Deep Learning Techniques for Fake News Detection

Kairui Fang

Department of Mathematics, University of California, San Diego, La Jolla, United States

kafang@ucsd.edu

Abstract. Social media has recently become the primary source for people to consume news. Plenty of users prefer to go to social media apps such as Twitter, Facebook, and Snapchat to obtain the latest social events and news. Meanwhile, traditional media is emulating the new media to post their news on the aforementioned apps. This prevalence is a double-edged sword, for the advantage is that users can easily gain access to the news articles they look for on social media. However, it also provides an ideal platform for fake news propagation. The spread of fake news is extremely fast on social media and can cause adverse effects in real life. The unregimented, incomplete censorship and the absence of fact-checking processes make fake news easy to propagate and hard to control. Therefore, fake news detection on social media has become a trending topic that draws tremendous attention, as shown in figure 1. Nevertheless, as pundits dig into the realm of deep learning, some of the studies utilize deep neural networks (DNN) to build frameworks that would help detect fake news. Although impressive progress on the topic has been made, the lack of a review dissertation that summarizes and synthesizes the overall development of the study would be problematic. Hence, this paper aims to summarize different models implemented in recent studies that improve the veracity of fake news detection.

Keywords: Object detection, convolutional neural network, deep learning.

1. Introduction

Given that people are inclined to consume news on social media [1] for its convenience and perpetual activeness, it inevitably becomes the fertile soil for fake news to grow exponentially [2]. In recent years, fake news on social media has incessantly caused enormous problems online and in people's real lives. For instance, during the 2016 U.S. presidential election, former president Donald Trump's election team attempted to hinder the electoral result by creating and sharing tons of unverified news that was in favor of Donald Trump and against Hillary Clinton [3]. The former eventually won the election and became the 45th president of the U.S. According to a post-election investigation, of the approximately 30 million tweets that contain a link to a news website, 10% point to sites with fake news or conspiracy theories, and 15% of them point to sites with news that is blatantly biased [4]. The upsurge of fake news on social media has adverse effects on individuals. People would tend to rationalize harmful behaviors [5], lose fundamental justification, and believe in science — no longer trust its ability to provide answers [6]. Generally, one can quickly create an account for posting endless content they intend to show to people on apps like Twitter, Facebook, and Snapchat since it is rather difficult to manually identify and regulate such users on social media. Therefore, detecting fake news on social media with the help of computers is now a necessary task for researchers.

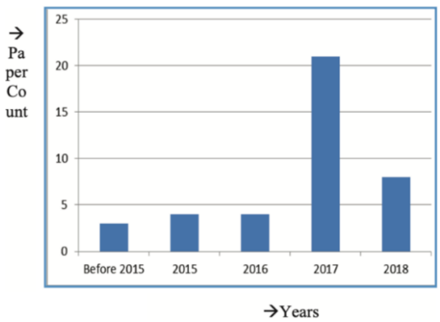

Figure 1. Studies on fake news detection

Machine learning and data mining were utilized to detect fake news in the first place. Such methods require manually constructing feature engineering of the specific task, which captures features from raw data to obtain representations of the task and is fundamental in enhancing model accuracy, then introducing advanced machine learning algorithms to classify the target task and make predictions based on the computer's decisions. In recent years, we have testified many fake news detection models pertaining to the above methods as their ability to justify the accuracy of the news by factors such as emotions, the speaker's political affiliations, and jobs, based on a machine learning trained expert system [7]; The support vector machine method (SVM) which classifies data (i.e., news) and conduct regression analysis using supervised learning models [8]. Although heuristic results are produced, there are issues regarding the methods. The efficiency of ML is one of the protuberant problems because its feature-building process requires colossal manual work that is challenging and complex. Also, there are no significant results of fake news detection applying ML methods. The deep neural network (DNN) is introduced to the field with better performance in featuring and classifying data.

