

Fake News Detection Using Machine Learning approaches: A systematic Review

Syed Ishfaq Manzoor
Research Scholar
Lovely Professional University
Phagwara Punjab India
esishfaq@gmail.com

Dr Jimmy Singla
Assistant Professor
School of Computer Science &
Engineering
Lovely Professional University
Phagwara Punjab India

Nikita
Former Assistant Professor
School of Computer Science &
Engineering
Lovely Professional University
Phagwara Punjab India

Abstract –The easy access and exponential growth of the information available on social media networks has made it intricate to distinguish between false and true information. The easy dissemination of information by way of sharing has added to exponential growth of its falsification. The credibility of social media networks is also at stake where the spreading of fake information is prevalent. Thus, it has become a research challenge to automatically check the information viz a viz its source, content and publisher for categorizing it as false or true. Machine learning has played a vital role in classification of the information although with some limitations. This paper reviews various Machine learning approaches in detection of fake and fabricated news. The limitation of such and approaches and improvisation by way of implementing deep learning is also reviewed.

Keywords—fake news, machine learning , deep learning.

INTRODUCTION:

Classification of any news item /post / blog into fake or real one has generated great interest from researchers around the globe. Several research studies have been carried out to find effect of falsified and fabricated news on masses and reactions of people upon coming through such news items. Falsified news or fabricated pos newst is any textual or non-textual content that is fake and is generated so the readers will start believing in something which is not true. For Example recently a news item was floated on social media networking platform Facebook by a accredited Journalist of Srinagar J&K Titled “ Beasts in White Aprons ” regarding mismanagement and carelessness of doctors in a local Pediatric hospital of Srinagar with a Image (refer to left part in Figure 1 Original but misinterpreted) .



Newborn burnt to death at GB Pant Hospital Srinagar by doctors and paramedics who kept her very close to blower, the bereaved family said. To hide the murder, hospital administration rushed body out without ambulance to her home in Kreeri. Kindly share for public awareness
#BeastsInWhiteAprons
#MJH



The post was shared 1.2 K times in span of 01 day. Later on the news was proved to totally false and misrepresented. The two key components of this post are (see Figure 1):

- Author of the post and
- Image associated with post.

The following are the psychological factors which convince a readers to believe in any falsified post or news item :

Author: Author of post makes a compelling statement about negligence and careless of doctors in the hospital. Generally nobody will do a fact checking on such news when it is posted by an accredited journalist and is supported by the a valid image.

Image: The impact of a false or fabricated news is increased many fold when it is supported by an image. In this post the image supports what post is alarming about. Image depicts the burned leg of an infant.

Source/Author : The author plays a vital role about the credibility of the news/post.

Platform : Since the userbase of the platform , Facebook in this case , is very huge and diversified which results in its exponential growth.

Although there are multiple news carrier platforms which include SNS , stand alone websites , blogs and microblogs emails etc. But this paper will emphasize on Social Media networking sites (Facebook and Twitter).

TYPE OF DATA IN SOCIAL MEDIA POSTS

As discussed by the authors of [1] there are three major forms in which social media networking Sites read a news item :

Text (Multilingual) is analyzed by computational linguistics which focuses the genesis of text semantically and systematically. since much of the posts are produced in the form of texts much work has been carried out on its analysis.

Multimedia: Multiple forms of media are integrated in a single post. This may includes audio, video, images, and graphics. This is very much attractive and it fetches the attention of the viewers without bothering about the text.

Hyperlinks enable the originator of the post to cross reference to different sources and thus gains viewers the trust by certifying genesis of the post. Even cross reference to other social media networking sites and embedment of snapshots is in practice .

FAKE NEWS TYPES

The various types of fake news by Authors of paper [2] , in their recent paper is summarize below.

1. **Visual-based:** These fake news posts use graphics a lot more in as content, which may include morphed images, doctored video, or combination of both [5].
2. **User-based:** This type of fabricated news is generated by fake accounts and is targeted to

specific audience which may represent certain age groups, gender, culture, political affiliations .

3. **Knowledge-based :** these types posts give scientific (so called) explanation to the some unresolved issues and make users to believe it is authentic. For example natural remedies of increased sugar level in human body .
4. **Style-based** posts are written by psedo-journalists who pretend and copy style of some accredited journalists
5. **Stance-based:** It actually is representation of truthful statements in such a way which changes its meaning and purpose.

