Fake News Detection Methods

Ruochen Kong
ruochen.kong@emory.edu
Emory University
Atlanta, Georgia, USA

ABSTRACT

Fake news has spread for years on social media as user engagement increases, which diminishes the credibility of real news, and is extremely harmful in the current condition with Covid-19. Although this topic remains novel, several research papers have already developed powerful models for detecting fake news. Their approaches varied significantly on the features, including the linguistic features, the contained figures, the user profiles, and the network of users. In this survey, I aim to compare these models with their accuracy, interpretability, utility, and time complexity. The result would improve future research on fake news detection and would assists researchers to extract features that could be more clearly distinguished.

KEYWORDS

Fake news, Social media, Data mining, Machine learning, Neural networks

1 INTRODUCTION

Because of the threat of Covid-19, the demand for timely news has increased significantly. The most efficient way to obtain the news is to browse through social media. However, there exists numerous fake news. As noted in "Inoculating Against Fake News About COVID-19" (Sander van der Linden et al. 2020), about 50% of the population has engaged in fake news and 25% of Americans believe that vaccines are planted with microchips [5]. The abundance of fake news diminishes the credibility of real news, which is extremely harmful in the current pandemic period. Therefore, a robust automatic fake news detection method is urgently demanded. Current research papers have provided possible features of fake news, including the content, the writing style, the resolution of the figure, the relevance of the figure, the location of the users, the age of the user, and the social network of the user. These papers have investigated various features which are not fully listed here but will be discussed in the following sections. Notice that the current papers are generally based on the datasets that contain the fake news of the 2016 presidential election, so some important features in these models may not maintain the same importance on the Covid-19 datasets, but it is still cost-worthy to compare and summarize the exited models.

2 PROBLEM DEFINITION

The previous papers used different definitions of the fake news detection problem but shared some common variables. The following definition is a preliminary summarization of the previous definitions.

Given the dataset $\mathcal{X}$, the label set $\mathbb{Y}$ is either collected along with the dataset or determined through several processes. The dataset $\mathcal{X}$ can also be separated into $\mathcal{X}^R$ and $\mathcal{X}^F$, where R represents real news and F represents fake news. The features $\mathcal{Z}$ may be extracted from $\mathcal{X}$ along with $\mathcal{Z}^R$ and $\mathcal{Z}^F$. Then the final goal is to develop a function $\mathcal{M} : \mathcal{Z} \rightarrow \mathbb{Y}$.

The format of $\mathcal{X}$ and $\mathcal{Z}$ varies from matrix to graphs, which will be further explained when investigating actual models. The format of $\mathbb{Y}$ is also different based on the specific goals.

3 METHODS

3.1 FNDNet

FNDNet [1] is a machine learning approach which extracted the features from the raw data by utilizeing an exist natural language processing method, GloVe [3]. GloVe calculates the clossness of two words in context and is storded as a matrix, $\mathcal{Z}$. $\mathcal{Z}$ is then passed through sevral steps of CNN and pooling, but lack of explanations. The authors provide an accuracy of 98% of this model. This value, however, would be less believable due to the simultaneously high performance of the baselines, but it would also be because of the importance of GloVe.

3.2 TI-CNN

Different from FNDNet which considers only the text content, TI-CNN includes the text content, the metadata and the attached images [7]. According to the authors, the statistical features of words, sentences, and images are proved to distinguish from real news to fake news, which forms the dataset $\mathcal{X}$. Then $\mathcal{X}$ is passes through CNNs to form the training data $\mathcal{Z}$. Finally, the neural network $\mathcal{Z}$ is trained with back-propagation algorithm. In this paper, the importance of attached images is convincingly proved.

3.3 MMFD

Multi-source Muilti-class Fake News Detection (MMFD) [2] considered even more sources, for example the history of the user engaged in fake news. Additionally, this model used several output classes to represent the degree of fakeness instead of the binary separation. The challenge of this model is the way to combine the features from each source into a coherent dimension, and the method to reliably determine the degree of fakenees for the training dataset.

3.4 Network Based & User Profiles

The previous papers considered mostly on the content of fake news, while the network based approach [8] and the user profiles approach [4] provide another possibility. The features used in the netwoke based approach are carefully compared with each other by the authors which increase the interpretability. The user profiles approach, similarly, also shows the importance of several features. Both papers provide insight of further research.

3.5 Unsupervised Detection Model

Most of the existing models are supervied model, Yang et al. (2019) provied a possible unsupervised framework [6]. However, the framework is stated only with statistical equations which is less explainable in some aspects.

4 CHALLENGES

The most common challenge is that, in general, the users who create fake news deliberately mix real news with artificial information to increase credibility. Thus, models would be blinded if only considering the content. To utilize broader sources such as social networks or user profiles would resolve this problem, but result in another challenge with the time complexity. Thus, creating an optimal balance between them woulg be the current goal.

REFERENCES

[1] Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang, and Soumendu Sinha. 2020. FNDNet – A deep convolutional neural network for fake news detection. *Cognitive Systems Research* 61 (2020), 32–44. https://doi.org/10.1016/j.cogsys.2019.12.005

[2] Hamid Karimi, Proteek Roy, Sari Saba-Sadiya, and Jiliang Tang. 2018. Multi-source multi-class fake news detection. In *Proceedings of the 27th international conference on computational linguistics*. 1546–1557.

[3] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 1532–1543. https://doi.org/10.3115/v1/D14-1162

[4] Kai Shu, Xinyi Zhou, Suhang Wang, Reza Zafarani, and Huan Liu. 2019. The role of user profiles for fake news detection. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining*. 436–439.

[5] Sander van der Linden, Jon Roozenbeek, and Josh Compton. 2020. Inoculating Against Fake News About COVID-19. *Frontiers in Psychology* 11 (2020). https://doi.org/10.3389/fpsyg.2020.566790

[6] Shuo Yang, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu, and Huan Liu. 2019. Unsupervised fake news detection on social media: A generative approach. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 5644–5651.

[7] Yang Yang, Lei Zheng, Jiawei Zhang, Qingcai Cui, Zhoujun Li, and Philip S Yu. 2018. TI-CNN: Convolutional neural networks for fake news detection. *arXiv preprint arXiv:1806.00749* (2018).

[8] Xinyi Zhou and Reza Zafarani. 2019. Network-based fake news detection: A pattern-driven approach. *ACM SIGKDD explorations newsletter* 21, 2 (2019), 48–60.