Nevertheless, there are currently very few reviews on fake news detection models based on DNN. This results in that many novices are not well informed of the latest progress of the task and what new methods have been proposed when they are exposed to it. In addition, a review can be seen as a phased summary that contributes to the development of follow-up research. Therefore, we have elucidated and summarized the development of this task in recent years. We first distinguish the definition of the task and then introduce the commonly used datasets and evaluation metrics for the task. In addition, we also explain the mainstream deep learning methods used for the task. Finally, we introduce some state-of-the-art methods and potential challenges in this task.

2. Background

2.1. Task Definition

2.1.1 Deep Learning Model

The definition of the task is detecting fake news using models that are based on DNN. Previous studies drew attention to applying ML in fake news detection. However, the weaknesses of ML were explored more as research fields enlarged in the area. First, building up features of the task, which is the most critical procedure for the machine learning method, is based on hand-crafted data [9]. The extracted feature may be biased because the feature extraction process is manual and convoluted. Also, ML methods fail to capture features from complex semantic information, resulting in the lack of outstanding and accurate predictions in fake news detection. Deep learning (DL) methods defeat the traditional ML methods for their extraordinary ability in feature extraction. In contrast, DL models have the following advantages:

1) Feature engineering becomes unnecessary. Deep learning can execute the construction process on its own by analyzing the data and searching for features automatically.

2) Unstructured data (data in different forms) can also be learned. Deep learning transfers these data into real-valued vectors and looks for correlation among them [10].

3) Efficient while delivering high-quality outcomes. The deep neural network can rapidly process thousands of complex and repetitive tasks while guaranteeing high-quality outcomes.

4) Deep learning models can capture concealed representations from simpler inputs such as news content and context varieties [9].

5) DNN requires shorter data classification time than other methods like SVM [11].

Figure 2 illustrates the structure of the DL model used in fake news detection. The initial procedure is to find a dataset from public dataset websites, and data pre-processing then transfers the received data to a neural network [12]. The word vector model maps the data (news content) into real-valued vectors. The most frequently used word embedding techniques in previous studies are TF-IDF, Word2Vec, and FastText [13]. Fine-grained data is sent to the prediction-making process, while other data is trained by the neural network-based training model and labeled predictions.

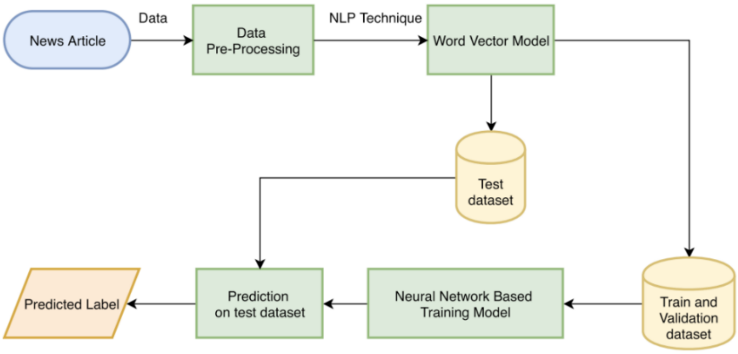

Figure 2. DL structure in fake news detection [9]

2.1.2 Types of Methods

Once the models are considered sophisticated, social media apps would apply them to help detect fake news such that misinformation could be diminished. The fake news detection frameworks can be categorized into two types: the content-based method and the network-based method.

1) The *content-based methodology*. It is somewhat straightforward to focus on the content since it is the essence of news. The content-based method captures the textual feature of a news event [14], as news on social media often includes textual descriptions. Generally, the content-based model learns a series of text characteristics, such as language styles [15], writing consistency [16], social emotions [17], and topic features [18]. However, the traditional method has drawbacks in capturing visual features. Thus, recent models apply DNN to learn both textual and visual feature representations by extracting semantic information in the news [19]. Different types of DNN, such as convolutional neural networks [20], are also applied in the method. Moreover, attention mechanism has procured a secured position in utilization for its automaticity in obtaining complex contextual patterns of the content [21]. There are issues regarding the method; we will elaborate more in 3.