BACK GROUND :

As per authors in [13] the Information fabrication is not new. The authors have referred to columnist of Guardian Natalie Nougayrède who says “ The use of propaganda is ancient, but never before has there been the technology to so effectively disseminate it”. Falsified information, distorted information propaganda based information and fun based false information have been human communication features since Roman times .

The impact and penetration of social media have dramatically changed reach of falsified information. The introduction of smart gadgets and very low cost internet cost have added to its reach. In India even the remotest village has access to smart phones and internet services. Although there are numerous advantages of thses services but it comes at a cost , the cost of rapid dissemination of falsified information along with substantiate information.

Last 10 years have witnessed manifold increase in the number of users on the social media and microblogging. The data/text available on these sites in the form of news , blogs, posts , reviews, opinions, suggestions, arguments, comments etc. provides growth in the field of techniques and methods in the authenticity of these posts. Many researches have been conducted where machine learning have been used to automatically detect the fake news items. Also few research works have been conducted using deep learning for auto feature extraction in fake news detectors.

LITERATURE SURVEY ON FACE NEWS DETECTION

The available literature has described many automatic detection techniques of fake news and deception posts. Since there are multidimensional aspects of fake news detection ranging from using chatbots for spread of misinformation to use of clickbaits for the rumor spreading . There are many clickbaits available in social media networks including facebook which enhance sharing and liking

of posts which in turn spreads falsified information. Lot of work has been done to detect falsified information.

Various detection techniques have been introduced by authors in [1] . the authors have introduced following Detection Methods

1. Linguistics basis

Deception modelling

2. Clustering
3. Predictive modelling
4. Content cue based methods
5. Non text cue based methods

The authors have tabulated Fake news methods for different fake news types as :

Fake News Type	Fake news detection Method					
	Linguistic Modeling	Deceptive	Clustering	Predictive Modeling	Content Cues	Non-Text Cues
Visual-based	No	No	No	No	No	Yes
User Base	No	No	No	Yes	Yes	Yes
User Post Based	Yes	Yes	Yes	Yes	No	Yes
Social Network Based	No	No	No	No	Yes	No
Knowledge Based	No	No	Yes	No	No	No
Style Based	Yes	No	No	Yes	Yes	No
Stance Based	No	No	No	No	No	No

The Authors have shown accuracy of these models between 63 to 70 percent only.

The Authors in [2] have described Linguistic Cue Approaches with Machine Learning , Bag of words approach ,Rhetorical Structure and discourse analysis ,Network analysis approaches and SVM classifiers. These are models are text based only and have very little or negligible improvement on existing methods.

The authors of [3] have classified every tweet/post as binary classification Problem. The Classification is purely on the basis of source of the post/tweet. The Authors used manually collected data sets using twitter API , DMOZ . The following algorithms where used on data sets

1. Naïve Bayes
2. Descision trees
3. SVM
4. Nueral Networks
5. Random Forest
6. XG Boost.

The results show 15 percent fake tweets, 45 % real tweets , rest posts where undecided.

The Authors of [4] in their paper suggested deception detection using labelled benchmark data set ‘ LIAR ’ with evident improved efficiency in detection of fake posts/news. The Authors argued the use of corpus for classification of stance ,opinion mining, rumor detection, and political NLP research.

The authors of [7] have Introduced Need for hoax detection . They Used ML approach by combining news content and social content approaches. The authors Claim the performance is good as compared to described in literature. The authors Implemented it with Facebook messenger chatbot. Three different datasets of Italian news posts of Facebook where used. Both content based methods with social and content signals using Boolean crowd sourcing algorithms where implemented . The following Methods where used by the :

1. Content based
2. Logistic regression on social signals.
3. Harmonic Boolean label crowdsourcing on social signals.

The authors in [9] observed about 14 million messages retweeted about 400 thousand times on Twitter during and following the 2016 U.S. Presidential campaign and election by bots. The methods to categorize the posts spread by bots where described.

The Authors in [10] have described Tabloidization in the form of Click baiting. They have described Click baiting as a form of rapid dissemination of rumor and misinformation online . The authors have discussed potential methods for automatic detection of clickbait as a form of deception. Content cues which includes lexical and semantic level of analysis where implemented by the authors.

