Review of Online Fake News: Characterization, Detection and Mitigation.

Varchita B. Lalwani
Independent Researcher

Author Note

Natural Language Processing and Computer Vision Researcher.

Worked at a start-up for 4 months as AI/ML Engineer.

Attended Masters of Science by Research in NLP for 3.5 years under Dr. Manish Gupta (Director of Google Research, India) at International Institute of Information and Technology.

Abstract

Advancement in innovation gives us the opportunity to establish or exchange ideas, data, professional interests, and many articulations through virtual platforms and organisations. One such domain is online news but with great there comes awful too. From one perspective, online media is minimal expense, simple access, and quick dispersal of data lead people to look and consume news from social media. Then again, it empowers the widespread use of "fake news". The broad spread of false news affects people and society negatively. Therefore fake news detection is very much important and has recently become an emerging research. Many point to the 2016 U.S. presidential election campaign as having been influenced by fake news. Since then, "fake news" has been the dominant expression for online deception and disinformation. Fake news detection is technically challenging for some reasons. Reality and purpose of any assertion can't be evaluated by computers alone, so efforts should rely upon collaboration of people and innovation. Next, fake news is purposely composed to mislead readers to accept the false information, which makes it difficult and nontrivial to detect based on news content; in this way, we need to incorporate assistant data, such as user social engagements on social media, which is also challenging as users' social engagements with fake news produce data that is big, incomplete, unstructured and noisy. This literature review mainly focuses on characterizations, existing detection methods, mitigation strategies.

Keywords: Fake news detection, fake news characterization, fake news mitigation, social media, online fake news

Introduction

Fake news is not a new concept. It exists from the day the news circulated. Online social media platforms such as Twitter and Facebook can help users all around the world share real-time information. People are increasingly turning to social media for news rather than traditional news agencies. Consumption of online news is simple since it is less priced, easy to obtain, and easy to share, like, and comment on. Fake news is more popular and extensively disseminated on social media than legitimate news (Balmas, 2012). An article in News India Express said according to the poll, India is one of the most mobile-focused economies, with 73 percent of respondents using smartphones to get news. Overall, 82% of respondents get their news from the internet, including social media, and 63% get their news from social media platforms. It said WhatsApp and YouTube were the leading social media channels where 53% of people acquired their news, despite worries about widespread misinformation on these platforms. There is no agreed definition of fake news. Stanford University defines fake news: 'the news articles that are intentionally and verifiably false, and could mislead readers' (Allcott & Gentzkow, 2017). According to Wikipedia: 'Fake news is false or misleading information presented as news. It often has the aim of damaging the reputation of a person or entity, or making money through advertising revenue' (Wikipedia contributors, 2021). (Zhang & Ghorbani, 2020) defines 'fake news refers to all kinds of false stories or news that are mainly published and distributed on the internet, in order to purposely mislead, befool and lure readers for financial, political or other gains'. Since 2016, the term "fake news" has been used to describe internet misinformation and disinformation, both in journalistic and political debate, as well as scholarly research. Computer science and social sciences are the disciplinary fields with the largest number of documents published (*View of Four Years of Fake News: A Quantitative Analysis of the Scientific Literature | First Monday*, n.d.).

Due to the negative impact of fake news on people, detection is important. It's crucial to distinguish between the numerous types of false news, which include clickbait, hoaxes, propaganda, satire and parody, and others (Name-theft, framing, journalism deception), when spotting fake news (*Trends in Combating Fake News on Social Media a Survey*, n.d.), truth discovery, spammer and bot detection are also related to the problem of fake news detection (Shu et al., 2017). (*Trends in Combating Fake News on Social Media a Survey*, n.d.), (Shu et al., 2017) has a detailed summary of various sorts of fake news and areas related to it.

