SOSC 4300 Literature Review

Fake news detection using computational methods

Group members: Hui Ka Ming (20506919); Lam Wing Yi (20510178)

Law Po Yi (20510673); Lo Wing Ching (20521892)

Contributions of group members

1. Introduction	Lo Wing Ching
2. Traditional Methods/Data/Limitation	Hui Ka Ming
3. Computational Methods/Data (3.1) Content Analysis	Lam Wing Yi
Computational Methods/Data (3.2) Account analysis	Lo Wing Ching
Computational Methods/Data (3.3) Crowed-sourcing	Law Po Yi
Computational Methods/Data (3.4) Mixed method	Hui Ka Ming
4. Conclusion	Hui Ka Ming

1. Introduction

1.1 Background

“Fake news” is referring to the untrue and fabricated information that followed the formats of news media content but not authoritative or with malicious intentions (Lazer et al., 2018). Social media acted as a prominent source of information and news consumption. Yet, fake news has become pervasive with the growth of social media platforms, especially on Twitter and Facebook News Feed. The rise of social media enable sharing among a large group of people online and replicating messages rapidly, in which it accelerated the speed of spreading misinformation. The spread of fake news is rapid and even more faster than that of real news (Vosoughi et al., 2018). While fake news is not a currently emerged issues but became prevalent in the last few years, people investigated a lot of methods to distinguish the fake news. In the past, traditional methods like manual checking was used to identify fake news. Yet, traditional methods have many limitations and with low accuracy to examine the falsity of information. As a result, computational methods and data such as machine learning will be used to detect fake news nowadays.

1.2 Importance of this topic

Fake news is competing with real news, lowering the credibility and trust of actual information. In addition, consuming fake news can alter individuals’ beliefs and affect their behaviours. Public opinions will be easily manipulated by the disinformation, causing disruption on the public fairness and rationality (Nguyen et al., 2020). It happened a lot in real world, notably during critical events such as political elections. For example, the fake news received higher sharing than the news from major legitimate media on Facebook in the 2016 United States presidential election, in which altered the decisions of electorates and outcome of the election (Allcott & Gentzkow, 2017) . Many research has revealed that over time exposure to various types of fake news might lower individuals' resistance against fallacious messages, thus being more impressionable.

Therefore, it is important to stop the spreading of fake news from misleading people. Apart from the major fact-checking organizations such as FactCheck, Politifact, and Snopes combating fake news, many researchers have also devoted efforts in limiting the disperse of fake news by identifying falsity.

1.3 Overview

This paper summarized several studies in which adopted computational methods and data to detect fake news online. Traditional methods and data used for detecting fake news were examined, followed by an introduction to three major computational methods for fake news detection, including (1) content analysis, (2) account analysis, and (3) crowd-souring method. Previous studies related to fake news identification using these methods were also explained. Lastly, the advantages and shortcomings of each computational methods were discussed.

2. Traditional Methods/Data/ Limitations

Manual fact checking requires a substantial amount of time and human effort. Despite devotion in identifying disinformation, online data are being generated at a tremendous speed, solely with manual fact checking method, information pollution could not be well-contained as the magnitude of data is too big. Besides, the manual analysis method would be costly to scale given the amount of data. Machines would be needed to aid the process, either by automating the process or acquire human-in-the-loop methods.

3. Computational Methods

3.1 Content Analysis

With the explosive growth of fake news in the social media platforms, people nowadays emphasize the importance of fact-checking. Content analysis is one of the widely-used methods to detect fake news on social media and blogs. It is a discourse-dependent approach which evaluates the truthfulness of an article's content (Figueira & Guimarães & Torgo, 2018). The validity of the news content is determined by four aspects: (1) the false information it contained, (2) its writing style, (3) its propagation patterns and (4) the credibility of the sources where it gets the information from (Zhou & Zafarani, 2020). Zhou and Zafarani defined fake news into two categories: (1) false news, which means the news media mislead the public by posting unreal statements and figures. They are often related to public figures and organizations. (2) Another type of fake news is deceptive news, which can be more harmful and difficult to identify. It focuses on the intention and news authenticity. This kind of fake news is often intentionally published by the news media to address the public's perceptions. However, news or articles that indicate their disagreement on others' point of view or interests are not perceived as false news.

