

Machine Learning Based Fake News Detection

Mahesh Pawar, Anuj Kasar, Rohit Satpute, Dr. M. R. Bendre
Pravara Rural Engineering College, Loni, Maharashtra, India

Abstract: *In Social media, it is a popular medium for the dissemination of real-time news all over the world. Easy and quick information proliferation is one of the reasons for its popularity. An extensive number of users with different age groups, gender, and societal beliefs are engaged in social media websites. Despite these favorable aspects, a significant disadvantage comes in the form of fake news, as people usually read and share information without caring about its genuineness. It is imperative to research methods for the authentication of news. To address this issue, this article proposes a two-phase benchmark model named FakeNews based on word embedding (WE) over linguistic features for fake news detection using machine learning classification. The first phase preprocesses the data set and validates the veracity of news content by using linguistic features. The second phase merges the linguistic feature sets with WE and applies classification. To validate its approach, this article also carefully designs a novel FakeNewsdata set with approximately thousands articles, which incorporates different data sets to generate an unbiased classification output.*

Keywords: Fake News, User Profile, Trust Analysis, machine learning, Social Media

I. INTRODUCTION

The reliability of information diffused on the World Wide Web (WWW) is a central issue of modern society. In particular, in the recent years the spreading of misinformation and fake news on the Internet has drawn increasing attention, and has reached the point of dramatically influencing political and social realities. As an example, showed the significant impact of fake news in the context of the 2016 US presidential elections; analyzed the most viral tweets related to the Boston Marathon blasts in 2013, finding that the share of rumors and fake content was higher than the share of true information. As an increasing amount of our lives is spent interacting online through social media platforms, more and more people tend to seek out and consume news from social media rather than traditional news organizations. The reasons for this change in consumption behaviors are inherent in the nature of these social media platforms: (i) it is often more timely and less expensive to consume news on social media compared with traditional news media, such as newspapers or television; and (ii) it is easier to further share, comment on, and discuss the news with friends or other readers on social media. For example, 62 percent of U.S. adults get news on social media in 2016, while in 2012; only 49 percent reported seeing news on social media¹. It was also found that social media now outperforms television as the major news source². Despite the advantages provided by social media, the quality of news on social media is lower than traditional news organizations.

However, because it is cheap to provide news online and much faster and easier to disseminate through social media, large volumes of fake news, i.e., those news articles with intentionally false information, are produced online for a variety of purposes, such as financial and political gain. It was estimated that over 1 million tweets are related to fake news³ by the end of the presidential election. Given the prevalence of this new phenomenon, "Fake news" was even named the word of the year by the Macquarie dictionary in 2016. The extensive spread of fake news can have a serious negative impact on individuals and society. First, fake news can break the authenticity balance of the news ecosystem. For example, it is evident that the most popular fake news was even more widely spread on Facebook than the most popular authentic mainstream news during the U.S. 2016 president election. Second, fake news intentionally persuades consumers to accept biased or false beliefs. Fake news is usually manipulated by propagandists to convey political messages or influence. For example, some report shows that Russia has created fake accounts and social bots to spread false stories. Third, fake news changes the way people interpret and respond to real news. For example, some fake news was just created to trigger people's distrust and make them confused; impeding their abilities to differentiate what is true from

what is not6. To help mitigate the negative effects caused by fake news-both to benefit the public and the news ecosystem-It's critical that we develop methods to automatically detect fake news on social media.

Detecting fake news on social media poses several new and challenging research problems. Though fake news itself is not a new problem-nation or groups have been using the news media to execute propaganda or influence operations for centuries-the rise of web-generated news on social media makes fake news a more powerful force that challenges traditional journalistic norms. There are several characteristics of this problem that make it uniquely challenging for automated detection. First, fake news is intentionally written to mislead readers, which makes it nontrivial to detect simply based on news content. The content of fake news is rather diverse in terms of topics, styles and media platforms, and fake news attempts to distort truth with diverse linguistic styles while simultaneously mocking true news. For example, fake news may cite true evidence within the incorrect context to support a non-factual claim. Thus, existing hand-crafted and data-specific textual features are generally not sufficient for fake news detection. Other auxiliary information must also be applied to improve detection, such as knowledge base and user social engagements. Second, exploiting this auxiliary information actually leads to another critical challenge: the quality of the data itself. Fake news is usually related to newly emerging, time-critical events, which may not have been properly verified by existing knowledge bases due to the lack of corroborating evidence or claims. In addition, users' social engagements with fake news produce data that is big, incomplete, unstructured, and noisy. Effective methods to differentiate credible users extract useful post features and exploit network interactions are an open area of research and need further investigations.