This literature review focuses on characterizations in terms of fundamental theories related to user and news, existing detection algorithms that are classified into practical based approaches and existing research based approaches, mitigation strategies such as removing malicious accounts and others based on existing information diffusion models. (D'Ulizia et al., 2021) Collecting reliable databases of fake and trustworthy news is not an easy undertaking. First and foremost, it causes news fact-checking so that articles can be marked as true or untrue. This procedure can be carried out in four ways: through human verification through expert-oriented fact-checking; through computational fact-checking; through crowd evaluation as in crowdsourced fact-checking; and through fact-checking assessment sites. Such a dataset is tough to come by as high-quality dataset for detecting fake news as the real-world online dataset is usually big, incomplete, unstructured, unlabeled, and noisy (Shu et al., 2017); and every day, social media generates a big amount of incorrect material with various purposes and linguistic characteristics. (D'Ulizia et al., 2021) gives detailed summary and analysis of 27 datasets with 11 features taken in mind.

Fake News Characterization

’ To comprehend the extension and assortment of online fake data, some significant viewpoints for characterizing the fake news are shown in Figure 1 (Zhang & Ghorbani, 2020). The core is the term ‘Fake news’ and has 4 major components: News Content, Social Context, Target Victims, Creator/Spreader. These deceive and exploit online social users by propagating false information and messages, thus jeopardising social belief and trust.

Creator/Spreader

The creator of online fake news can be non-humans and real human beings. Non-humans include cyborg and social bots, while actual humans include benign authors and users who publish fake news unintentionally and users who create fake content on purpose.

Non-human

Bots are computer algorithms designed to exhibit human-like behaviors, and automatically produce content and interact with humans on social media (Ferrara et al., 2016b). Bots are used to automate the tasks that humans do on their own such as, displaying and fetching information, making doctor appointments, and more. Many bots are made purposely to distribute rumors, misinformation, just noise etc (Ferrara et al., 2016b). Many bots are created to support U.S. election 2016, injecting thousands of tweets pointing to websites with fake news (Shu et al., 2017). Cyborg refers to human-assisted bots or bot-assisted humans (Chu et al., 2012).

Real-humans

Humans only do the development of bots and fake news. Bots just automate generation of fake news. Thus the ultimate creators of fake news are humans. It is very difficult to differentiate between fake and real information only by content and linguistic analysis. The following news: ‘FBI agent suspected in Hillary E-mail leaks found dead in apparent murder-suicide’ is completely false but is share don facebook over half a million times (Allcott & Gentzkow, 2017).

Target Victims

Victims are the prime targets of the online fake news. Users of online social media or other online news outlets could be among them. Students, voters, parents, senior citizens, and other groups may be targeted depending on the news's goals. Each piece of news comprises physical news content and non-physical news content. Voters and citizens may be the target users of false political claims; online customers may be the target users of false reviews and adverts; parents may be the target users of false educational information; and elderly citizens may be the target users of false health information.

