded that every researcher has performed opinion mining considering a specific *application* and faced a lot of *challenges*, e.g., evaluating sarcasm.

The work on rumor detection helped us identify misinformation based on two methods. First is to classify tweets using a feature-based approach. At the *user-level*, the features are mostly quantitative. Rumor spreaders may be newly registered [4]. Other such quantitative features include RT count, Favorites Count, source of origination of the rumor (single/few or many people supporting the claim), geographical location where the tweet got posted, link to support credibility [1], hashtags and emoticons. The other method is to classify tweets using linguistic approach. At the *content-level* [3], we use NLP to present tweet with 2 patterns: lexical and POS. The labeling concept we use for categorization of tweets in: affirms, denies, questions, and unrelated, is proven to be quite effective in previous papers [2, 3]. Reading through the comments of the tweet [4] can be helpful for collecting enough evidence to confirm about the rumored topic. At *network-level,* we learnt about the pattern of propagation of a rumor graphically and also about how to identify structural and linguistic differences between rumor and non-rumor [5].

2. METHODOLOGY

We propose a general framework, which given a tweet predicts (1) whether it is related to a 'claimed' rumor, and if so (2) whether the user believes it or not, by comparing the tweets message with a news article which is scraped from a reliable news website. We categorized two methods for the purpose of detecting rumored data before it starts to diffuse to a large community. The feature-based approach returns quantitative results which are not based on the language or tone of tweet message. The linguistic approach, which we have decided to use in our work, returns qualitative results after processing the language and the content of the message of the tweet and categorizing the words in positive and negative models based on whether the word or gestures (hashtags, emoticons expression, and URL content) are denying the rumor or endorsing it.



Tweet Details
Sentence Extractor
POS Tagging
Identifying Nouns
Linguistic Approach: Extracting the concepts /topics that the tweet message is trying to convey
Using concepts/topics as keys performing a Search (look-up) in local RKB
RKB: Rumor-Knowledge Base
Search: Linguistic Similarity based search
Match found in Search?
Yes
Old rumor
No
Draw Propagation Graph: sender id as source vertex, an edge for each tens of RTs
Using concepts/topics as keys performing a Search (web look-up) on news websites
Match found in web Search?
Yes
Old rumor
No
Add the rumor to the local RKB
Not a rumor

Figure 1.Proposed algorithm to find whether tweet is a rumor or not

We group all the tweets belonging to a single topic as an input and collectively measure the score of sentiment (e.g., anger, sad) and linguistic (e.g., negate) categories. We build a model for three-way classification task of classifying sentiment into positive, negative and neutral. Our detection approach is based on RKB which is a repository of tweets related to different rumor topics.

We have decided not to clean the data during preprocessing because (1) punctuation marks tell if the rumor is questioned or if the user is surprised to read such a rumor (2) capitalization is necessary as it shows the impact of the rumor on the user. (3) We do not want all the stopwords to be removed because words like "can", "how", etc. are important to analyze the opinion of the user.

In Fig 1, Tweet Details is the input which contains all the attributes related to a tweet that can be retrieved using Twitter (search and streaming) API. The attributes are the tweet message, id, timestamp for when the tweet got posted, registration details of sender, location of the tweet, favorite count, RT count, source of status, etc.

The first step is to use the linguistic approach to apply text preprocessing methods to determine concepts/topics that the tweet message tries to convey. Thus, three tasks: (1) Sentence Extractor (2) Part-of-speech tagging (3) Nouns (named entity recognition) Identification. In the second step, using concepts/topics as keys we perform Linguistic-Based Search by looking-up in the Rumor-Knowledge-Base (RKB) to find out if the incoming tweet could potentially be related to any rumor topic and then finding this tweet's sentiment polarity by comparing it with similar tweets which are already present in RKB. In the last step, for an old rumor, draw propagation graph with sender tweetid as the source vertex and an edge for each tens of retweets. If no match is found, we propose to use concepts/topics as keys performing search in popular news agencies by scraping the websites for ascertaining the trustworthiness of tweet contents. If a match is found in the web search, label the rumor as *new* and add the rumor to the local RKB for future reference. Otherwise, label the tweet as a *non*-rumor.

3. RESULT

This is an on-going work as part of our B. Tech final year project. In accordance with the proposed approach we are almost complete with the linguistic approach of extracting the concepts/topics that the message of the tweet is trying to convey and moving towards the step of maintaining a database (RKB) in which the search can be performed.

4. IMPLICATION

The proposed approach presented in the poster on complete implementation will give us a rumor detection system which could be used by the microblogging network system as well as individuals to look out for falsified information. This algorithm will be implemented as a web application. The main purpose of doing rumor detection is to secure people from receiving wrongful information over the OSNs.

5. REFERENCES

[1] Castillo, C., Mendoza , M., and Poblete, P. 2011. Information Credibility on Twitter. *In Proceedings of the 20th international conference on World wide web (2011)*, pages 675–684, 2011.

[2] Mendoza, M., Poblete, B, and Castillo, C. 2010.Twitter Under Crisis: Can we Trust what we RT? In *1st Workshop on Social Media Analytics (SOMA '10)*, 25 July 2010

[3] Oazvinian,VB. Rosengren,E., and Radev, R. 2011. Rumor has it: Identifying Misinformation in Microblogs. In *Proceedings of the 2011 Conference on Empirical Methods in NLP*

[4] Kwon, S., Cha, K., Jung, W.C., and Wang, Y. 2013.Aspects of Rumor Spreading on a Microblog Network. *SocInfo, volume 8238 of Lecture Notes in Computer Science, page 299-308.Springer (2013)*

[5] Kwon, S., Cha, K., Jung, W.C., and Wang, Y. 2013. Prominent Features of Rumor Propagation in Online Social Media. *Data Mining (ICDM), 2013 IEEE 13th International Conference Dec. 2013*