A Research on Online Fake News Detection using Machine Learning Techniques

G. Purna Chandar Rao[1] and V. B. Narasimha[2]

[1]Research Scholar, Department of Computer Science and Engineering, Osmania University, Hyderabad, Telangana-500 007.
[2]Asst Professor, Department of Computer Science and Engineering, Osmania University, Hyderabad, Telangana-500 007.

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract

With the advent of modern day journalism and social media at peak, fake news may spread faster around the world. Therefore, it is important to detect the fake news and considered as a popular research topic among the community.Any changes on a particular news articles includes editorial, news report, expose, etc are predicted by using Fake News Detection (FND) techniques. Nowadays, fake news is defined as one of the major threats to economies, democracy and journalism.But, the reliability identification of online information is the most important difficult process in FND, which leads the researchers and technical developers to design an efficient techniques for improving the FNDs' performance. In this study, a discussion for detecting fake news on social media, including different kinds of news platforms, fake news characterizations, different types of data for fake news and finally, existing algorithms from a data mining perspective. In addition, the study also presents the open research problems for FND on social media.

Keywords: Detection, Democracy, Fake News, Journalism, News Articles, Reliability, Social Media.

1. Introduction

Recently, the topic of FND gained much interest among numerous researchers, where several social studies provides the impact of fake news and people are reacted to the fake news content. The fake news can be defined as any content can able to believe the readers in news which is actually not true [1,2]. A serious negative impact is created among society and individuals by the excessive spread of fake news. Initially, the authenticity balance of the news ecosystem can be modified or broken by this kind of fake news. The readers are made forcefully to accept the false or biased beliefs by the characteristics of fake news. In general, political messages or influence are conveyed by manipulating the fake news using propagandists [3,4]. Finally, the way of people's interaction and responding to real news are changed by fake news. Therefore, it is important to design an effective methods for automatically identify the fake news on social media, which will leads to the elimination of negative impacts of those fake news [5]. However, the FND on various social media has several challenging research problems. It is quite challenge for automatic FND due to the characteristics of news. Initially, readers are misled by fake news, which make it difficult for detecting the fake news according to news content.

The fake news contents are diverse on the basis of styles, media platforms and topics, where these fakes news tried to distort truth by using diverse linguistic styles [6,7]. Due to insufficient of corroborating claims or evidence, the existing knowledge bases are failed to verify the fake news properly, when these news is related to time-critical events. Moreover, a data (i.e. unstructured, noisy, incomplete and big data) is developed by fake news that are engaged with users' social media [8,9]. In recent years, researchers tried to identify the issues of fake news, their credibility on social media namely Twitter, YouTube, Facebook and television [10]. The identification of political/product opinions, user's feelings, natural phenomena in progress, events occurring around the world and satisfaction of users with health care services are analyzed by using data [11]. Hence, the useful post features are extracted and network interactions are exploited by developing effective methods for various credible users. In this research study, a characteristics of fake news, their types and detection approaches are presented. The suitable clarifications about fake news are provided for the better guidance on further researches of FDN applications. The study of traditional FDN with its advantages and limitations are discussed, where various challenges for the fake news on social media is also explained. But, there are several issues exists in the FDN on social media, which should need further investigations.

The remaining paper of the research study is composed of: Section 2 describes the taxonomy of FND. The various types of data in news are presented in Section 3. The various categories of fake news are depicted in Section 4. The explanation of methodologies to detect the fake news is illustrated in Section 5. The discussion of existing

techniques used to identify the news is presented in Section 6 and open challenges in detecting the fake news are represented in Section 7. Finally, the conclusion of this study with future development are given in Section 8.

2. Taxonomy of Fake News Detection

In this section, fake news are identified by using detection techniques from various platforms and datasets. Initially, the descriptions for various platforms are given:

2.1. Platforms:

Any new contents are provided to the end users by using carrier platforms, where specific platforms list are discussed in this section that are popular among the majority of readers. The major platforms are presented as below:

Standalone website:A new stories can be produced by any sites and a separate URL are dedicated to each story. When an user want to share or create a social media post, these URL are directly used. In general, sites are classified into three types namely blog, media and popular news sites. The authentic content are generated by the popular new sites, because they have their own social media presence. The best place to get the wrong information is the blog sites, which is highly based on user-generated content and unsupervised content. According to the media-rich content, media sites allows the users to design their sites by content creation based on style and user's interest.