2) The network (propagation)-based methodology requires a certain amount of user interactions and extracts feature through activities such as comments, retweets, posts, and following numbers [22] of these accounts. Essentially, it learns the propagation patterns of fake news to determine the authenticity. To construct such a model, homogeneous networks would be a suitable tool. Homogeneous networks are comprised of graphs with the same types of edges and nodes. Hence, it vividly demonstrates the relationships between the propagation feature and other related information.

2.2. Datasets

Datasets are obtained from fact-checking sources such as PolitiFact, Snopes, and FactCheck. The data is generally comprised of a tremendous number of posts on different social media platforms. Some of the widely used datasets are:

1) CREDBANK: The dataset specifically gathers Twitter news stream. It is divided into four files and is made up of crowdsourced data from 60 million tweets. It uses a 30-dimension vector of believability classes to classify tweets into 1049 events and has verified 1000 topics.

2) Reuters: This is a multi-label dataset with 90 classes. It contains 7769 training documents, of which 3019 are testing materials. The documents are culled from a single source, which increases the quantity of biased data for detection.

3) LIAR: The fact-checking website PolitiFact provided the data for this dataset. Instead of whole articles, it classifies 12,836 brief assertions into the following categories: pants-fire, true, half-true, barely true, and false [1].

4) BuzzFace: It divides 2263 articles into four categories: mostly true, mostly false, a mixture of true and false, and no factual information. The dataset collects several types of news feeds, including texts, images, comments on social media apps, and URL links.

Evaluation Metrics

The evaluation of the outputs has three parts: Accuracy (A), Precision (P), and Recall (R). The elements in the evaluation matrix of the outcome are true positive, true negative, false positive, and false negative.

The accuracy score is calculated by:

$$A = \frac{TruePositive + TrueNegative}{SumofPredictions} \tag{1}$$

Where A is the percentage of accurate predictions made by the model out of the total predictions.

The precision score is calculated by:

$$P = \frac{TruePositive}{Positive + FalsePositive} \tag{2}$$

Where P is the number of actual accurate predictions divided by both true and false positive results (total accurate predictions), which includes falsely recognized accurate predictions.

The recall score is calculated by:

$$R = \frac{TruePositive}{TruePositive + FalseNegative} \tag{3}$$

Where R is the number of actual accurate predictions divided by the total accurate predictions which include predictions that are falsely recognized as inaccurate.

3. Challenges

Although the deep learning method outperforms other methods in the veracity of data prediction, there are still improvable aspects in the purview of fake news detection:

1) Model selection significantly affects the prediction outcomes. Researchers should select models in accordance with particular features of the data. RNN is a suitable tool for the long textual feature, but there are merely a few studies that utilize such a model [9].

2) Nowadays, news feeds on social media are not only in text but in visual types. The prominent ability of feature catching for deep learning allows researchers to apply visual data, such as images and videos. Studies in this domain are scarce. Therefore, it is necessary to implement visual data in order to reach more comprehensive results.

3) The content-based method that applies deep learning is problematic. First, the approach that captures content features is proven ineffective in the actual detecting process [23] because it is somewhat difficult to contrast the content of fake news from the real ones. Second, the lack of background information checking in the method results in inaccurate outcomes while making predictions.

4. Recent Researches

In this section, we revise several recent proposed models for fake news detection. We not only demonstrate the problems addressed by them, but also revisit their neural-based methods in detail.

Literature [24] demonstrates a fake news detection model that takes realistic conditions into account. This work addresses several problems in previous methods: Firstly, the content-based approach relies on the length of fake news. Long texts like news articles are optimal objects to learn. However, social media tweets are mostly short texts. Secondly, some advanced models require enormous user comments to learn opinions from users, which are crucial in identifying fake news. However, users on social media do not leave their comments as often. Thirdly, Retweeting is useful for classifying information. However, it is rather hard to get the diffusion structure due to privacy concerns. Finally, Existing models cannot explain who the potential users are to support fake news.

Therefore, a novel model GCAN is designed in the paper to operate the detection process better. First, they obtain user features such as word embeddings by browsing through user profiles and social interactions. Second, based on user features, CNN is utilized to study the representation of retweet propagation. Third, they build a graph and its convolution network to model and learn interactions between users. Fourth, a dual co-attention mechanism is developed for learning correlations and co-influences between the original tweet and the retweets. Lastly, GCAN generates a binary prediction based on the features.

