Automatic Rumor Spread Prediction on Twitter

Jung Kyu (Rhys) Lim (jklim@umd.edu)
Sanchari Chowdhuri (sanchari@umd.edu)
Sohan Shah (sosha@terpmail.umd.edu)

Contents

1. INTRODUCTION

Through our research project, we aim to build a model with which we can predict whether tweets are rumors or not, and indicate the accuracy of these predictions. Information available on social networking websites like Twitter and Facebook has a global outreach, underlying the crucial role played by social media in modern day communication. With the increasing number of active users on social media, the need for accurate and unbiased information is vital. Rumors are easily spread through such websites, impacting people and organizations across the world. While social media is great platform to broadcast news, rumors create a void in the availability of credible information. We attempt to fill this void by developing a system that can catch such rumors, and also validate other cases that are genuine.

1.1. APPROACH

Leveraging Twitter APIs, we are extracting all tweets that are related to 60 different rumor cases, and tweets related to 60 cases that are not rumors (120 cases in total). These rumor and non-rumor cases are handpicked. With a total of 120 cases, we would have over a thousand tweets collected by scraping data off of twitter. Following are few of the parameters which we are considering namely:

- Number of retweets
- Number of favorites
- Number of comments
- Hashtags and URLs used in the tweets and comments
- Person/group/bot that tweeted the information
- Network pattern

- Linguistic pattern
- Temporal frequency

With a combination of these parameters, machine learning and natural language processing; we want to design our predictive system.

2. LITERATURE REVIEW

Rumor propagation research from different disciplines has identified features of rumor spreading. First, social psychologists found the short attention time span that rumormongers have, because rumor can flourish only during the time frame where the information void exists (Shibutani, 1966). Such temporal feature was confirmed on rumor and non-rumor cases on social media (Kwon, Cha, Jung, Chen, & Wang, 2013). Indeed, rumor cases tend to have multiple and periodic spikes, while non-rumors cases typically have a single prominent spike. Kwon and Cha (2014) modeled temporal pattern with external shock and exclusivity.

Second, rumor spreading demonstrated structural properties: Gossip spread more widely in sparse network structure, and denser network is less vulnerable than sparse network (Foster & Rosnow, 2006). Such structural property was identified on social network (Kwon et al, 2013). Kwon and his colleagues (2013) found that rumor cases involved larger fraction of singletons than non-rumor cases. Furthermore, Vosoughi (2015) detailed it into fraction of low-to-high diffusion, fraction of nodes in largest connected component (LCC), average depth to breadth ratio, ratio of new users, ratio of original tweets, fraction of tweets containing outside links, and the fraction of isolated nodes.

Third, rumor spreads showed the linguistic feature. Rumors are expected to be dominated by certain sentiments such as anxiety, uncertainty, credulity, and outcome-relevant involvement (Rosnow, 1991). Such linguistic characteristic was identified on social network: While some found that the certain sentiments were related to rumor (Kwon et al, 2013), some further specified the linguistic characteristics, including ratio of tweet containing negations, average formality and sophistication of the tweets, ratio of tweets containing opinion and insight, and ratio of inferring and tentative tweets (Vosoughi, 2015), as well as the number of exclamation and question marks and URLs, external sentiment classification MetaMind and Part-of-Speech (POS) tagger (Zeng, Starbird, & Spiro, 2016).

Fourth, social media showed users' profile and their online activities, which helps identifying rumors. Vosoughi (2015) detailed the user identity into: user identity, controversy, originality, credibility, influence, role, and engagement, and found user profile is available identifying rumor spreads, later further confirmed in real time (Liu, Nourbakhsh, Li, Fang, and Shah, 2015).

Lastly, beyond identifying the rumor that has ended, research attempted to identify rumor in real time. Kwon, Cha, and Jung (2017) found that user and linguistic features act as a good indicator during initial propagation phase, while structural and temporal features were not available at the initial phase. Liu and colleagues identified source credibility, source identity, source diversity, source location and witness, message belief, and event propagation (Liu, Nourbakhsh, Li, Fang, and Shah, 2015).

3. RESEARCH QUESTIONS

Guided from the previous literature, we raise the following research questions:

Research Question 1: Predict whether a given tweet is a rumor or not.

Building a classifier which can correctly predict if a given tweet is a rumor or not based on the text of the tweet and other user related parameters like number of followers of the users, number of friends, and number of likes for the tweet.

Research Question 2: Predict user demographic information of those who are spreading rumors

Building a classifier which can correctly predict the gender of users from parameters like text,user followers ,user friends.