Methodology

In the study of Zhou and Zafarani, they applied automatic fact-checking technique instead of manual fact-checking which is time-consuming and too slow in progress along with the newly posted information. Before creating an automatic detective machine, they divided the detection of fake news into four perspectives: (1) Knowledge-based method, which checks whether the text of the news is consistent with fact; (2) Style-based method, which is based on whether there are extreme tone or emotions behind the content; (3) Propagation-based method, which is depends on the way that the news spread online and; (4) Source-based method, which examines the trustworthiness of the sources where the news get information from. After setting these criteria, they grouped fake news models by the machine-learning methods based on the aforementioned four perspectives of fake news detection to create a new fake news database with the contribution of existing dataset for automatic fake news detection.

The fact-checking system separates the procedure into two stages. The first stage is Fact Extraction, which classifies whether the source of the news relies on is reliable, valid and complete. For example, if the knowledge of a news article is extracted from Wikipedia, which is a relatively low-credibility website providing raw facts is comparatively unreliable and may lead to incomplete knowledge. The second stage is the Fact-checking process, which is to access the authenticity of a news article. It compares the knowledge from a news article with the true knowledge to check if they are consistent.

Evaluation

Knowledge-based method

To identify whether the information from the news article complies with the truth knowledge is the most direct way. However, fact-checking based on the knowledge-based method largely relies on external resources to determine true and false claims for a computational-oriented approach. The overly-relying on external sources for automatic fact-checking machine may create another problem of reliability (Shu et al., 2017).

Style-based fake news detection

The accuracy is high since the authors write in a special and attractive style to motivate readers to continue reading and convince them to trust. The use of content analysis by machine learning systems to identify the factuality of a news spread on social media has been examined by Potthast et al. (2017). They analyze the writing style of news from extremely biased in favor of a particular political party and the result showed that the models with an accuracy of 78% in identifying extremely biased news.

However, only relying on a style-based method might be problematic and unreliable since the performance of it relies on how well the style of news content (text and images) can be captured and represented (Zhou & Zafarani, 2020).

Source-based methods

Although source-based methods can be an easiest and obvious method to check the factuality of a news article, there is a limitation in the method relying on the source of a news article extracted from that we cannot assume every document from an non-credible website is untrue and unreliable. Przybyla (2020) clarified that a non-credible website may contain true information, similarly, news articles sourced from credible news media may not be perfectly accurate.

3.2 Account analysis

Fake news are usually spread by spammer accounts or bot in which the accounts are registered as running online activities automatically. The false identities in online social platforms mimic humans and aim to spread propaganda or perform crimes online (Hakimi et al., 2019). To detect fake news, one of the most important steps is to look at the sources of the news or information (Baly et al., 2018). Hence, the credibility of social media accounts and its events has been brought to attention. Account analysis focus on identifying the credibility of social media accounts in order to diagnose whether the social media accounts are spam accounts (Alvaro et al., 2018). To perform account analysis, data can be extracted through (1) API based approach, (2) artificial data generation, (3) bot-crawled, and by (4) existing dataset study (Hakimi et al., 2019). In most studies, Twitter and Facebook accounts and its events were collected as data to detect fake accounts by using account analysis.

However, since the real-world data of Facebook API has authorisation restrictions on users, obtaining data from Facebook API is difficult without permission, as well as API-based approach and bot-crawler approach As a result, Hakimi et al.'s (2019) study used artificial data generation approach, fabricated sample data by referring to the network structure and features of existing data to conduct analysis. Prediction models of K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Neural Network (NN) algorithms were constructed using the annotated datasets for classifying true users and fake users.

In Castillo et al.'s (2011) study, they set up datasets of tweets about some of the trending themes. The datasets respecting the credibility of tweets were construed one by one using crowd sourcing approach and decision tree model was also used on four groups of features, including the messages, users, topics, and propagations (Castillo et al., 2011). Benevenuto et al. (2010) created a model that conduct manual annotated datasets for detecting spammers accounts. The datasets contains 1,000 spammer and non-spammer accounts' records. This method can detect 70% of spammer accounts and 96% of non-spammer accounts accurately (Benevenuto et al., 2010).

Chu et al. (2012) has did a similar study for detecting bot accounts. In their study, accounts were categorized into three distinct groups, including human accounts, bots accounts, and cyborgs accounts. With human-labelled dataset of 6,000 users collected, they constructed a system to analyse the accounts in four different fields such as measurement of entropy, detection of spammer account, analysis of account properties, and decision-making. Accuracy was used to evaluated the effectiveness of the system. Another recent study done by Dickerson et al. (2014) analysed the datasets from 2014 Indian election. They proposed a method that separate the accounts into human accounts and bots accounts by identifying special features of the account, such as network, linguistic and application-oriented features.