Social Context

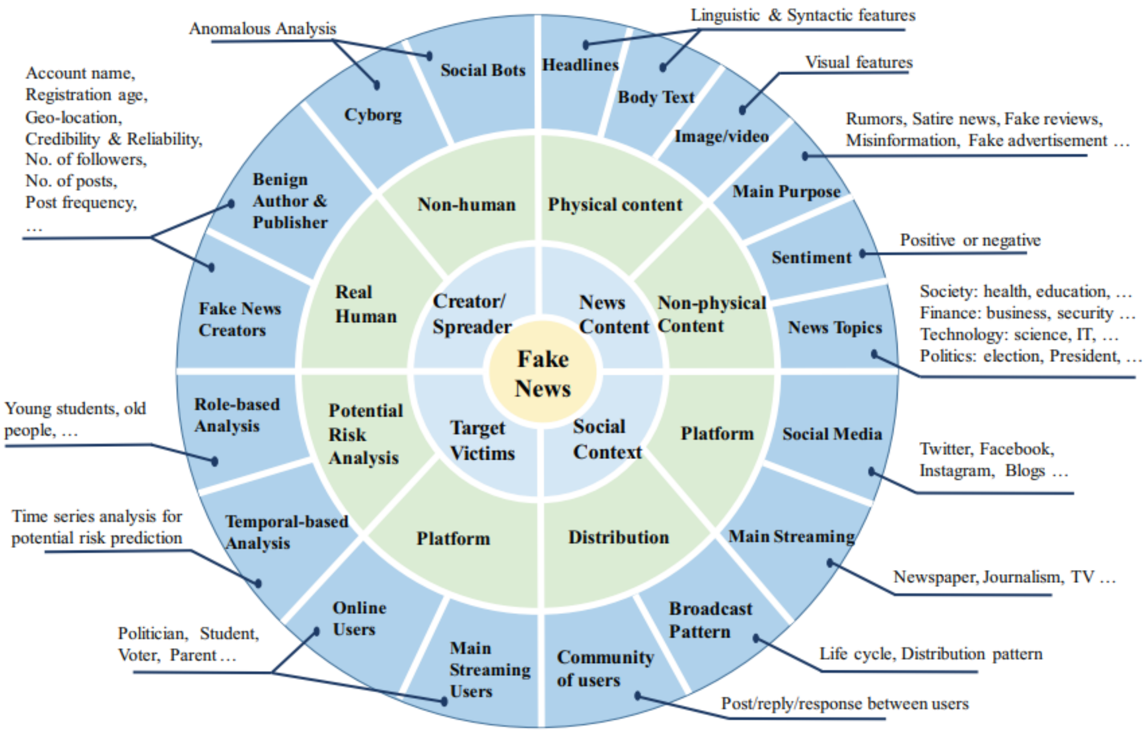
This refers to distribution of news through the internet i.e. the entire social environment and activity system. This includes how users are involved in the news and broadcast pattern analysis. Mainstream media platforms include TV, radio, newspaper. Online social media platforms include Twitter, Facebook, Instagram. Mainstream platforms are shifted to digital platforms such as The Times of India, Washington Post and so on, because of easy access and low cost of internet.

News Content

News content refers to what information it is carrying, also refers to the body of the news. It includes both physical (title, body text, multimedia) and nonphysical (purpose, sentiment, topics) content. Fake news also targets different domains (e.g. financial, IT etc.)

Figure 1

Fake news and everything related to it



Physical News

Facebook posts, tweets, Instagram posts become a powerful medium of information sharing in today's wide social network. Online social messaging are good representations of hot trending events since they happen in real time. URL of a webpage, emoji, image, video etc recognized as physical content of the news also called as carries and format of the news. Those components are key aspects for fake news detection since they have specific meanings and functions.

Non-physical

Non-physical contents are the thoughts, feelings, attitudes, and sentiments that news writers seek to express. Fake news can be fake reviews, fake advertisements, fake political news, and so on. Poor quality and biassed reviews are major problems for both online customers and brands. Fake reviews not only have the potential to influence decision-making, but they can also easily destroy a brand's reputation. Fake reviews and fake advertisements are both hazardous to the credibility of online e-commerce. The main kernel of fake news is non-physical content, which contains all the important ideas, feelings, and points of view that the authors want to convey to the readers. Another important feature for non-physical content for fake news is semantic polarity.

Fundamental Theories

There are many hypotheses about how and why people react to certain types of news. These theories of human cognition and behaviour developed across a variety of fields, including social sciences and economics, provide vital insights for analysing fake news. The hypotheses revolve around the news or its disseminators. (Zhou & Zafarani, 2020) lists the foundational theories in the social sciences and economics. There are some theories about news that claim fake news differs from the truth in aspects of, for example, writing style and quality, quantity such as word counts (according to the information manipulation theory), and conveyed attitudes. It's worth noting that these forensic psychology theories focus on false remarks or testimonies (i.e., misinformation), not fake news, despite the similarities in principles.

