

# Rumor Detection from Social Media

## Problem Statement

The objective is to detect and ascertain the rumor on social media. The extensive spread of rumor has the potential for extremely negative impacts on individuals and society. Social psychology literature defines a rumor as a story or a statement whose truth value is unverified or deliberately false. False rumors are damaging as they cause public panic and social unrest thereby creating a chaos in the city. Automatically predicting the veracity of information on social media is of high practical value. Debunking rumors at an early stage of diffusion is particularly crucial to minimizing their harmful effects. To distinguish rumors from factual events, individuals and organizations often have relied on common sense and investigative journalism. Rumor reporting websites like snopes.com and factcheck.org are such collaborative efforts. However, because manual verification steps are involved in such efforts, these websites are not comprehensive in their topical coverage and also can have long debunking delay.

Individual microblog posts are short in nature, containing very limited context. A claim is generally associated with a number of posts that are relevant to the claim. In this research rumor ascertaining will be done on aggregate level instead of individual level. Therefore, predicting the veracity of each post is not a complete solution that's why here concentrating on detecting rumors at the event-level, comprised of a set of relevant posts.

## Background

The rumor detection on social media has recently become an emerging research that is attracting tremendous attention. Social media for news consumption is a double-edged sword. Automatic rumor detection from social media is based on traditional classifiers that detect misinformation stemming from the pioneering study of information credibility on Twitter. Existing rumor detection models use learning algorithms that incorporate a wide variety of features manually crafted from the content, user characteristics, and diffusion patterns of the posts. Most of these prior works attempted to classify the veracity of spreading memes using information other than the text content, for instance, the popularity of a post (e.g., the number of retweets or replies of the post), the features relevant to determine a user's credibility, etc. However, feature engineering is painstakingly labor intensive. The RNN-based method disregards this completely yet can achieve better performance due to the effective representation learning capacity of deep neural models.

Deep neural networks have demonstrated clear advantages for many machine learning problems. The opportunities can be explored to automatically discover and exploit deep data representations for efficient rumor detection. The sequential nature of text streams in social media, recurrent neural networks (RNN) are suitable for rumor detection. This is because the connections between units in an RNN form a direct cycle and create an internal state of the network that might allow it to capture the dynamic temporal signals characteristic of rumor diffusion. The gradients of RNNs are computed via backpropagation through time. Because of vanishing or exploding gradients, the basic RNN cannot learn long-distance temporal

dependencies with gradient-based optimization. LSTM (Long Short-Term Memory) is used to deal with this while making an extension that includes “memory” units to store information over long time periods.

<https://www.ijcai.org/Proceedings/16/Papers/537.pdf>

## Methodology

Step1: Collection of data from post datasets using Twitter (www. twitter.com)

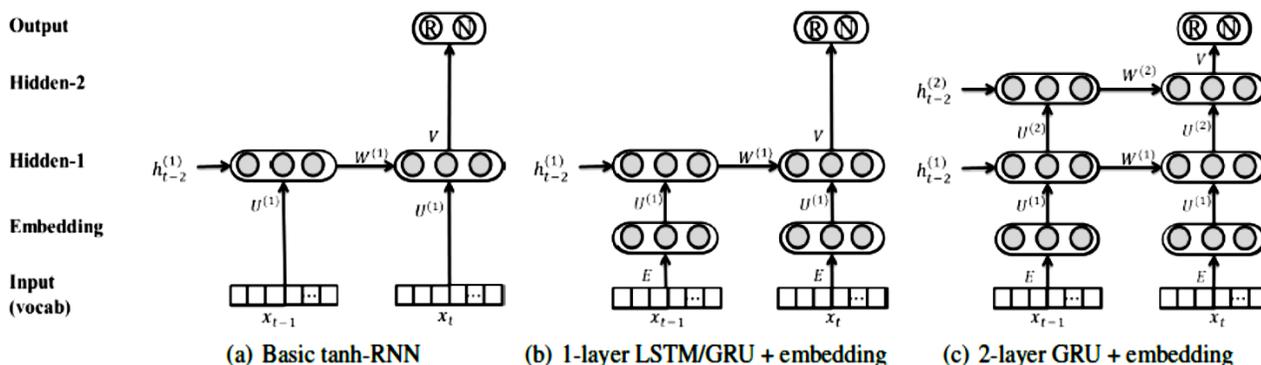
For the Twitter data, we confirmed rumors and non-rumors from www. snopes.com, an online rumor debunking service.

Step 2: For each event, we extract the keywords from the last part of the Snopes URL, e.g., <http://www.snopes.com/pentagon-spends-powerballtickets>. We refine the keywords by adding, deleting or replacing words manually, and iteratively until the composed queries can have reasonably precise Twitter search results.

Step3:

Apply a type of feed-forward neural network RNN that can be used to model variable-length sequential information such as sentences or time series. The RNN-based model will classify posts or microblog events into rumors and non-rumors. As follows

- Firstly converts the incoming streams of posts as continuous variable-length time series.
- Then describe RNNs with different kinds of hidden units and layers for classification.
- This method learns RNN models by utilizing the variation of aggregated information across different time intervals related to each event.
- Empirically evaluate the RNN-based method with three widely used recurrent units, tanh, LSTM and GRU, which perform significantly better.



<https://www.ijcai.org/Proceedings/16/Papers/537.pdf>

## **Experimental Design**

### ***Dataset:***

Post/Blog datasets using Twitter ([www.twitter.com](http://www.twitter.com))

### ***Evaluation Measures:***

This method learns RNN models by utilizing the variation of aggregated information across different time intervals related to each event. RNN-based method can be evaluated with three widely used recurrent units, tanh, LSTM and GRU, which perform significantly better.

### ***Software & Hardware Requirements:***

Python based Computer Vision and Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Navigator, Python, and libraries such as Tensorflow, and Keras will be utilized for this process.

## **References**

- [1][Allport and Postman, 1965] G.W. Allport and L.J. Postman. The psychology of rumor. Russell & Russell, 1965.
- [2] [Bengio et al., 1994] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, 1994.
- [3] [Castillo et al., 2011] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In *Proceedings of WWW*, 2011.
- [4] [Cho et al., 2014] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.
- [5] [Collobert et al., 2011] Ronan Collobert, Jason Weston, Leon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12:2493–2537, 2011.
- [4] [Devlin et al., 2014] Jacob Devlin, Rob Fergus, Oriol Vinyals, and Jeff Dean. Fast and robust neural network joint models for statistical

machine translation. In Proceedings of ACL, 2014.

[5] [Duchi et al., 2011] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods

for online learning and stochastic optimization. Journal of Machine Learning Research, 12:2121–2159, 2011.

[6] [Friggeri et al., 2014] Adrien Friggeri, Lada A Adamic, Dean Eckles, and Justin Cheng. Rumor

cascades. In Proceedings of ICWSM, 2014.

[7][Graves, 2013] Alex Graves. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850, 2013.

[8][Gupta et al., 2013] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In Proceedings of WWW companion, 2013.

[9] [Kombrink et al., 2011] Stefan Kombrink, Tomas Mikolov, Martin Karafiat, and Luk ´ as Burget. Recurrent neural net- ´ work based language modeling in meeting recognition. In Proceedings of INTERSPEECH, 2011.

[10] [Kwon et al., 2013] Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. Prominent features of rumor propagation in online social media. In Proceedings of ICDM, 2013.