Rumor Detection on Twitter Using a Supervised Machine Learning Framework

Hardeo Kumar Thakur, Netaji Subhas Institute of Technology, New Delhi, India
Anand Gupta, Netaji Subhas Institute of Technology, New Delhi, India
Ayushi Bhardwaj, Netaji Subhas Institute of Technology, New Delhi, India
Devanshi Verma, Netaji Subhas Institute of Technology, New Delhi, India

ABSTRACT

This article describes how a rumor can be defined as a circulating unverified story or a doubtful truth. Rumor initiators seek social networks vulnerable to illimitable spread, therefore, online social media becomes their stage. Hence, this misinformation imposes colossal damage to individuals, organizations, and the government, etc. Existing work, analyzing temporal and linguistic characteristics of rumors seems to give ample time for rumor propagation. Meanwhile, with the huge outburst of data on social media, studying these characteristics for each tweet becomes spatially complex. Therefore, in this article, a two-fold supervised machine-learning framework is proposed that detects rumors by filtering and then analyzing their linguistic properties. This method attempts to automate filtering by training multiple classification algorithms with accuracy higher than 81.079%. Finally, using textual characteristics on the filtered data, rumors are detected. The effectiveness of the proposed framework is shown through extensive experiments on over 10,000 tweets.

KEYWORDS
Keyword Extraction, Naive Bayes, Rumor, SVM, Text Classification, Twitter

INTRODUCTION

When it comes to people and connectivity online social networking sites like Twitter seem to have set their web wider than any other. Studies reveal that the swift information propagation potential and extensive reach to the masses has made twitter the top source of breaking news in most parts of the world (Rosenstiel, Sonderman, Loker, Ivancin & Kjarval, 2015). Amidst the news statements, misinformation or rumor also finds its route into the society and its widespread is one of the biggest challenges facing social media (The guardian, 2011). Rumors are ubiquitous and with vast public involvement, they have the capability to impose real damage to individuals, organizations, and the government. Viral rumors about individuals that condemn them for their actions may lead to hate campaigns and eventually harm their reputation. This may affect individual’s self-esteem and confidence level. Rumors accelerate the dynamic nature of share markets and consequently elaborate their effect on organizations (Time,2013). Sometimes misinformation about the outburst of a disease (Time,2014) might affect the tourism of a country and likewise other government sectors. Analysis of rumors led to its aspect of public participation through various perspectives, for example, political belief (Shin, Jian, Driscoll & Bar,2016), influence on markets (Cruz & Gomes, 2013; Yiwen,
Guizhong & Zongping, 2000) and crisis management (Onook, Manish & Raghav, 2013). For analysis of misinformation, its detection is the preliminary step.

The micro-blogging service averaged at 310 million users in the first quarter of 2016 (Statista - The statistics portal) and around 6000 tweets per second are captured on an average worldwide (Internet live stats). The task of detecting a rumor in such an enormous database is challenging. Researchers have been dealing with this problem by looking into various properties of rumors on social media. One of the methods involves a baseline method being framed by seeking related information in other major media to confirm the topic reliability (Hashimoto, Kuboyama & Shirota, 2011). Another method involves employing temporal characteristics wherein a periodic time series model is proposed that considers daily and external shock cycles, where the model demonstrates that rumor likely has fluctuations over time (Sejeong, Meeyoung, Kyomin, Wei & Yajun, 2013). However, waiting for other major media for confirmation or observation of temporal shock cycles might give ample time for the rumor to spread. Another study is on user relationships networks in rumor and non-rumor cases showing that rumor networks are dense i.e. users are closely linked whereas non-rumor networks are scattered (Ling-han, Jia-yin & Chao-yang, 2013) but again a distinguishable network formation buys time for rumor propagation. In another research, Aspects of Rumor Spreading (Sejeong, Meeyoung, Kyomin, Wei, & Yajun, 2013), diffusion structure and linguistic properties of rumors are studied that are a promising aspect of rumor spreading, but this paper has given no framework for detection of misinformation. Another approach involves time series modeling technique to incorporate various social context information about rumors (Ma, Gao, Wei, Lu & Wong, 2015). Here, a method is proposed for detection of the decentralized phenomenon that rules out the major problems as explained in the following example. Humanity tends to be more susceptible when it comes to sensitive issues underlying a rumor and get patently swayed, therein upholding integrity becomes a matter of paramount and urgent importance, for example, Kolkata riots (Scroll.in, 2015) or rumors about Hindu exodus from a village. Such rumors demand a quick detection algorithm which the earlier mentioned methods lack. Clearly, there are two snags in the way, one being huge data of social sites for which this approach incorporates a filter mechanism and another being the identification of a feature that does not buy time to reveal, which is inspected to be queried reactions of viewers.

The two-fold approach focuses on a supervised machine learning model to detect rumors. Since experts instead of measuring falsity, elucidate rumors to be statements presenting facts that lack substantiation. Therefore, this becomes the base of the first classification model to filter factual or informative tweets. The bags of news-like tweets, being a potential rumor, are further studied. Moreover, research across sociology, psychology, and communication studies have widely varying definitions of rumor (Pendleton, 1998). Hayakawa (2002) has defined rumor as a kind of social phenomenon that a similar remark spreads on a large scale in a short time through chains of communication. Likewise, rumors, on social networks, in their unverified stage produce a distinctive burst in the number of retweets within the first few minutes (Zubiaga et al., 2016). Therefore, for any information on the social media, the first few judgments are passed as the reaction by its audience, and these reaction tweets help recognize rumor on early stages. Lastly considering the linguistic trait of rumor reaction tweets, queried threads are detected as rumor since it highlights the information as insufficient or doubt rising, while rest are put as non-rumor information.

**RUMOR DETECTION MODELS**

This section elaborates the two models and the framework that has been used to solve the problem of rumor detection as explained previously.

**Model 1: Personal and Non-Personal Classification**

The Twitter Landscape Report (Brandwatch, 2013), shows that a large number of tweets can be categorized as personal. Therefore, any tweet from the twitter database is broadly classified as
Dervin’s Sense-Making Theory
[www.igi-global.com/chapter/dervins-sense-making-theory/127122?camid=4v1a](www.igi-global.com/chapter/dervins-sense-making-theory/127122?camid=4v1a)

Efficiently Producing the K Nearest Neighbors in the Skyline on Vertically Partitioned Tables
[www.igi-global.com/article/efficiently-producing-the-k-nearest-neighbors-in-the-skyline-on-vertically-partitioned-tables/100041?camid=4v1a](www.igi-global.com/article/efficiently-producing-the-k-nearest-neighbors-in-the-skyline-on-vertically-partitioned-tables/100041?camid=4v1a)