Psychological Foundations Of Fake News

Consumers are the primary targets of fake news, which exploits their specific vulnerabilities. Because of two fundamental criteria, consumers are naturally prone to bogus news: I Naive Realism: consumers believe that their perceptions of reality are the only ones that are right, and that anyone who disagrees is uninformed, irrational, or biassed. (ii)

Confirmation Bias: Customers prefer information that validates their present opinions . Because of cognitive biases that are inherent in human nature, consumers commonly confuse fake news with real news. Once a misunderstanding has been established, it is extremely difficult to change it. Correcting erroneous information (e.g., fake news) with the presenting of accurate, factual information, according to psychological studies, is not only useless in diminishing misperceptions, but may significantly enhance them, especially among ideological groups.

Social Foundations Of Fake News Ecosystem

Decision making, according to prospect theory, is a process in which people make decisions based on the relative rewards and losses compared to their current circumstances. This motivation to maximise the reward of a decision also extends to social benefits, such as continuous acceptance by others in a user's immediate social network. This preference for social acceptance and affirmation, as described by social identity theory and normative influence theory, is critical to a person's identity and self-esteem, leading users to choose "socially safe" options when consuming and disseminating news information, adhering to community norms even if the news is fake.

Echo Chamber Effect: Consumers are selectively exposed to certain sorts of news as a result of the recommendation and the way feeds show on people's social media homepages, aggravating the psychological hurdles of recognising fake news. Users like to join groups with like-minded people on social media, polarising their opinions and producing an echo chamber effect. Because the echo chamber effect makes it simpler for people to consume and believe bogus news. (1) social credibility, which indicates that people are more likely to believe a source is credible if others believe it is credible, especially when there is insufficient information to determine the source's honesty.; and (2) frequency heuristic, consumers may naturally prefer information they hear frequently, even if it is fake news. Increased exposure to an idea is enough to produce a favourable perception of it.

Formulating Fake News As Game Theory

In order to describe this rational theory of false news exchanges from an economic game theoretical perspective, the news generation and consumption cycle can be modelled as a two-player strategic game. Assume that the information ecosystem has two types of players: publisher and consumer. The mapping of the source signal ‘s' to the output news ‘b' with distortion signal ‘a', b = [-1, 0, 1] demonstrating [left, no, right] bias effects on ‘s' is the procedure for publishing news. Two opinions on the value of a publisher: (i) short-term utility: the desire to maximise profit, which is inversely proportional to the number of customers reached.; (ii) Confirmation bias and prospect theory are examples of psychology utility: obtaining news that satisfies their prior ideas and social demands. In this strategy game of news consumption, both the publisher and the consumer aim to maximise their overall utility. We can capture the fact that fake news occurs when a publisher's total utility is dominated by short-term utility, and the consumer's overall utility is dominated by psychology utility, and an equilibrium is maintained. (Zhang & Ghorbani, 2020); (Zhou & Zafarani, 2020) shows detailed understanding of different theories.

Fake News Detection

Fake news detection is classified into two approaches(Zhang & Ghorbani, 2020) (i) the practical-based detection approaches, from the perspective of Internet users (ii) the existing research-based detection methods, from the perspective of academia and research.

Practical-based Detection Approaches

Online Fact Checking Resources

FackCheck.org is a non-profit “consumer advocate” website for voters that strives to decrease political dishonesty and ambiguity in the United States (Factcheck). They assess the veracity of assertions or remarks made by important political figures in the United States. They eventually analyse and report the veracity of each piece of information with the help of reputable experts. FactCheck.org also has various other components, such as SciCheck for fact-checking science-based assertions, Health Watch for fact-checking health-care debates, and the Facebook Initiative for refuting fake news on Facebook.

Factmata.com is a Google-funded project that uses statistical fact-checking and claim recognition to identify false claims (Factmata). The claim checking work is fully dependent on artificial intelligence and machine learning algorithms, which is the project's most notable characteristic. This platform's goal is to detect and check disinformation in order to provide a more informed perspective on the world.

Hoaxy.iuni.iu.edu is a platform for gathering, detecting, and analysing online misinformation, as well as related fact-checking initiatives. An interactive graphic depicts a preliminary examination of a sample of public tweets including both fake news and fact-checking information on this site. They can also see the networks and actions of fact-checkers and fake news spreaders.

