A Survey on Fake News Detection Methods

Rashmi Rane[1] and Dr. R. Subhashini[2]

[1]Research Scholar, School of Computing, Sathyabama Institute of Science and Technology, Chennai, India.

rashmirane09@gmail.com

[2]Professor, School of Computing, Sathyabama Institute of Science and Technology, Chennai, India.

subhashini.it@sathyabama.ac.in

Abstract: Nowadays with the extensive use of Social media, people are sharing contents in an unprecedented scale. As there are two sides of every coin, people are using Social Media to consume news as its very cost effective fast method for getting the latest news on the tips of their fingers. But it also propagates fake news, i.e. low quality news with intentionally false information. In this paper we will see the different methods which are being used to detect the fake news also known as rumors.

Keywords: Rumor, Fake News Detection, Online Social Networks, Rumor Classification.

Introduction

In the last decade, most of the people rely on the Internet as a source of news. A study by the Pew Research Center [6] states that today around seven-in-ten Americans use social media to connect with one another, engage with news content, share information and entertain themselves. With the development of social network, the amount of information has been growing exponentially. However, the eminence of information does not change. Social network is full of all kinds of incorrect information, especially rumor information. Therefore, automatic valuation of information trustworthiness has received considerable attention in recent years.

Rumor definition

According to Merriam-Webster [1], a rumor is defined as “a statement or report current without known authority for its truth,” which may refer to a “talk or opinion widely disseminated with no discernible source” or “information or a story that is passed from person to person but has not been proven to be.” In this paper, we make use of the definition for rumor from [2], i.e., “a statement or report current without known authority for its truth a rumor is defined as a statement whose truth value is unverifiable or deliberately false,” because most of the researchers in the computational rumor detection area agree this definition as well as its practicality for guiding the design of computational algorithms for automatic rumor detection. Rumor is defined in the dictionary as a currently circulating story or report of uncertain or doubtful truth.

Rumor classification

As per the architecture defined in [4], the entire rumor classification process can be divided into four components.

(1) Rumor Detection: The first component of rumor classification system is rumor detection where the system has to recognise whether a piece of information establishes a rumor. We can give input to this component a stream of social media posts, where a binary classifier will determine whether each post is a rumor or non-rumor.

(2) Rumor Tracking: Once a rumor is identified, this second component collects and filters posts discussing the rumor by scanning the relevant posts on social media.

(3) Stance Classification: This component determines how each post is oriented to the rumor's veracity.This is an optional component of the architecture.

(4) Veracity classification: The final component of the classification model determines the actual truth value of the rumor.The veracity classification component can get its input either from rumor tracking component or Stance classification component. The output of this component can be the predicted truth value which can be supported by relevant URLs.

Methodologies used in Rumor Classification

Many different methodologies have been used to solve the problem of Rumor classification.

Recurrent Neural Networks(RNNs) have been used widely in the work done in this field. In [5] Provenance based approach is proposed. In this approach, an important source of information used is the provenance (origin or source) of the events appearing as links in the posts on social media. The authors have proposed a provenance based approach to classify events into rumors and non-rumors. The temporal dependencies between the posts are captured using Recurrent Neural Networks. Here the textual content as well as the provenance information is combined to classify an event into rumor or non-rumor.

Most work that exists on rumor detection from social media focus on mining features or rules manually. The authors in [9] proposed a deep learning framework for rumor debunking. In their method, the RNN models learn by making use of the disparity of combined information across different time intervals related to each event. Three widely used recurrent units, tanh, LSTM and GRU, have been used to evaluate th RNN-based method, which perform significantly better than the state-of-the-arts. They have added multiple hiddenlayers and embedding layers for further enhancement.

Vosoughi et al. [6] attempted veracity classification task using three categories of features-linguistic, user oriented and temporal propagation dynamics. They also used various machine learning approaches inspired by speech recognition, such as Dynamic Time Wrapping (DTW) and Hidden Markov Models (HMMs). The results of the assessments performed on Twitter datasets gathered by the author showed that HMMs were superior to DTWs. The temporal propagation category features outperformed the linguistic and user oriented features as per the testification given by the authors.

Chen et al. [3] have treated rumor as anomalies since rumors account for a very tiny amount of all the posts shared on social media while most of the microblogs can be regarded as trustworthy posts. They have further included crowd wisdom (the collective opinion of a group rather than that of a single person or expert, in this case other users' comments on doubtful microblogs) to increase the detection performance of rumors. They have additionally considered more comment based features to describe certain differences in user behaviors when critically observing the rumor posts and genuine posts. To take into consideration of the variation of features over time at which the comments are posted, the authors employed Recurrent Neural Networks (RNN) to analyze the features. Thereafter, a combination of these time dependent features and the time independent features extracted from the original microblogs were given input into a variant Autoencoder (AE) for further anomaly detection. Experimental results showed that the combination of RNN and the variant AE based on individual users used in this model can achieve a high accuracy and F1 measure. The authors have developed an unsupervised model which does not require labeled data. This makes it particularly useful when rumor data is not so easy to obtain for training.

Liu et al. [10] introduced a new method for detecting rumors flowing freely in social media by distinguishing their proliferation patterns from those of trustworthy messages. They articulated the social media rumor recognition task as a microblog classification problem by presenting an information propagation model built upon a diverse user representation. The derived microblog classifier followed a hypothesis that rumors and credible messages tend to propagate in a social media environment following measurably differentiable patterns among a disparate population of end users. By extracting user status-sensitive features from individual users' profiles, their method estimated the retweeting probabilities of a message when it is a rumor versus a credible message respectively. The proposed information propagation model generates a classification message whether the message is a rumor or a truthful message based on the estimate given above.