Social media: To circulate the content on these sites, the most common way is the sharing. Nearly, 70% of user shares the content for daily news source. The information are shared by users with the most common social media sites namely Twitter, Facebook and Whatsapp. The fake news can be reach to larger audience by creating paid ads for any post in Facebook, where a user create a tweet with limited character and share a popular tweet with other user by retweet in Twitter.

Emails: To receive news, consumers uses a emails as an effective platform, but it is a challenging issues for validating the news emails authenticity.

Broadcast networks (Podcast):A small number of users only uses the podcasts services for news, where this service is a kind of audio multimedia category.

Radio service: The validation of audio truthfulness is a major challenging tasks in radio services, because these services act as an effective sources of news.

3. Different Types of Data in News:

The new stories are made of different types of data, which are discussed in this below section. In general, users consume the news by four major formats that are described as below:

Text: The linguistic of text is used to analyze the content of string or text, which is mainly focused on the text as a communication system. The analysis of discourse is carried out by the characteristics includes grammar, tone and pragmatics, which is much more than words and sentences.

Multimedia: The media are integrated into various forms that includes audio, graphics, images and video. At very first sights, it catches the viewers' attention due to its visual representation.

Hyperlinks or Embedded Content: The link off to various sources are enabled by hyperlinks for users and the hypothesis of news story are used to gain readers trust. A snapshot of relevant social media posts such as tweet, sound cloud clip, Facebook posts, YouTube video, etc are embedded by writers using the advent of social media.

Audio: For a news source, audio has a standalone medium, which is a one of the parts of multimedia category. The news are delivered to the greater audience by using this medium that has various media such as radio services, podcast and broadcast network.

4. Types of Fake News:

A fake news are studied from various perspectives by social science researchers, then a general categorization of various types of fake news are provided. The below statement presented these categorization:

i. **Visual-based:**A graphical representation of video, photo-shopped images or combination of both is used in the content to describe the categories of fake news [12].
ii. **User-based:**In this type, the target audience can be obtained by creating fake accounts, where the audience are represented by particular gender, age groups, culture, etc.
iii. **Post-based:**This kind of fake news will be mainly appeared on social media platforms, such as Facebook post with video or image caption, memes, tweets, etc.

iv. **Network-based:**A certain members of a specific organization are connected with this kind of fake news, where this idea is mainly applied to group of connected individuals on LinkedIn and friends-of-friends on Facebook.

v. **Knowledge-based:**To spread fake information, this kind of new stories will be designed, where the documents consists of reasonable explanation or scientific information to an unresolved issues.

vi. **Style-based:**The fake news can be written by the people, who are well-being to write the information in different style, where this style-based news focused only on the presentation of fake news to the end-users.

5. Detection Methods for identifying fake news

The characterization of fake news and various types of news data are explained in the previous section, where several existing techniques are developed based on the feature extraction. In the below statements, different types of feature based techniques are discussed:

5.1. Linguistic Features based Methods

The key linguistics features are extracted by using the linguistic based approaches. There are various features such as syntax, Ngrams, psycho-linguistic, punctuations and readability features, where the most important features are depicted as follows:

- **Ngrams:**In a story, the collection of words are used to extract the unigrams and bigrams. These extracted features are stored as Term Frequency Inverse Document Frequency (TFIDF) for retrieving the information.
- **Punctuation:**The difference between truthful and deceptive texts are illustrated by using the punctuation in the FND algorithms.
- **Syntax:**According to Context-Free Grammar (CFG), a set of features are extracted by this technique. Based on the lexicalized production rules which is a combination of parent and their grandparent nodes, these set of features are used.

5.2. Deception Modeling based Methods

According to two theoretical approaches namely Vector Space Modeling (VSM) and Rhetorical Structure Theory (RST), the process of clustering the truthful vs. deceptive stories are conducted. By applying RST, a set of rhetorical relations will be obtained by analyzing every text in a hierarchical tree. Finally, the results of RS relations are identified by using VSM. When compared with similarity based cluster analysis, the RST-VSM method provides an curating data edge that is based on the distance between samples.