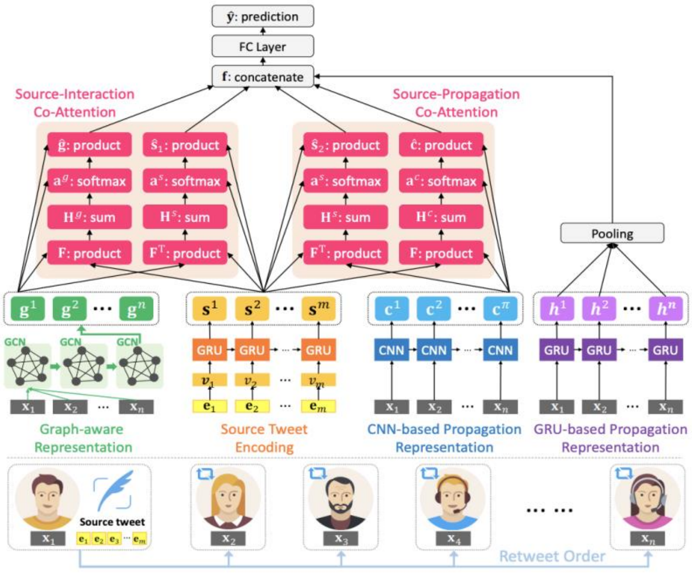

Figure 3. The architecture of the GCAN model [26]

Literature [25] fixates on detecting fake news by the evidence-based approach where external evidence is provided to verify a claim's accuracy (fake news). Models using this approach must explore and pack up sufficient information in evidence during the claim verification process. The models of the existing methods, which first capture the semantics of claims evidence and then model their interactions to observe the semantic coherence, are insufficient due to the neglect of investigating the exquisite semantics of claims and evidence. Specifically, the issues of previous methods are: The long-distance semantic dependency is hard to capture; Superfluous information is preserved by the model, which is harmful to the claim-evidence interaction process.

To solve the issues, the Graph-based sEmantic sTructure (GET) mining model is proposed to explore exquisite semantics. GET models both claims and evidences by graph structure. Nodes in the graph represent phrases, and edges indicate the connection between two phrases. In the structure, unnecessary nodes are discarded based on the complex linguistic structure to reduce the redundant information in semantics. Then GET implements the attention mechanism used in previous studies to decipher information from nodes and transfer it into representatives of claims and evidence. Lastly, the claim verification prediction would be conducted using the integrated information.

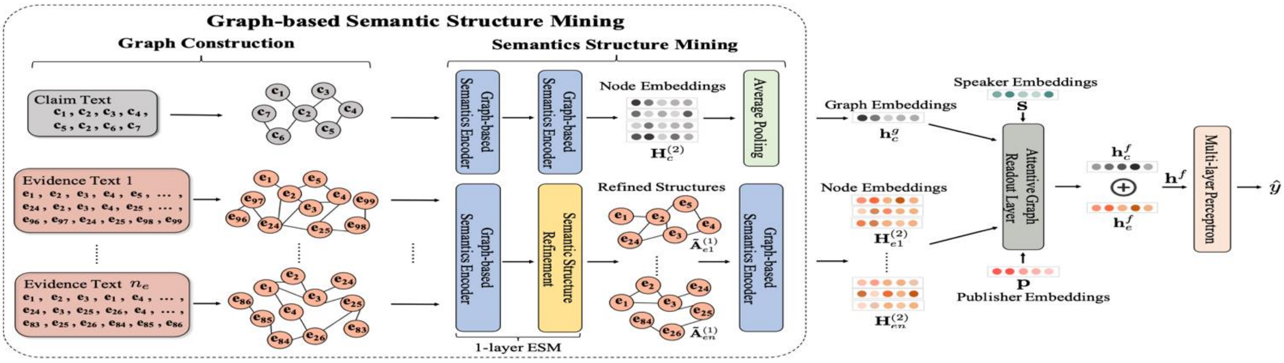

Figure 4. The GET model [27]

Literature [26] is proposed to identify fake news by the user-based method. Compared to the spam, fake news has more significant impact on society, and more interactions with users for they can actively receive and share fake news on social media.