The models built with big data in account analysis can attain high accuracy levels in most studies. In Benevenuto et al.'s (2010) study, their method can detect 70% of spammer accounts and 96% of non-spammer accounts accurately. For Chu et al.'s (2012) study, their system attained 96% in the human class. In Perex-Rosas et al.'s (2017) study, they built a fake news datasets using crowd-sourced and their dataset attained 78% accuracy level in the best model.

3.3 Crowded-sourced

'Crowd-sourced' is an alternative type of fact-checking method. The linkage between flagging systems and crowd-sourcing is one of the current major methods in detecting fake news. Users act as a fact-checker rather than professional fact-checkers to assess the reliability of the news through a flagging system (Coscia & Rossi, 2020). There are several studies which aimed to investigate the relation between fakes news and crowded misinformation flagging. Studies have shown that news consumption patterns vary across the political spectrum, which means people tend to believe the misinformation that is consistent with their political ideology. (Pennycook & Rand, 2019) Therefore, most cases of investigating users flagging activity conducted by the politically ideological alignment. To identify the reliability of the news, we have to understand the relation between main agents (1) news sources (2) users ,and the power of the crowd (The users flagging's activity).

Measurement of main agents

In the study of Pennycook & Rand, the news can be categorized as (1) mainstream media outlets, (2) hyperpartisan website (3)websites that produce blatantly false content. In 2020, Coscia & Rossi has classified the value of news as popularity, polarity and truthfulness. They mentioned that truthfulness is correlated with the news source's polarity, which means more polarized sources tend to be less truthful. Unlike news sources, users embed in social networks that could be involved in three activities, including reshare, flag as well as consumption, corresponding to their political stance.

Methodologies

In the study of Coscia & Rossi (2020), it created the Bipolar model to replace the Monopolar model. They had run the models for 50 times and got the average value. The structure of the Bipolar model includes a user-source network and social network. Each of them were given polarity and popularity in which picking the required number of individuals with polarity values regarding the sources polarity. Also, LFR benchmark will generate the social network that shares nodes. In short, the Bipolar model can demonstrate how polarization affects the flagging system, more popular items are shared more and flagged more, even adding the factors of source popularity, reputation. While the traditional monopolar model analyses users activity without considering the user's polarisation, the approximation is not a realistic representation of reality.

Apart from the Coscia & Rossi (2020), Sebastian et.al has elaborated more on investigating polarization by performing Bayesian inference to analyse the user flaging activity, regarding user abstaining from flagging activity, user's accuracy in flagging the news and user's observed activity. The data were collected from Facebook users and normalized in which averaged over 5 runs. There are few areas they focused on (1) social network graph and new generation (2) news spreading (3)users' parameters (4)Algorithms. In short, this research shows user's flagging behaviour can be observed through detective algorithms. (Tschiatschek, Singla, Rodriguez, Merchant, & Krause, 2018)

Evaluation

Both of the studies have drawbacks. Relying on the social media data to identify the effectiveness of flagging systems on detecting fake news, the soundness and reliability of the result has been doubted. The diversity of the sources might be neglected, for example, Coscia & Rossi's sources are neutral and there are fewer polarized sources, resulting in lacking extremists data. Replacing the traditional monopolar model by the Bipolar model is a better assumption on people's flagging behavior, including flag untruthful news and also truthful news items, in which the traditional methods might not cover. The judgement from this model is often subjective and involves pre-existing polarization of both sources and users. At the same time, the Bipolar model ignores the role of originality or of spreaders' effort in making content go viral, reward and cost function for both users or news sources are neglected.

On the other hand, Sebastian et.al's study has focused on the user's activity while the relation between source and users has been neglected, in particular how the bias of the source affects the user's flagging activity. However, this shed the light on the importance of observing the users' behaviour before analysing the relation between sources and users that will improve the accuracy and effectiveness.

3.4 Mixed method

Given flaws of each and individual method, scholars have been attempting to combine selected methods into different solutions.

Nakov, Nguyen, Kan and Sugitama (2020) proposed Factual News Graph(FANG) a novel disinformation detection framework which can be implemented with Graph Neural Network(GNN). They highlighted that FANG takes social context into account, heterogeneous relationships between different social entities (i.e. users, news sources) are captured. As a result, the complexity of the model based on FANG could be translated into explainability.

