Fake News Detection in Data Science

1.Disha Saraswat 2.Dikshant Mathur

PCE19CS053 PCE19CS052

Computer Science and Engineering Computer Science and Engineering

Poornima College of Engineering, Poornima College of Engineering Jaipur, Rajasthan Jaipur,Rajasthan

[2019pcecsdisha53@poornima.org](mailto:2019pcecsdisha53@poornima.org) [2019pcecsdikshant52@poornima.org](mailto:2019pcecsdikshant52@poornima.org)

3. Hardik Singh 4. Abhishek Sharma

PCE19CS065 PCE19CS011

Computer Science and Engineering Computer Science and Engineering

Poornima College of Engineering, Poornima College of Engineering,

Jaipur, Rajasthan Jaipur, Rajasthan

[2019pcecshardik65@poornima.org](mailto:2019pcecshardik65@poornima.org) [2019pcecsabhishek11@poornima.org](mailto:2019pcecsabhishek11@poornima.org)

|  |  |
| --- | --- |
| **ABSTRACT**  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 enables the wide spread of “fake news”, i.e., low quality news with intentionally false information. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society. Therefore, fake news detection on social media has recently become an emerging research that is attracting tremendous attention. Fake news detection on social media presents unique characteristics and challenges that make existing detection algorithms from traditional news media ineffective or not applicable. First, fake news is intentionally written to mislead readers to believe false information, which makes it difficult and nontrivial to detect based on news content; therefore, we need to include auxiliary information, such as user social engagements on social media, to help make a determination. Second, exploiting this auxiliary information is challenging in and of itself as users’ social engagements with fake news produce data that is big, incomplete, unstructured, and noisy. Because the issue of fake news detection on social media is both challenging and relevant, we conducted this survey to further facilitate research on the problem. In this survey, we present a comprehensive review of detecting fake news on social media, including fake news characterizations on psychology and social theories, existing algorithms from a data mining perspective, evaluation metrics and representative datasets. We also discuss related research areas, open problems, and future research directions for fake news detection on social media. | **INTRODUCTION**  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 is 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 this 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 news “Pizzagate”3 by the end of the presidential election. Given the prevalence of this new phenomenon, the word “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 4 . 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. |

|  |  |
| --- | --- |
| **Literature Survey**  Network-based: Users form different networks on social media in terms of interests, topics, and relations. As mentioned before, fake news dissemination processes tend to form an echo chamber cycle, highlighting the value of extracting network-based features to represent these types of network patterns for fake news detection. Network-based features are extracted via constructing specific networks among the users who published related social media posts. Different types of networks can be constructed. The stance network can be built with nodes indicating all the tweets relevant to the news and the edge indicating the weights of similarity of stances [37; 75]. Another type of network is the cooccurrence network, which is built based on the user engagements by counting whether those users write posts relevant to the same news articles [69]. In addition, the friendship network indicates the following/followee structure of users who post related tweets [42]. An extension of this friendship network is the diffusion network, which tracks the trajectory | **Knowledge-based**  Since fake news attempts to spread false claims in news content, the most straightforward means of detecting it is to check the truthfulness of major claims in a news article to decide the news veracity. Knowledgebased approaches aim to use external sources to fact-check proposed claims in news content. The goal of fact-checking is to assign a truth value to a claim in a particular context [83]. Fact-checking has attracted increasing attention, and many efforts have been made to develop a feasible automated factchecking system. Existing fact-checking approaches can be categorized as expert-oriented, crowdsourcing-oriented, and computational-oriented. • Expert-oriented fact-checking heavily relies on human domain experts to investigate relevant data and documents to construct the verdicts of claim veracity, for example PolitiFact11, Snopes12, etc. However, expertoriented fact-checking is an intellectually demanding and time-consuming process, which limits the potential for high efficiency and scalability. • Crowdsourcing-oriented fact-checking exploits the “wisdom of crowd” to enable normal people to annotate news content; these annotations are then aggregated to produce an overall assessment of the news veracity. For example, Fiskkit13 allows users to discuss and annotate the accuracy of specific parts of a news ar |

**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**

[1] Sadia Afroz, Michael Brennan, and Rachel Greenstadt. Detecting hoaxes, frauds, and deception in writing style online. In ISSP’12. [2] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. Technical report, National Bureau of Economic Research, 2017. [3] Solomon E Asch and H Guetzkow. Effects of group pressure upon the modification and distortion of judgments. Groups, leadership, and men, pages 222–236, 1951. [4] Meital Balmas. When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism. Communication Research, 41(3):430–454, 2014. [5] Michele Banko, Michael J Cafarella, Stephen Soderland, Matthew Broadhead, and Oren Etzioni. Open information extraction from the web. In IJCAI’07. [6] Alessandro Bessi and Emilio Ferrara. Social bots distort the 2016 us presidential election online discussion. First Monday, 21(11), 2016.