PolitiFact.com is a website that rates the accuracy of claims or assertions made by politicians, pundits, columnists, bloggers, political analysts, and other members of the media in the United States. PolitiFact.com is an online fact-checking system for political news and information that is independent and nonpartisan (Politifact). True, Mostly True, Half True, Mostly False, False, and Pants on Fire are the judgments they employ to rate the truthfulness of a proposition.

Other comparable websites include OpenSecrets.org, OpenSources, FakeNewsWatch.com, FakeSpot.com, ReviewMeta.com, and more. Overall, fully automated false news identification is still difficult to achieve and has a long way to go. (Zhang & Ghorbani, 2020) compares existing fact-checking resources in terms of "Topic coverage," "Source of fake news," "Rating levels," "Dashboard & visualisation," "API," "Detection technology," and "Others."

Social Practical Guide For Fake News Detection

The present fact-checking resources have some flaws and limitations, such as the detection process being time-consuming, the results always being delayed, and a significant amount of manual labour being required. Practical social counsel, like the definition of false news, can be defined as a creator-based approach, new content-based approach, and social context-based approach.

Creator-based Approach

More and more studies believe that news sources are the ideal place to look for fake news. For instance, what if the news source is from a well-known or unknown online domain? It is also feasible to detect a malicious website solely by inspecting the

URL's lexical properties, such as strange domain names (e.g., ".com.co") or suspicious tokens. The "About Us" or "Disclaimer" section of a webpage can sometimes provide relevant information about the page and can be used as a credibility indicator.

News Content Based Approach

Much research has demonstrated that Internet users have poor ability to distinguish between real and incorrect information. Here are some practical sociological theories that can assist users in identifying potentially harmful news material.

Do Not Stop At Headlines. To attract more clicks and attention, fake news headlines are always sensational and eye-catching. When defining any suspicious web material, reading the entire piece rather than just the headline is a solid method.

Check The Supporting Resources. To persuade readers, news writers constantly provide a plethora of information, such as specialist knowledge, statistics or survey data, and so on. Take the time to examine the accompanying materials, as they may assist Internet users in determining the accuracy of the news information.

Check The Sentiment And Sensitive Topics Of The News. The majority of fake news plays on the worries, anxiety, pity, curiosity, and other emotions that readers have. Users should be accountable for identifying the sensitive sentiment level of the news, such as if it makes you furious or sad, before following the feelings and viewpoints expressed in the news. Also, internet users must identify some sensitive themes and constantly ask themselves if the news is too humorous or intriguing to be true; if the news contains any depressing stories; if the news foreshadows a future tragedy such as an earthquake or an epidemic sickness, and so on.

Social Context Based Approach

Another useful method is to record the sceptical social context of online news. Verify if the news from the same source is believable or not; check the date of the news; check if any other online news platforms are reporting the same or similar stories, and so on.

Research Based Approached

We'll go through and discuss the most recent research on spotting fake news. The three types of models for research-based approaches are component-based category, data mining-based category, and implement-based category.

Component Based Category

There are three types of fake news detection methods: creator and user analysis, news content analysis, and social context analysis.

Creator And User Analysis. Malicious social media accounts have distinct traits that set them apart from real users.

User Profiling Analysis. This includes the language spoken by the account, the account's geographic locations, if the account is verified or not, how many posts/tweets the account has, and so on (Ferrara et al., 2016b). The user profiling analysis describes how active and suspicious a social account is, and it has been shown to be successful in spotting suspicious social accounts.

Temporal and posting behavior analysis. Signal resemblance to a Poisson process, average time between two successive posts, reaction frequency, and so on are all examples of temporal behaviour. Due to timers or automatic mechanisms, suspicious accounts like social bots and cyborgs are more active during specified periods .

Genuine human users, on the other hand, display a wide range of temporal patterns (Chu et al., 2012). It's probable that the great volume of responses and mentions are connected. This indicates how suspicious a social account is.

Credibility-related information. The number of friends and followers can also be utilised to tell the difference between fake and real accounts. Social bots have significantly more friends than followers.