5.3. Clustering based Methods

A vast amount of data will be compared and contrasted is defined as the process of clustering. While running a huge amount of clusters, a small number of clusters are formed or sorted with the help of k-nearest neighbor and agglomerative clustering approach. According to the normalized frequency of relations, the news reports are clustered and the deceptive value of a news story is identified on the basis of coordinate distances principle. The accurate results are not provided by this approach, when it is applied on recent fake news story, because similar reports are not available.

5.4. Non-Text Cues based Methods

The non-text content of news are used to convince the readers for having a faith in contaminated news, which is mainly focused by this technique. Here, two various analyses are used that are described as:

- **Image Analysis:**The emotion in observers are manipulated by using known key method, which is also known as strategic use of images.
- **User Behavior Analysis**: The behavior of readers (i.e. how they engaged with news) are assess by using content-independent method called user behavior analysis. The main idea of the method is to understand the user behavior and their teasing images on social media.

5.5. Content Cues based Methods

According to the ideology of user's choices to read the news and way of writing the news for users by journalists, this method will be developed. These news stories are developed by forwarding the same messages more than one sources, but written in various ways. Two various analyses are presented in this method as explained below:

- **Lexical and Semantic Levels of Analysis:**In the story, readers should believe the fake news as real story by convincing them with choice of vocabulary. The difference between two journalistic formats are identified by extracting the stylometric features of text using automated methods.
- **Syntactic and Pragmatic Levels of Analysis:**In the discourse, the reference for upcoming parts are identified by using the pragmatic function. The leveraging ensuring texts are filled with empty thoughts by writing headlines.

6. Literature Review

In this section, the survey of recent techniques used for detecting the fake news on social media are presented. The Table 1 describes the methodology, advantage and limitation of existing techniques. The parameter evaluations for validating the FNDs techniques are also explained.

Table 1: Comparative Analysis of Existing Techniques for fake news detection

Year	Authors	Methodology	Advantage	Limitation	Parameter Evaluation
2019	F. A. Ozbay, and B. Alatas, [13]	Two-step methods such as combination of text mining algorithms and supervised artificial intelligence algorithms were proposed to detect the fake news.	The structured dataset are obtained from un-structured datasets by using document-term matrix and TF weighting method	Among all the supervised algorithms, KLR provides very poor performance in all parameter evaluation and failed to detect the fake news in real-world datasets.	Accuracy, recall, precision and F-measure are used to validate these combined algorithms.
	K. Xu, *et al.,* [14]	Term frequency-inverse document frequency are applied to identify the content characterizations. Latent Dirichlet Allocation (LDA) are designed to detect the fake news.	The detection of fake and real news on social media are effectively carried out by studied the domain reputations and content understanding.	The similarity and dissimilarity of the content are captured only for few important words of each article.	Document similarities between real and fake news is used as parameter evaluation for identifying the effectiveness of LDA.
	M. BalaAnand, *et al.,* [15]	From the large volume of Twitter data, the fake users and news are detected by designing an enhanced graph-based supervised learning algorithm as EGSLA.	EGSLA algorithm effectively predicted the fake user and news on Twitter by extracting the important features, which is identified on the weighted graph.	The method worked effectively only on labeled data, where unlabeled data are not considered by EGSLA algorithm.	Accuracy, precision, sensitivity, specificity, Mathews Correlation coefficient and F-measure are used to validate the EGSLA against decision tree, SVM and KNN.
	M. Visentin, *et al.,* [16]	identified the fake news transfer of individuals' perceptions to an adjacent brand advertisement.	The difference between percerived credibility of the news and real news are studied and manipulated by observing the changes	The perceived credibility of the sources are highly affected due to the impact of fake news on user behavioral intentions.	The direct effect of behavioral intentions towards brands is used to test the effects on fake news.