Research Question 3: Predict the topic/industry of rumors.

The rumor dataset will contain tweets related to various topics .Building a classifier which can correctly predict the topic of rumor tweet based on text of tweet.

4. DATA SET COLLECTION

Tweets collected are related to two types of events - rumors and non-rumors. We ascertained a given event as a rumor or non-rumor based on our previous knowledge regarding the situation. We have made an assumption that until a certain piece of information is not authenticated officially; it would be treated as a rumor. To collect tweets we used Twitter REST APIs. We have developed code in python using **Tweepy,** an open-sourced (library) hosted on GitHub, which enables **Python** to communicate with Twitter platform and use its API to collect tweets with desired keywords, language and geolocation, among many other parameters.

In order to create our dataset we collected tweets related to certain rumor events. Some of the rumors cases which are actively trending on twitter are "***tweets about iPhone 7 explosions***" and "***News of Trump's allegiance with Russia and his real estate transactions there."*** For non-rumors we considered cases like "***NASA's discovery of seven new Exoplanets", "announcements of wrong best pictures at Oscars***" and "***ISRO's recent 104 satellite launches".*** Using tweepy we collected tweets about these incidents and stored in separate csv files.

Our code contains a for loop with an API call which iterates 2000 times to retrieve tweets specific to the keywords and hashtags we are interested in. Each iteration would give tweets related to the event starting with the most recent tweet and dating back as long as 7 days. The API response contains the actual text in the tweet along with other attributes of interest, such as tweet ID, date, number of retweets, favorite count, user ID, user followers count, and geolocation, to name a few. The entire API response is stored in a new pandas dataframe, which we then save as a csv file.

4.1. DATA CLEANING

For each keyword that is used in the query, the API responses are stored in one csv file. At times, we need tweets that contain two or more keywords (or hashtags), requiring csv files to be merged, with the resulting file containing only the relevant tweets. When more than one csv file encapsulates an event, we merged the csv files based on the 'tweet ID'. Merging was done using a combination of IF and VLOOKUP functions in excel.

As a result of the merging process, each rumor and non-rumor event contains only 1 source csv file, which we then intend to import into R and perform data analysis.

5. CONTRIBUTION OF TEAM MEMBERS

Task	Jung Kyu Rhys LIM	Sanchari Chowdhuri	Sohan Shah
Brainstorming and Research Questions Formulation	✓	✓	✓
Data Collection		✓	✓
Data Cleaning			✓
Report Writing	✓	✓	✓
Literature Review	✓		
Website Updates		✓	
Team Review of all tasks	✓	✓	✓
Project Presentation	✓	✓	✓

6. CONCLUSION & FUTURE SCOPE:

Milestone 1 was mainly focused on raising formulating research questions which we hope to address and also identifying key parameters and their correlation with our research questions. We were focused on developing a code to create our dataset for rumor and non-rumor tweets collection. We hope to replicate this with other rumor, non-rumor tweet collections which will be collected day to day basis.

Reference

- Foster, E. K., & Rosnow, R. L. (2006). Gossip and network relationships. *Relating difficulty:Theprocess of constructing and managing difficult interaction*, 161-180.
- Kwon, S., & Cha, M. (2014, May). Modeling Bursty Temporal Pattern of Rumors. In *ICWSM*.
- Kwon, S., Cha, M., & Jung, K. (2017). Rumor Detection over Varying Time Windows. *PLOS ONE, 12*(1), e0168344.
- Kwon, S., Cha, M., Jung, K., Chen, W., & Wang, Y. (2013, December). Prominent features of rumor propagation in online social media. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on* (pp. 1103-1108). IEEE.
- Liu, X., Nourbakhsh, A., Li, Q., Fang, R., & Shah, S. (2015, October). Real-time rumor debunking on twitter. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (pp. 1867-1870). ACM.
- Rosnow, R. L. (1991). Inside rumor: A personal journey. *American Psychologist*, *46*(5), 484.
- Shibutani, T. (1966). *Improvised news: A sociological study of rumor*. Ardent Media.
- Vosoughi, S. (2015). *Automatic detection and verification of rumors on Twitter* (Doctoral dissertation, Massachusetts Institute of Technology).

- Zeng, L., Starbird, K., & Spiro, E. S. (2016, March). # Unconfirmed: Classifying Rumor Stance in Crisis-Related Social Media Messages. In *Tenth International AAAI Conference on Web and Social Media*.

****