Sentiment-related analysis. Malicious accounts might exaggerate facts and mislead legitimate users by eliciting unusual emotional responses. Sentiment analysis is a valuable tool for explaining how online social media conveys emotion, attitude, and opinion. Psychological keyword analysis is a popular method of determining the original author's mood and feeling. Arousal-valence dominance score, happiness score, emotion score and polarisation and strength score are some of the ways offered.

News Content analysis. Using in-depth news content analysis, we can assess language patterns and writing styles for both true and false news, and then capture the most discriminative components for online fake news detection.

Linguistic-based analysis. The goal of linguistic analysis is to match the news creator's language ability by looking at language formats and finding writing trends. "Bag-of-words" and "n-grams" are the most used approaches for representing raw news texts. Other techniques for natural language representation and document classification have been proposed and used, such as deep learning techniques, embeddings, LSTMs, and so on.

Semantic-based analysis. Fake news writers frequently use exaggerated titles to attract readers' attention. True news, on the other hand, should have a title that corresponds to the article's content. The dishonest reviewers have no knowledge of how the things work or what services they provide. As a result, semantic-based analysis can provide important clues for assessing the level of suspicion in online news. Combining the "n-gram" model with the "deep syntax" model, this can establish the degree of compatibility and consistency between the news creator's personal experience and the news content. Researchers can utilise the combined information from news creator analysis and semantic based analysis to assess the compatibility between the user's background and the news content, which is a significant improvement.

Knowledge-based analysis. Knowledge-based analysis refers to attempts to directly check the truthfulness of the major statements in a news report (Shu et al., 2017). Despite the recent development of various creative ways for automatically detecting fake news, today's fact-checking tasks still rely significantly on human knowledge. Second, we can create supervised machine learning models for discriminating fake from real news by approaching fake news detection as a binary classification problem.

Style-based analysis. Social media is used by legitimate internet users to communicate their thoughts, emotions, and feelings about certain things, events, and services . On the other hand, malicious online identities hide deceptive information by obfuscating their writing style or attempting to imitate other users. By striving to capture the distinctive aspects of writing styles of both accounts, style-based analysis plays an important role in spotting online false news. (1) Physical style analysis is the practise of identifying influential physical clues to distinguish fake news from authentic news. These features, such as the frequency of verbs and nouns, emotion words, and informal terminology, might disclose the news writer's writing style, text grammar, and personal attitude. The presence of suspicious tokens in social communication data is also a useful feature for determining authorship and analysing writing style. (2) Non-physical style looks at the text's complexity and readability. fake news creators take longer to write and make more errors. The keys "backspace" and "delete," for example, are regularly used by false news creators when they want to publish erroneous statements. The author's credibility is crucial to the credibility of a piece of news or a document.

Social context analysis. The study of how rapidly and extensively social data is distributed, as well as how online users interact with one another, is known as social context analysis.

User network analysis. The number of internet users who interact with the news creator on a regular basis can be used to gauge the news's accuracy. If a lot of unusual or untrustworthy accounts "like" or "comment" on a piece of news, it's more probable that the content is erroneous and misleading. Furthermore, the reliability of the news creator's social network may be a reliable indicator of the news's authenticity

Distribution pattern analysis. The study of user network analysis uncovers how internet users interact, whereas distribution pattern analysis investigates how information is distributed. Anomaly pattern identification for social data is a hot topic among researchers, as it can help internet users locate data that is unusual, strange, or unexpected (Chu et al., 2012). Detecting questionable distribution patterns for fake news is challenging and difficult due to the various and dynamic nature of online social behaviours. As a result, advanced visualisation systems are always paired with traditional machine learning methods to address the aforementioned challenges.

Implement Based Category

Fake news detection can be classified as real-time detection or offline detection, depending on how the system is implemented.

Offline Detection. Batch-sized machine learning models are widely employed to detect bogus news. They're useful for classifying online fake news since they can investigate anomalous information in a descriptive manner, such as identifying the most influential qualities for differentiating false information amid large amounts of social communications. Based on the types of online information, offline classification can be divided into fake review detection, satire news detection, hoaxes detection, and political news detection.