			in the behavior of user.		
	D. K. Vishwakarma, *et al.,* [17]	Four integrated units such as entity extractor, extraction of text, scraping the web and processing unit is designed for the FND.	The similarities between extracted entities and page title for selected keywords is identified to remove the false positives.	The classification of local news are not focused by these approaches. In addition, the method is insufficient to extract the text due to presence of image characteristics along with text.	Accuracy, precision, recall and f-measure are used to validate the effectiveness of novel method.
	L. Wang, *et al.,* [18]	designed a principled automated approach to distinguish these different cases while assessing and classifying news articles and claims.	The method provide an analysis of different kinds of fake contents, considering both linguistic characteristics of user posts and the sharing dynamics in Twitter.	The adversarial behavior is not considered for the prediction and also the reactions on social media, which is used to understand the intents behind misinformation.	The parameters such as Accuracy, mean squared error (MSE) and F1-score are used to validate this approach
	P. Shi, *et al.,* [19]	designed a novel method for accurately detecting malicious social bots in online social networks.	The method analyzes the time feature of user behavior and transaction probability of their clickstreams.	The specific intensions of the malicious social bots in online social platforms are not identified.	Accuracy, precision, recall and F-score are used to evaluate the results of proposed algorithm against SVM technique.
2018	T. Mondal, *et al.,* [20]	developed a content-based analysis for ensuring the extracted tweets contributions.	The proposed rumor detection technique performed well and is able to find out rumors at early stages, even before contradicting or interrogating posts are posted.	After early detection of a rumor, the method didn't plan to devise effective rumor control strategies.	Accuracy, precision, recall and f-score are used as a parameter metrics for validating the performance of proposed model.
2017	P. Dewan, and P. Kumaraguru, [21]	identified the malicious posts in real-time Facebook data by designing an Facebook Inspector called REST browser.	Facebook Inspector detects malicious posts in real time without depending on any engagement metrics associated with a post (likes, comments, or shares)	The current architecture of FBI is restricted to public Facebook posts only. The FBI will not helpful to address the Zero-attack problem.	Accuracy, response time, precision, recall and ROC_AUC are used for validating the performance of FBI.

2016	Z. Jin, *et al.*, [22]	developed a different features of visual and statistical patterns for the identification of fake news	The image statistics and attribute information are summarized by statistical features in news events. The verification performance are improved by capturing the image distribution pattern quantitatively.	The dependency information is missed, when the proposed methodology trained the models on image and non-image dataset separately.	The parameters such as accuracy, f1-score, precision and recall are used for validation process.

7. Open Research Challenges

The main challenges of FND that will lead to future research are given in the following statements:

Multi-modal Verification Method: The fake news can be detected by developing various methods using linguistic approaches, but people highly believes the fake news content via visual presentation. Therefore, it paves a ways to verify not only languages, but also audio, images, hyperlinks and embedded content such as tweets, embedded video, Facebook post.

Source Verification: In the proposed existing techniques, the source of news story is not concentrated for effective identification of FDN. This leads to the development of new FDN techniques to verify the sources of fake news stories.

Author Credibility Check: The tones of a story are identified by one third of the total existing methods for fake news validation. Therefore, a system should design for verifying the author credibility, so that the chain of news with same group of authors are detected.

Multi-modal Data-set:A complete multi-modal collection of fake news are not provided by any of the standard datasets, which is considered as one of the major challenging issues. Therefore, it will leads to the creation of new multi-modal dataset which covers all various types of fake news data.

8. Conclusion

The traditional news media highly relies only on news content, where the additional information as extra social context auxiliary information are provided for the FDN by social media. Due to the popularity of social media, nowadays, people fetch the news content from the Facebook, YouTube, Twitter, etc rather than traditional news media. But, the fake news are highly spread by the social media that had a negative impact on group of people or individual users. In this research study, there are two various phases includes characterization and detection that are described to explore the problem of fake news by existing techniques. The basic principles and concepts of fake news in traditional media as well as social media are illustrated in the characterization phase, where the existing FDN approaches with its advantages and limitations are presented in detection phase. In addition, evaluation metrics as well as research problems in detecting the fake news are presented.

Reference

[1]K. S. Adewole, T. Han, W. Wu, H. Song, and A. K. Sangaiah, (2018). "Twitter spam account detection based on clustering and classification methods". *The Journal of Supercomputing*, 1-36.

[2] M. Z. Asghar, A. Ullah, S. Ahmad, and A. Khan, (2019). "Opinion spam detection framework using hybrid classification scheme". *Soft Computing*, 1-24.