II. RELATED WORK

Literature survey is the most important step in any kind of research. Before start developing we need to study the previous papers of our domain which we are working and on the basis of study we can predict or generate the drawback and start working with the reference of previous papers.

“In this section, we briefly review the related work on fake news detection system and their different techniques.

In this paper [1], the results of a fake news identification study that documents the performance of a fake news classifier are presented. The Textblob, Natural Language, and SciPy Toolkits were used to develop a novel fake news detector. Advantages-1. Used natural language processing2. Fake news detection based on attribute classification Disadvantages-Time consuming process.

This paper [2] introduce the datasets which contain both fake and real news and conduct various experiments to organize fake news detector. Advantages is 1. Used Natural Language Processing, Machine learning and deep learning techniques to classify the datasets2.Accuracy is better and disadvantages is use Limited dataset.

This paper [3] proposed a distributed framework to implement the proposed truth discovery scheme using Work Queue in an HTCCondorsystem. Advantages is 1. Find trustworthy information on Social media2.Proposed truth discovery scheme using Work Queue in an HTCCondor system and disadvantages is Accuracy is low

This Paper [4] Studied various detection techniques i.e. content based, social context based and hybrid based. Advantages is Proposed content-based, social context-based and hybrid-based methods and disadvantages is only survey state of the methods.

This paper [5] Present a new fake news detection model using unified key sentence information which can efficiently perform sentence matching between question and article by using key sentence retrieval based on bilateral multi perspective matching model. Advantages is Implement natural language processing using key sentence retrieval and disadvantages is Fake news detection accuracy is low.

This Paper [6] classifies fake news messages from Twitter posts using hybrid of convolutional neural networks and long-short term recurrent neural network models. Advantages is Implement hybrid CNN and RNN Models and Accuracy is much better. Disadvantages is only consider tweet headlines.

This paper [7] Compare news to other sources in 2016 year. Advantages is 1.detect 2016 election fake news spread through social media2. Goal in this paper is to offer theoretical and empirical background to frame this debate. Disadvantages is1. Limited dataset used2. Limited to 2016 news only.

This paper [8] shows a new approach for fake news detection using naive Bayes classifier. Use Implement naïve bayes machine learning algorithm but accuracy is low.

This paper [9] introduced the basic concepts and principles of fake news in both traditional media and social media. In the detection phase, we reviewed existing fake news detection approaches from a data mining perspective, including feature extraction and model construction. Advantages is in this paper, they explored the fake news problem by reviewing existing literature in two phases i.e. characterization and detection but on Use static data.

This study [10] contributes to the scientific knowledge regarding the influence of the interaction between various types of media use on political effects. Advantage is Used multiple news sources for fake news detection and disadvantage is Focus on only political data

III. PROPOSED APPROACH

Online news can be collected from different sources, such as news agency homepages, search engines, and social media websites. However, manually determining the veracity of news is a challenging task, usually requiring annotators with domain expertise who performs careful analysis of claims and additional evidence, context, and reports from authoritative sources. Generally, news data with annotations can be gathered in the following ways: Expert journalists, Fact-checking websites, Industry detectors, and Crowd-sourced workers.

News content features describe the Meta information related to a piece of news. A list of representative news content attributes is listed below:

Source: Author or publisher of the news article.