Real-time Detection. Real-time analysis approaches are used to determine whether current social information is false or not. By using predictive analytics approaches on real-time data, real-time analysis can improve the applicability of offline methods and add practical value to online false news prediction. A mechanism for assessing false information on Twitter in real time and internet architecture for tracking real-time disinformation and the fact-checking records that go with it are developed.

Data Mining Based Category

Categorized into supervised and unsupervised techniques.

Supervised Learning. For online hoaxes, frauds, and deceptive information classification, supervised machine learning algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), Logistic Regression, and K-nearest Neighbor have been extensively used in previous literatures (A detailed list of papers related to models is in (Zhang & Ghorbani, 2020)). The majority of existing algorithms treat the challenge of detecting fake news as a classification problem, in which the goal is to predict whether a news storey is fake or not. True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), Precision (Pr), Recall (Re), False Positive Rate (FPR), False Negative Rate (FNR), F-score, and Accuracy (Acc) are some of the most commonly used metrics (Shu et al., 2017)

Unsupervised Learning. A supervised learning model's performance is heavily influenced by the quality of a tagged dataset. As previously stated, obtaining the data's ground truth label is difficult. As a result, an unsupervised learning model is more practical and realistic for handling real-world problems. The majority of the research focuses on sentiment analysis or semantic similarity analysis. For online false reviews, unsupervised technique is based on similarity measurement. By combining word similarity and word-order similarity, they present a method for identifying near-duplicate online evaluations with high accuracy. Other techniques in unsupervised sentiment is the analysis framework for social media photos. Using the correlations between visual information and important contextual information, their method can evaluate the sentiment of social pictures from two large-scale datasets. An unsupervised generative Bayesian model is also studied to automatically discriminate between honest and bogus reviews.

Mitigation

Let us look at a few interventions targeted at reversing the effect of fake news by carefully adding accurate news to social media platforms, allowing people to get to the truth and reducing the impact of fake news on user opinion.The purpose of utilising users' flagging activity on Facebook in collaboration with fact-checking organisations is to minimise the spread of misinformation by leveraging users' flagging behaviour. (Sharma et al., 2019); (Pierri & Ceri, 2019) present the understanding and mitigation strategies

The Independent Cascade (IC) and Linear Threshold models as well as point process models like the Hawkes Process model are all well-known models for representing information dissemination in social networks. Each edge in the IC model is represented by a parameter $p_{u,v}$ which indicates the strength of user u's effect on user v. The diffusion process begins with the activation of a set of seed nodes in the first timestep. A node u activated at time step t makes a single activation attempt on each inactive neighbour v at each time step of the diffusion process. The activation is successful with probability p, and once activated, a node remains active during the diffusion process. The edge parameters in the LT model reflect a weight of influence, and each user has their own (uniformly and independently) random threshold parameter, and the sum of incoming edge weights should be less than one. As a result, a user is activated when the weights of all its active neighbours reach its threshold at a certain time step in the diffusion process.

Decontamination

The IC or LT model is used to simulate the diffusion process. A greedy method is used to identify the optimal selection of seed users from whom to begin the diffusion process for real news, with the goal of decontaminating at least a β-fraction of the users. The technique is to select the next best user to add to the seed set iteratively based on the marginal advantages received by including the user (i.e. the number of users that will be activated or reached by the true news in expectation, if the seed set did additionally include the chosen user). Iterative seed user selection continues until the goal of decontaminating a β-fraction of the infected users in expectation is met.

Limitation

It is being advocated as a response to the spread of fake news. Also the quantity of seeds is not fixed, and if the damage is significant, the expense of restoring the actual news cascade by activating or incentivizing more seed users may be prohibitive.