[3] C. Boididou, S. Papadopoulos, M. Zampoglou, L. Apostolidis, O. Papadopoulou, and Y. Kompatsiaris, (2018)."Detection and visualization of misleading content on Twitter".*International Journal of Multimedia Information Retrieval*, 7(1), 71-86.

[4] Y. Boshmaf, D. Logothetis, G. Siganos, J. Lería, J. Lorenzo, M. Ripeanu, and H. Halawa, (2016). "Íntegro: Leveraging victim prediction for robust fake account detection in large scale OSNs". *Computers & Security*, 61, 142-168.

[5] K. Dhingra, and Sumit Kr Yadav. "Spam analysis of big reviews dataset using Fuzzy Ranking Evaluation Algorithm and Hadoop." *International Journal of Machine Learning and Cybernetics* (2017): 1-20.

[6] Jang, Boyeon, SihyunJeong, and Chong-kwon Kim. "Distance-based customer detection in fake follower markets."*Information Systems* 81 (2019): 104-116.

[7]E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, (2019). "A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making". *Industrial Marketing Management*.

[8] D. Plotkina, A. Munzel, and J. Pallud, (2018). "Illusions of truth—Experimental insights into human and algorithmic detections of fake online reviews".*Journal of Business Research*.

[9] Y. Qin, W. Dominik, and C. Tang, (2018)."Predicting future rumors".*Chinese Journal of Electronics*, 27(3), 514-520.

[10] H. Allcott, and Matthew Gentzkow. "Social media and fake news in the 2016 election."*Journal of economic perspectives* 31.2 (2017): 211-36.

[11] C. S. Atodiresei, A. Tănăselea, and A. Iftene, "Identifying Fake News and Fake Users on Twitter".*Procedia Computer Science*, 126, 451-461, 2018.

[12] Li, Yuenan. "Image copy-move forgery detection based on polar cosine transform and approximate nearest neighbor searching." *Forensic science international* 224.1-3 (2013): 59-67.

[13] F. A. Ozbay, and B. Alatas. "Fake news detection within online social media using supervised artificial intelligence algorithms." *Physica A: Statistical Mechanics and its Applications*, 123174, 2019.

[14] K. Xu, F. Wang, H. Wang, and B. Yang, "Detecting fake news over online social media via domain reputations and content understanding".*Tsinghua Science and Technology*, 25(1), 20-27, 2019.

[15] M. BalaAnand, N. Karthikeyan, S. Karthik, R. Varatharajan, G. Manogaran, and C. B. Sivaparthipan, "An enhanced graph-based semi-supervised learning algorithm to detect fake users on Twitter". *The Journal of Supercomputing*, 75(9), 6085-6105, 2019.

[16] M. Visentin, G. Pizzi, and M. Pichierri. "Fake News, Real Problems for Brands: The Impact of Content Truthfulness and Source Credibility on consumers' Behavioral Intentions toward the Advertised Brands." *Journal of Interactive Marketing*, 45, pp. 99-112, 2019.

[17] D. K. Vishwakarma, D. Varshney, and A. Yadav. "Detection and veracity analysis of fake news via scrapping and authenticating the web search." *Cognitive Systems Research* 58, pp. 217-229, 2019.

[18] L. Wang, Y. Wang, G. De Melo, and G. Weikum, "Understanding archetypes of fake news via fine-grained classification".*Social Network Analysis and Mining*, 9(1), 37, 2019.

[19] P. Shi, Z. Zhang, and K. K. R. Choo. "Detecting Malicious Social Bots Based on Clickstream Sequences." *IEEE Access* 7, pp. 28855-28862, 2019.

[20] T. Mondal, P. Pramanik, I. Bhattacharya, N. Boral, and S. Ghosh, "Analysis and early detection of rumors in a post disaster scenario".*Information Systems Frontiers*, 20(5), 961-979, 2018.

[21] P. Dewan, and P. Kumaraguru. "Facebook Inspector (FbI): Towards automatic real-time detection of malicious content on Facebook." *Social Network Analysis and Mining* 7.1 (2017): 15.

[22] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel visual and statistical image features for microblogs news verification". *IEEE transactions on multimedia*, 19(3), 598-608, 2016.