Headline: Short title text that aims to catch the attention of readers and describes the main topic of the article

Body Text: Main text that elaborates the details of the news story; there is usually a major claim that is specifically highlighted and that shapes the angle of the publisher

3.1 System Diagram

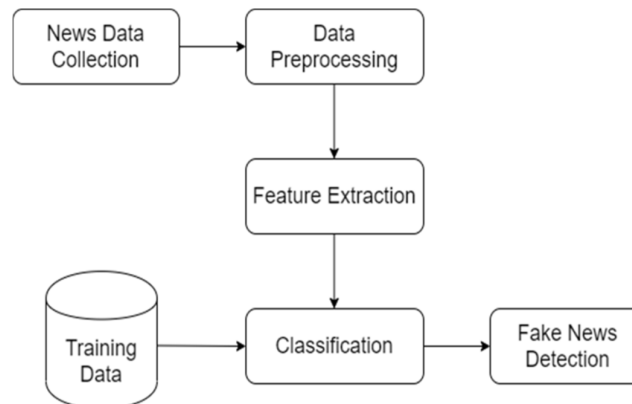


Fig 1. System Architecture

IV. CONCLUSION

Growing popularity of social media, more and more people consume social media news instead of traditional media. However, social media have also been used to disseminate false news, which has strong negative impacts on individual users and the wider society. Here to explore the problem of false news by reviewing existing literature in two phases: characterization and detection. In the characterization phase, we introduce the basic concepts and principles of false news in both traditional media and social media. In the detection phase, we reviewed the current false news detection approaches from a machine learning perspective, including feature extraction and model building. We also discuss evaluation metrics, and future promising directions in fake detection research and expand the field to other applications.

REFERENCES

- [1]. Terry Traylor, Jeremy Straub, Gurmeet, Nicholas Snell” Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as a Supervised Learning Estimator”2019.

- [2]. Rohit Kumar Kaliyar” Fake News Detection Using A Deep Neural Network”2018.
- [3]. Daniel (Yue) Zhang, Dong Wang, Nathan Vance, Yang Zhang, and Steven Mike” On Scalable and Robust Truth Discovery in Big Data Social Media Sensing Applications” 2018.
- [4]. ZaitulIradahMahid,SelvakumarManickam,ShankarKaruppayah” Fake News on Social Media: Brief Review on Detection Techniques” 2018.
- [5]. Namwon Kim, DeokjinSeo, Chang-Sung Jeong” FAMOUS: Fake News Detection Model based on Unified Key Sentence Information” 2018.
- [6]. OluwaseunAjao, DeepayanBhowmik, ShahrzadZargari” Fake News Identification on Twitter with Hybrid CNN and RNN Models”2018.
- [7]. Hunt Allcott Matthew Gentzkow” SOCIAL MEDIA AND FAKE NEWS IN THE 2016 ELECTION” 2017
- [8]. MykhailoGranik, VolodymyrMesyura” Fake News Detection Using Naive Bayes Classifier”2017
- [9]. Kai Shu , Amy Sliva , Suhang Wang , Jiliang Tang , and Huan Liu” Fake News Detection on Social Media: A Data Mining Perspective”2016
- [10]. MeitalBalmas” When Fake News Becomes Real: Combined Exposure to Multiple News Sources and Political Attitudes of Inefficacy, Alienation, and Cynicism” 2014
- [11]. Daniel (Yue) Zhang, Chao Zheng, Dong Wang, Doug Thain, Chao Huang, Xin Mu, Greg Madey .”Towards Scalable and Dynamic Social Sensing Using A Distributed Computing Framework.” Department of Computer Science and Engineering Department of Aerospace and Mechanical Engineering University of Notre Dame Notre Dame, IN, USA IEEE 2017.
- [12]. Daniel (Yue) Zhang, Dong Wang, Hao Zheng, Xin Mu, Qi Li , Yang Zhang.” Large-scale Point-of-Interest Category Prediction Using Natural Language Processing Models.” Department of Computer Science and Engineering Department of Aerospace and Mechanical Engineering University of Notre Dame Notre Dame, IN, USA IEEE 2017.
- [13]. Daniel (Yue) Zhang, Rungang Han, Dong Wang, Chao Huang.” On Robust Truth Discovery in Sparse Social Media Sensing.” Department of Computer Science and Engineering University of Notre Dame Notre Dame, IN, USA IEEE 2016