The model transfers the detecting process to the classification of news content where it determines the trustworthiness of articles, subjects, and creators concurrently through a prediction model. Furthermore, FAKEDETECTOR learns the feature representation of news content and other user information (subjects and creators) through an innovative hybrid feature learning unit, HFLU. Then, the gated diffusive unit (GDU) embedded in deep diffusion model is involved to complete the identification process. Specifically, FAKEDETECTOR is comprised of two parts:

1) Representation feature learning: learn features from text data through HFLU; obtain the explicit/potential feature vector for articles, creators, and subjects; append the explicit and potential feature vector to form representations (inputs) of news articles (A), target subjects (S) and news creators (C).

2) Credibility label inference: the deep diffusive graph neural network model demonstrates the correlation between the credibility of news articles, target subjects and news creators. GDU models the relationship of the obtained three vectors simultaneously and outputs new vectors called the output vector of A, S and C. Lastly, the model transforms the output vectors to their credibility labels.

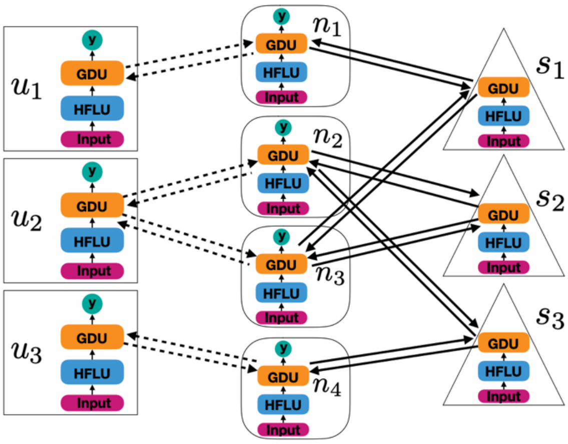

Figure 5. The architecture of FAKEDETECTOR [28]

Literature [27] focuses on the propagation-based method of fake news detection. The paper argues that previous works use static network structures that merely capture the graph structures with no temporal dynamic process (not a time-aware structure) to detect fake news, resulting in the insufficiency of real-world information diffusion networks. Hence, a novel temporal propagation-based fake news detection framework, TGNF in short, is designed to obtain temporal information in static networks and help identify false news. The TGNF system contains two main components:

1) News propagation modeling: capture the adaptable structure, syntax, and temporal information by TGAT; use continuous-time dynamic graph since the propagation evolves with time.

2) Temporal difference network (TDN): capture the information on variation between interactions.

5. Conclusion

We analyze the necessity of fake news detection on social media. Recent studies use machine learning and deep neural networks as based models. We briefly introduce and compare the methodology of both systems and how it functions in detecting fake news. Moreover, we summarized the classification of the DNN-based fake news detection approaches. The broadly used dataset and evaluation metrics are introduced. Based on DNN, several models such as GCAN, GET, and TGNF that enhance the accuracy of fake news detection are mentioned in the paper to help summarize recent research on the topic.

References

[1] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19(1), 22-36.

[2] Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. Information Processing & Management, 57(2), 102025.

[3] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of economic perspectives, 31(2), 211-36.

[4] Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. Nature communications, 10(1), 1-14.

[5] Sadio, A. J., Gbeasor-Komlanvi, F. A., Konu, R. Y., Bakoubayi, A. W., Tchankoni, M. K., Bitty-Anderson, A. M., ... & Ekouevi, D. K. (2021). Assessment of self-medication practices in the context of the COVID-19 outbreak in Togo. BMC public health, 21(1), 1-9

[6] Eichengreen, B., Aksoy, C. G., & Saka, O. (2021). Revenge of the experts: Will COVID-19 renew or diminish public trust in science?. Journal of Public Economics, 193, 104343.