Competing Cascades

A real news cascade is provided to compete with the fake news cascade as the fake news emerges and begins to spread through the network, rather than after it has spread. An influence blocking maximisation objective as determining the best strategy for disseminating true news in the face of a misinformation cascade by strategically selecting k seed users, with the goal of minimising the number of users who are activated by the fake news campaign instead of the true news campaign at the end of the diffusion. The model implies that once a user has been active by either the false or actual cascade, that user will remain activated by that cascade. Each edge has its own set of diffusion model parameters, one for real news and one for fraudulent news. It's reasonable to study the two forms of diffusion processes independently because the intensity of influence under factual news and sharing fraudulent news may differ.

Limitation

The true news cascade's seed users are chosen once at the start of the process in reaction to a detected false news cascade, after which both cascades propagate independently without external moderation. Second, because the model does not distinguish between exposure and re-sharing, a user being "activated" represents both exposure and re-sharing, as it implies that an active user is exposed and always tries to activate its neighbours through forced re-sharing.

Multi-stage Intervention

The method is a multi-stage intervention strategy to allow external interventions to adjust as needed to the observed spread patterns of fake news, in which previous news sharing activities trigger future news sharing events based on the intensity of effect and the time delay between events.

Each user i (news sharing) events are triggered by their base intensity and past (news sharing) events with some decay function that depends on the time gap between the current time and the past event, as well as the past events of other users j, whose influence is captured by i,j in multivariate point processes.

Limitation

The strategy implies that bogus news has already been recognised and is being tracked across the network, which is a difficult task. Furthermore, it is plausible that, once fake news has been found, we should simply provide direct recommendations of factual news to users who have been exposed to the fake news, and remove the fake news content from the platform to avoid future spread.

Conclusion

In this study, we looked into the subject of false news by looking at current literature in two phases: characterization, detection, and mitigation. During the characterisation phase, we looked at the components of fake news and hypotheses that help us understand how and why users react to it. We looked at existing research-based strategies as well as practical ways throughout the detection phase. After that, the Independent Cascade and Linear Threshold, and Hawkes Process models were utilised to look at mitigation options including decontamination and competing cascades. We also looked at features for detecting fake news, which were classified into three categories: creator/user-based features, news content-based features, and social context-based features.

References

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Balmas, M. (2012). When Fake News Becomes Real. *Communication Research*, *41*(3), 430–454. https://doi.org/10.1177/0093650212453600

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg? *IEEE Transactions on Dependable and Secure Computing*, *9*(6), 811–824. https://doi.org/10.1109/tdsc.2012.75

D'Ulizia, A., Caschera, M. C., Ferri, F., & Grifoni, P. (2021). Fake news detection: a survey of evaluation datasets. *PeerJ Computer Science*, *7*, e518. https://doi.org/10.7717/peerj-cs.518

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016b). The rise of social bots. *Communications of the ACM*, *59*(7), 96–104. https://doi.org/10.1145/2818717

Pierri, F., & Ceri, S. (2019). False News On Social Media. *ACM SIGMOD Record*, *48*(2), 18–27. https://doi.org/10.1145/3377330.3377334

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating Fake News. *ACM Transactions on Intelligent Systems and Technology*, *10*(3), 1–42. https://doi.org/10.1145/3305260

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22–36. Fake News Detection on Social Media: A Data Mining Perspective: ACM SIGKDD

Explorations Newsletter: Vol 19, No 1

Trends in combating fake news on social media â“ a survey. (n.d.). Taylor & Francis. Retrieved September 3, 2021, from https://www.tandfonline.com/doi/full/10.1080/24751839.2020.1847379

View of Four years of fake news: A quantitative analysis of the scientific literature | First Monday. (n.d.). firstmonday.Org. Retrieved September 3, 2021, from https://firstmonday.org/ojs/index.php/fm/article/view/11645/10152

Wikipedia contributors. (2021, August 29). *Fake news*. Wikipedia. https://en.wikipedia.org/wiki/Fake_news

Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, *57*(2), 102025. https://doi.org/10.1016/j.ipm.2019.03.004

Zhou, X., & Zafarani, R. (2020). A Survey of Fake News. *ACM Computing Surveys*, *53*(5), 1–40. https://doi.org/10.1145/3395046