The authors in [13] observed the principles, methods and algorithms employed for classification of falsified and fabricated news items, authors and subjects from online social networks and evaluating the corresponding reach and performance. The paper also suggested the research challenges through the undiscovered characteristics of fake news and diverse connections among news articles, authors and subjects. The Authors of the paper discuss automatic fake news inference model named as FakeDetectorIt is based on textual classification and builds a deep diffusive network model to learn the representations of news articles, authors and subjects simultaneously

FakeDetector addresses two main components: representation feature learning, and credibility label inference, which together will compose the deep diffusive network model FakeDetector

CONCLUSION:

Although there is evident success in detection of fake news and posts using various Machine learning approaches. However everchanging characteristics and features of fake news in social media networks is posing a challenge in categorization of fake news . However the main characteristic feature of deep learning is to compute hierarchical features . With the implementation of deep learning research and applications in recent past, lots of research works is going to implement deep learning methods, like convolutional Neural Networks [14], deep Boltzmann machine [15], Deep neural network [16, 17] and Deep autoencoder model [18], in various applications, like audio and speech processing [19, 20], Natural language processing and modelling [21, 22], information retrieval [23, 15], objective recognition and computer vision [24], as well as multimodal and multi-task learning [25, 26] in categorization of news posts.

REFERENCES

- [1]. Parikh, S. B., & Atrey, P. K. (2018, April). Media-Rich Fake News Detection: A Survey. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 436-441). IEEE.
- [2]. Conroy, N. J., Rubin, V. L., & Chen, Y. (2015, November). Automatic deception detection: Methods for finding fake news. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community (p. 82). American Society for Information Science.
- [3]. Helmstetter, S., & Paulheim, H. (2018, August). Weakly supervised learning for fake news detection on Twitter. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 274-277). IEEE.
- [4]. Wang, W. Y. (2017). "liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648.
- [5]. Stahl, K. (2018). Fake News Detection in Social Media.
- [6]. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436.
- [7]. Della Vedova, M. L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., & de Alfaro, L. (2018, May). Automatic Online Fake News Detection Combining Content and Social Signals. In 2018 22nd Conference of Open Innovations Association (FRUCT) (pp. 272-279). IEEE.
- [8] Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & de Alfaro, L. (2017). Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506.
- [9]. Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. arXiv preprint arXiv:1707.07592, 96-104.
- [10]. Chen, Y., Conroy, N. J., & Rubin, V. L. (2015, November). Misleading online content: Recognizing clickbait as false news. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection (pp. 15-19). ACM.
- [11]. Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1), 1.
- [12]. Haiden, L., & Althuis, J. (2018). The Definitional Challenges of Fake News.
- [13]. Zhang, J., Cui, L., Fu, Y., & Gouza, F. B. (2018). Fake news detection with deep diffusive network model. arXiv preprint arXiv:1805.08751.
- [14]. G. Hinton, S. Osindero, and Y. Teh. A fast learning algorithm for deep belief nets. *Neural Comput.*, 2006.
- [15]. R. salakhutdinov and G. Hinton. Semantic hashing. *International Journal of Approximate Reasoning*, 2009.
- [16].] H. Jaeger. Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach. Technical report, Fraunhofer Institute for Autonomous Intelligent Systems (AIS), 2002
- [17]. A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.
- [18] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 2010.

- [19] L. Deng, G. Hinton, and B. Kingsbury. New types of deep neural network learning for speech recognition and related applications: An overview. In ICASSP, 2013
- [20] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury. Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine, 2012.
- [21] E. Arisoy, T. Sainath, B. Kingsbury, and B. Ramabhadran. Deep neural network language models. In WLM, 2012
- [22] A. Mnih and G. Hinton. A scalable hierarchical distributed language model. In NIPS. 2009.
- [23] G. Hinton. A practical guide to training restricted boltzmann machines. In Neural Networks: Tricks of the Trade (2nd ed.). 2012.
- [24] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. Nature, 521, 2015. <http://dx.doi.org/10.1038/nature14539>.
- [25] J. Weston, S. Bengio, and N. Usunier. Large scale image annotation: Learning to rank with joint word-image embeddings. Journal of Machine Learning, 2010.
- [26] J. Weston, S. Bengio, and N. Usunier. Wsabi: Scaling up to large vocabulary image annotation. In IJCAI, 2011
- [27] Riedel, B., Augenstein, I., Spithourakis, G. P., & Riedel, S. (2017). A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. arXiv preprint arXiv:1707.03264.
- [28] Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2018). TI-CNN: Convolutional neural networks for fake news detection. arXiv preprint arXiv:1806.00749.
- [29] Huh, M., Liu, A., Owens, A., & Efros, A. A. (2018). Fighting fake news: Image splice detection via learned self-consistency. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 101-117).