[7] Fitzpatrick, B., Liang, X., & Straub, J. (2021). Fake news and phishing detection using a machine learning trained expert system. arXiv preprint arXiv: 2108.08264.

[8] Aphiwongsophon, S., & Chongstitvatana, P. (2018, July). Detecting fake news with machine learning method. In 2018 15th international conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON) (pp. 528-531). IEEE.

[9] Mridha, M. F., Keya, A. J., Hamid, M. A., Monowar, M. M., & Rahman, M. S. (2021). A Comprehensive Review on Fake News Detection with Deep Learning. IEEE Access.

[10] Xu, Z., & Sun, J. (2018). Model-driven deep-learning. National Science Review, 5(1), 22-24.

[11] Hiramath, C. K., & Deshpande, G. C. (2019, July). Fake news detection using deep learning techniques. In 2019 1st International Conference on Advances in Information Technology (ICAIT) (pp. 411-415). IEEE.

[12] Girgis, S., Amer, E., & Gadallah, M. (2018, December). Deep learning algorithms for detecting fake news in online text. In 2018 13th international conference on computer engineering and systems (ICCES) (pp. 93-97). IEEE.

[13] Deepak, S., & Chitturi, B. (2020). Deep neural approach to Fake-News identification. Procedia Computer Science, 167, 2236-2243.

[14] Song, C., Yang, C., Chen, H., Tu, C., Liu, Z., & Sun, M. (2019). CED: credible early detection of social media rumors. IEEE Transactions on Knowledge and Data Engineering, 33(8), 3035-3047.

[15] Popat, K. (2017, April). Assessing the credibility of claims on the web. In Proceedings of the 26th international conference on world wide web companion (pp. 735-739).

[16] Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2017). A stylometric inquiry into hyperpartisan and fake news. arXiv preprint arXiv:1702.05638.

[17] Guo, C., Cao, J., Zhang, X., Shu, K., & Yu, M. (2019). Exploiting emotions for fake news detection on social media. arXiv preprint arXiv:1903.01728.

[18] Castillo, C., Mendoza, M., & Poblete, B. (2011, March). Information credibility on twitter. In Proceedings of the 20th international conference on World wide web (pp. 675-684).

[19] Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2021). Detecting fake news with capsule neural networks. Applied Soft Computing, 101, 106991.

[20] Ajao, O., Bhowmik, D., & Zargari, S. (2018, July). Fake news identification on twitter with hybrid cnn and rnn models. In Proceedings of the 9th international conference on social media and society (pp. 226-230).

[21] Chen, Y., Sui, J., Hu, L., & Gong, W. (2019, November). Attention-residual network with CNN for rumor detection. In Proceedings of the 28th ACM international conference on information and knowledge management (pp. 1121-1130).

[22] Liu, Y., Jin, X., & Shen, H. (2019). Towards early identification of online rumors based on long short-term memory networks. Information Processing & Management, 56(4), 1457-1467.

[23] Shu, K., Mahudeswaran, D., Wang, S., & Liu, H. (2020, May). Hierarchical propagation networks for fake news detection: Investigation and exploitation. In Proceedings of the international AAAI conference on web and social media (Vol. 14, pp. 626-637).

[24] Lu, Y. J., & Li, C. T. (2020). GCAN: Graph-aware co-attention networks for explainable fake news detection on social media. arXiv preprint arXiv: 2004.11648.

[25] Xu, W., Wu, J., Liu, Q., Wu, S., & Wang, L. (2022, April). Evidence-aware Fake News Detection with Graph Neural Networks. In Proceedings of the ACM Web Conference 2022 (pp. 2501-2510).

[26] Zhang, J., Dong, B., & Philip, S. Y. (2020, April). Fakedetector: Effective fake news detection with deep diffusive neural network. In 2020 IEEE 36th international conference on data engineering (ICDE) (pp. 1826-1829). IEEE.

[27] Song, C., Shu, K., & Wu, B. (2021). Temporally evolving graph neural network for fake news detection. Information Processing & Management, 58(6), 102712.