The framework incorporates various methods of fake news detection including account analysis, identifying reliable news sources, network-based method, that captures relationships among heterogeneous social entities and linguistic (content or textual) analysis, that analyze sentiments and stances as well as news texts. The attempt to capture meaningful ties between users, posts and news sources means models based on FANG would be more complex, however, thanks to GNN, complex graph models can be implemented while retaining explainability. As opposed to typical models that learn for the best approximation of predicting fake news based on features of users, posts and news sources, FANG models implemented with GNN does not make predictions solely based on specific sets of attributes of social entities, instead, the models learn the contexts of a piece of news or information, contributing to not only more accurate but more explainable results.

The paper summarized with results from a FANG based GNN model, the model yielded better overall results, outperformed other models, including textual-only model, network model that does not take the social context into account, without the need of substantial amount of training data. However, the framework seems to require high quality, well-preprocessed data, the application of such framework might not be applicable outside of well-structured social media platforms. The model based on FANG seems to require high data quality as the framework has various components, i.e. types of nodes and edges. The availability of data remains questionable while the data preparation process might be difficult.

4. Conclusion

Detection of disinformation is an open end problem, scholars have researched many approaches, numerous proven effective. It is obvious the development of this area relies heavily on the advancement of advancement in computer science. Especially for recent works, the advancement in the field of machine learning has contributed to novel efficient, scalable solutions, Natural language Processing enables accurate analysis of textual traits of disinformation, Graph Neural Network opens the door for innovative methods. With availability of larger datasets and breakthroughs in deep learning, it seems that mixed machine learning methods would remain the focus of disinformation detection research.

References

Andrews, E. (2019). How fake news spread like a real virus. Retrieved from https://engineering.stanford.edu/magazine/article/how-fake-news-spreads-real-virus

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives. 31. 211-236. 10.1257/jep.31.2.211.

Baly, R., Karadzhov, G., Dimitar, A., Glass, J., & Nakov, P. (2018). Predicting factuality of reporting and bias of news media sources. 10.18653/v1/D18-1389.

Benevenuto, F., Magno, G., Rodrigues, T., and Almeida, V. (2010). Detecting spammers on twitter. Collabora- tion, electronic messaging, anti-abuse and spam con- ference (CEAS), 6:12.

Castillo, C., Mendoza, M., and Poblete, B. (2011). Infor- mation Credibility on Twitter.

Chu, Z., Gianvecchio, S., Wang, H., and Jajodia, S. (2012). Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? IEEE Transactions on Depen- dable and Secure Computing, 9(6):811–824.

Coscia, M., & Rossi, L. (2020). Distortions of political bias in crowdsourced misinformation flagging. Journal of The Royal Society Interface. 17. 20200020. 10.1098/rsif.2020.0020.

Dickerson, J. P., Kagan, V., and Subrahmanian, V. S. (2014). Using sentiment to detect bots on Twitter: Are humans more opinionated than bots? ASONAM 2014 - Pro- ceedings of the 2014 IEEE/ACM International Confe- rence on Advances in Social Networks Analysis and Mining, (Asonam):620–627.

Figueira, Á. & Guimarães, N. & Torgo, L. (2018). Current state of the art to detect fake news in social media: global trendings and next challenges. 332-339. 10.5220/0007188503320339.

Gilani, Z., Kochmar, E., and Crowcroft, J. (2017). Classification of twitter accounts into automated agents and human users. Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 - ASO- NAM '17, pages 489–496.

Hakimi A.N. et al. (2019) Identifying fake account in facebook using machine learning. In: badioze zaman h. et al. (eds) advances in visual informatics. IVIC 2019. Lecture Notes in Computer Science, vol 11870. Springer, Cham. https://doi.org/10.1007/978-3-030-34032-2_39

Nguyen, V., Sugiyama, K., Nakov, P., & Kan, M. (2020). FANG: leveraging social context for fake news detection using graph representation. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*.

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences,116*(7), 2521-2526. doi:10.1073/pnas.1806781116

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

Przybyla, P. (2020, April). Capturing the style of fake news. *In Proceedings of the AAAI Conference on Artificial Intelligence*(Vol. 34, No. 01, pp. 490-497).

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.

Thorne, James, & Vlachos, Andreas. (2018). Automated Fact Checking: Task formulations, methods and future directions.

Tschiatschek, S., Singla, A., Rodriguez, M. G., Merchant, A., & Krause, A. (2018). Fake news detection in social networks via crowd signals. *Companion of The Web Conference 2018 on The Web Conference 2018 - WWW 18*. doi:10.1145/3184558.3188722

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, *53*(5), 1-40.