EasyChair Preprint

№ 8285

Fake News Social Media: a Data Science Perspective

Vishal Gir Goswami

June 18, 2022

Fake News Social Media: a Data science Perspective

Vishal Gir Goswami

Vishalgirgoswami.2016@gmail.com

Chandighar University, Gharuan, Punjab

*Navneet Kaur

ABSTRACT – Today, in the age of technology, and while we spend most of our interacting online through social media platform, it has surfaced as a huge monster. Fake news, information bubbles, news manipulation and the lack of trust in the media are growing problems within our society. However, in order to start addressing this problem, an in-depth understanding of fake news and its origins is required. Social media for news consumption is a double-edged sword. On the one hand, its low cost, easy access, and rapid dissemination of information lead people to seek out and consume news from social media. On the other hand, it enable the wide spread of “fake news”, i.e., low quality news with intensional false information. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society one of the main reasons is that often the interpretation of the news requires the knowledge of political or social context or ‘common sense’, which current natural language processing algorithms are still missing. Recent studies have empirically shown that fake and real news spread differently on social media, forming propagation patterns that could be harnessed for the automatic fake news detection.

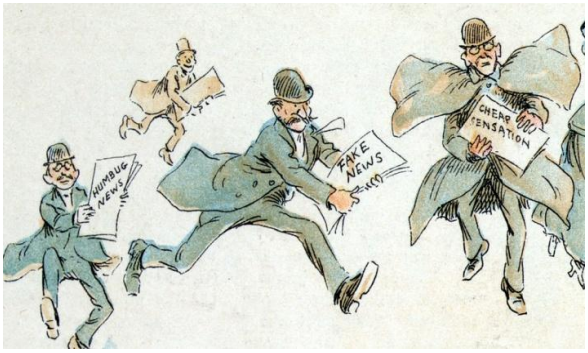
Introduction - Today, in the age of technology, and while we spend most of our interacting online through social media platform, it has surfaced as a huge monster. Fake news, information bubbles, news manipulation and the lack of trust in the media are growing problems within our society.. However, in order to start addressing this problem, an in-depth understanding of fake news and its origins is required**.** 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 it is easier to further share, comment on, and discuss the news with friends or other readers on social media. It was also found that social media now outperforms television as the major news source. 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

What is Fake news?

Fake news is a neologism. Fake news or fake news websites, have no basis in are presented as being factually accurate

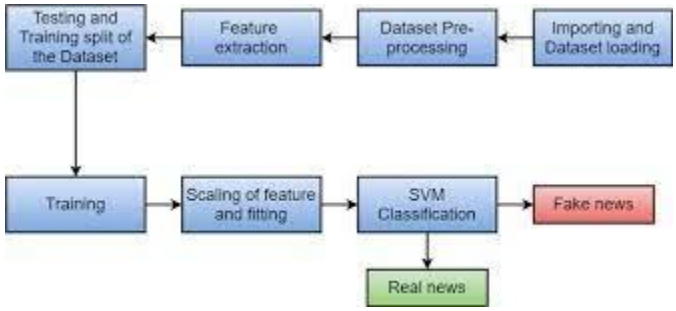
It's designed to deliberately spread hoaxes propaganda, and did information.

Fake News is different form satirical sites like 'the onion', a humour-based website. Fake News stories usually spread through social media sites like Face book, Twitter and Reedit.



FAKE NEWS DETECTION

In the previous section, we introduced the conceptual characterization of traditional fake news and fake news in social media. Based on this characterization, we further explore the problem definition and proposed approaches for fake news detection



In this subsection, we present the details of mathematical formulation of fake news detection on social media. Specifically, we will introduce the definition of key components of fake news and then present the formal definition of fake news detection. The basic notations are defined below

- Let a refer to a News Article. It consists of two major components: Publisher and Content. Publisher ~pa includes a set of profile features to describe the original author, such as name, domain, age, among other

attributes. Content ~ca consists of a set of attributes that represent the news article and includes headline, text, image, etc

. • We also define Social News Engagements as a set of tuples Ɛ = {eit} to represent the process of how news spread over time among n users U = {u1, u2, ..., un} and their corresponding posts P = {p1, p2, ..., pn} on social media regarding news article a. Each engagement € = {ui, pi, t} represents that a user ui spreads news article a using pi at time t. Note that we set t = Null if the article a does not have any engagement yet and thus ui represents the publisher.