G. Tong et al. [11] have studied the rumor blocking problem for online social networks since it is necessary to block the rumor as soon as it has been detected as a rumor to curb the further damage it can do. They have designed the R-tuple based sampling method and then presented a randomized rumor blocking algorithm. The proposed RBR algorithm conceptually dominates the existing rumor blocking algorithms, and as shown in the experiments it is very efficient without sacrificing the blocking effect.

B. Wang et al. [12] have proposed a rumor propagation model taking into account the following three elements: First, the global popularity of the rumor over the entire social network. Second, the attraction dynamics of the rumor to a potential spreader, i.e., the individual tendency to forward the rumor to its neighbors. Third, the acceptance probability of the rumor recipients. In their model which is inspired by the Ising model, they have combined all three factors together to propose a cooperative rumor propagation probability. They have considered the influence of blocking time to user experience in real world social networks in their rumor blocking strategies. They have proposed a blocking time constraint into the traditional rumor influence minimization objective function. Then they have proposed both greedy and dynamic blocking algorithms using the maximum likelihood principle.

Z. Tan et al. [13] proposed a novel rumor-propagation model, inspired by a ball elastic collision model called the elastic collision-based rumor-propagation model (ECRModel). In this paper, the authors have investigated the dynamics of ball elastic collisions, which is similar to the dynamics of rumor propagation between the nodes in Online Social Network. In the proposed model, the authors have divided the users into three groups according to three states, including `inactive and do not spread rumors', `active', and `inactive but have previously spread rumors'. In the ECRModel, multiple parameters are used for the transmission behaviors of the rumors. Individual features with detailed attributes and integral features with node-state densities have been harmoniously considered while designing the algorithm used to set the rumor propagation rules. They have analyzed the probability densities for different types of nodes and determined the steady state both analytically and through simulations. The results prove that the ECRModel is rational and more suitable for analyzing rumor propagationin social networks.

Discussion and Conclusion

The intension of this paper is to present a review of current work related to Fake News Detection and identify future research directions in the field of Fake News Detection. Most of the work is done in detecting whether an event is a rumor or genuine post. Various methodologies like Recurrent Neural Network, Autoencoders and Machine Learning Approaches are used in some works. We mainly tried to review the work done for accuracy improvement and performance improvement of Fake News Detection. The increasing interest in work on fake news detection combined with the challenges and special nature of social media content make this a timely survey that we hope will contribute to the further development of this area.

References

[1] Rumor: Merriam-Webster, : http://www.merriam-webster.com/dictionary/rumor

[2] V. Qazvinian, E. Rosengren, R. R. Radev, and Q. Mei, “Rumor has it: Identifying misinformation in microblogs”, Proc. Conf. Empirical Methods Natural Lang. Process., Edinburgh, U.K., 2011,1589–1599

[3] Weiling Chen, Yan Zhang, Chai Kiat Yeo, Chiew Tong Lau, Bu Sung Lee, “Unsupervised Rumor Detection Based on Users’ Behaviors using Neural Networks”, Pattern Recognition Letters, Elsevier ,2017

[4] Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, Rob Procter, "Detection and Resolution of Rumours in Social Media: A Survey" , ACM Computing Surveys 51, 2, Article 32, 2018,36 pages

[5] Duong C.T., Nguyen Q.V.H., Wang S., Stantic B., " Provenance-Based Rumor Detection", In: Huang Z., Xiao X., Cao X. (eds) Databases Theory and Applications. ADC 2017. Lecture Notes in Computer Science, vol 10538. Springer, Cham

[6] Soroush Vosoughi, Mostafa 'Neo' Mohsenvand, and Deb Roy, "Rumor Gauge: Predicting the Veracity of Rumors on Twitter" ACM Trans. Knowl. Discov. Data 11, 4, Article 50, July 2017, 36 pages.

[7] A. Dang, M. Smit, A. Moh'd, R. Minghim and E. Milios. "Toward Understanding How Users Respond to Rumours in Social Media", 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, 2016, pp. 777-784.

[8] The Pew Research Center .Social Media Fact Sheet http://www.pewinternet.org/fact-sheet/social-media/(2018)

[9] Jing Ma,Wei Gao,PrasenjitMitra,Sejeong Kwon, Bernard J. Jansen,Kam-Fai Wong, Meeyoung Cha, "Detecting Rumors from Microblogs with Recurrent Neural Networks", Proceeding IJCAI'16 Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence ACM Digital Library New York, New York, USA, 2016, Pages 3818-3824

[10] Y. Liu and S. Xu., "Detecting Rumors Through Modeling Information Propagation Networks in a Social Media Environment", IEEE Transactions on Computational Social Systems, 2016, Volume: 3, Issue: 2

[11] G. Tong , Weili Wu, Ling Guo, Deying Li, Cong Liu, Bin Liu, Ding-Zhu Du., "An Efficient Randomized Algorithm for Rumor Blocking in Online Social Networks", IEEE Transactions on Network Science and Engineering (Early Access),2017

[12] B. Wang, G. Chen, L. Fu, L. Song and X. Wang, "DRIMUX: Dynamic Rumor Influence Minimization with User Experience in Social Networks", IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 10, 2017, pp. 2168-2181

[13] Z. Tan, J. Ning, Y. Liu, X. Wang, G. Yang and W. Yang, "ECRModel: An Elastic Collision-Based Rumor-Propagation Model in Online Social Networks", IEEE Access, vol. 4, 2016, pp. 6105-6120.