ASSESSING DETECTION EFFICACY

Datasets:- 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

Evaluation Metrics

To evaluate the performance of algorithms for fake news detection problem, various evaluation metrics have been used. In this subsection, we review the most widely used metrics for fake news detection. Most existing approaches consider the fake news problem as a classification problem that predicts whether a news article is fake or not:

• True Positive (TP): when predicted fake news pieces are actually annotated as fake news;

• True Negative (TN): when predicted true news pieces are actually annotated as true news;

• False Negative (FN): when predicted true news pieces are actually annotated as fake news;

• False Positive (FP): when predicted fake news pieces are actually annotated as true news.

RELATED AREAS

We further discuss areas that are related to the problem of fake news detection. We aim to point out the differences between these areas and fake news detection by briefly explaining the task goals and highlighting some popular methods.

- Rumor classification
- Truth discovery
- Click bait detection
- Spammer and Bot detection
- We present some open issues in fake news detection and future research directions. Fake news detection on social media is a newly emerging research area, so we aim to point out promising research directions from a data mining perspective. Specifically, as shown in Figure 2, we outline the research directions in four categories: Data oriented, Feature-oriented, Model-oriented and Application oriented.

Data-oriented

Data-oriented fake news research is focusing on different kinds of data characteristics, such as: dataset, temporal and psychological.

Detecting fake news early can help prevent further propagation on social media. From a psychological perspective, different aspects of fake news have been qualitatively explored in the social psychology literature. But quantitative studies to verify these psychological factors are rather limited. For example, the echo chamber effect plays an important role for fake news spreading in social media.

Then how to capture echo chamber affects and how to utilize the pattern for fake news detection in social media could be an interesting investigation.

Feature-oriented

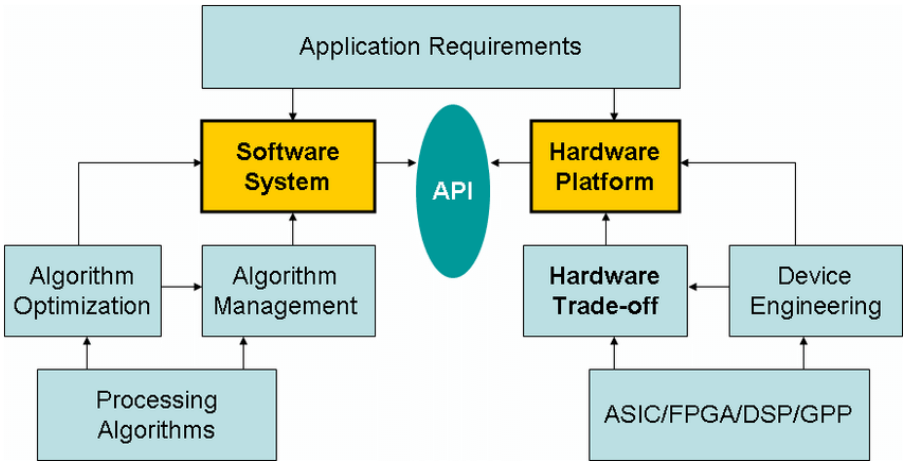
Features for detecting fake news from multiple data sources we have demonstrated that there are two major data sources: news content and social context.

Model-oriented

Model-oriented fake news research opens the door to building more effective and practical Feature-oriented fake news research aims to determine effective models for fake news detection. Most previously mentioned approaches focus on extracting various features, incorporating theses features into supervised classification models, such as naive Byes, decision tree, logistic regression, k nearest neighbor (KNN)and support vector machines (SVM), and then selecting the classifier that performs the best

Application-oriented

Application-oriented fake news research encompasses research that goes into other areas beyond fake news detection. We propose two major directions along these lines: fake news diffusion and fake news intervention. Fake news diffusion characterizes the diffusion paths and patterns of fake news on social media sites. Some early research has shown that true information and misinformation follow different patterns when propagating in online social networks



CONCLUSION

With the increasing popularity of social media, more and more people consume news from social media instead of traditional news media. However, social media has also been used to spread fake news, which has strong negative impacts on individual users and broader society. In this article, we explored the fake news problem by reviewing existing literature in two phases: characterization and detection. In the characterization phase, we 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. We also further discussed the datasets, evaluation metrics, and promising future directions in fake news detection research and expand the field to other applications

REFERENCES

- https://dl.acm.org/doi/abs/10.1145/3137597.3137600
- https://www.sciencedirect.com/science/article/pii/S1877050920300430
- https://ieeexplore.ieee.org/abstract/document/8777343
- **https://dl.acm.org/doi/abs/10.1145/3289600.3291382**