Fake News Detection: Literature Survey

Recently there have been several works related to fake news. Through this section, we summarize some of the existing research works about Fake News Detection in the field of Machine learning/deep learning.

Since most of the work has been focusing on either the text, the response an article receives, or the users who source it. Ruchansky et al. [2] incorporate all three and proposed the CSI model which is composed of three modules – Capture, Score and Integrate. First module is based on the response and text, it uses a Recurrent Neural Network to capture the temporal pattern of user activity on a given article. The second module learns the source characteristic based on the behaviour of users by using L2 regularization, and the two are integrated with the third module to classify an article as fake or not.

Fatima et al. [3] presented a dataset of fake news by scraping articles from multiple media sources around the Syrian War and used a semi-supervised fact-checking approach to label articles in their dataset. They used crowd sourcing to extract the ground truth and then used it to cluster the articles into two separate sets using unsupervised machine learning.

Granik et al. [1] implemented Naïve Bayes` Classifier as a software system and tested it on randomly shuffled fake news dataset retrieved from Facebook API and achieved an accuracy of 74%.

Traylor et al. [5] used Glaser and Strauss's grounded theory to look out for linguistic patterns and differences to develop a machine learning grammar and hypothesis. They developed a fake news identification corpus too and designed an algorithm or a scoring system called attribution-score or A-score to classify the fake and real document.

Yang et al. [4] used an unsupervised learning approach by treating truths of news and users' credibility as latent random variables, and exploiting users' engagements on social media to identify their opinions towards the authenticity of news. They [4] leverage a Bayesian network model to capture the conditional dependencies among the truths of news, the users' opinions, and the users' credibility. To solve the inference problem, they [4] propose an efficient collapsed Gibbs sampling approach to infer the truths of news and the users' credibility without any labelled data.

Ajao et al. [8] proposed a framework that detects and classifies fake news messages from Twitter posts using hybrid of convolutional neural networks and long-short term recurrent neural network models and achieved 82% accuracy.

Liu et al. [6] investigated the correlation between user profiles and fake/real news on social media and performed a comparative analysis over explicit and implicit profile features between these user groups, which reveals their potential to differentiate fake news. They [6] concluded that there are specific users who are more likely to trust fake news than real new and these users reveal different features from those who are more likely to trust real news. These observations ease the feature construction of profiles features for fake news detection.

Sliva et al. [15] used data mining perspective to detect fake news on social media by introducing features extracted from different sources, i.e., news content and social context and categorized existing methods based on their main input sources as: News Content Models and Social Context Models

Shu et al. [7] investigated the key achievements of user identity linkage across online social networks including state of-the-art algorithms, evaluation metrics, and representative dataset. They [7] introduced a unified framework for the user identity linkage problem, which consists of two phases: Feature extraction from profile, content and network information and Model construction in supervised, semi-supervised and unsupervised ways.

Zubiaga et al. [9] summarised studies reported in the scientific literature toward the development of rumour classification systems, defining and characterising social media rumours and has described the different approaches to the development of their four main components: (1) rumour detection, (2) rumour tracking, (3) rumour stance classification, and (4) rumour veracity classification.

Buntain et al. [10] used the structural features, user features, content features and temporal features are used for predicting accuracy using two datasets: CREDBANK, a crowdsourced dataset of accuracy assessments for events in Twitter, and PHEME, a dataset of potential rumours in Twitter and then evaluate how well each dataset predicts truth in the BuzzFeed News fact-checking dataset.

Ahmed et al. [11] used n-gram analysis, TF and TF-IDF feature extraction techniques and six different classification algorithms which are Stochastic Gradient Descent, Support Vector Machines, Linear Support Vector Machines, K-Nearest Neighbour and Decision Trees and achieves highest accuracy 92% when using unigram features and Linear SVM classifier.

Wang et al. [12] proposed an end-to-end framework named Event Adversarial Neural Network (EANN), which can derive event invariant feature and it consists of three main components - the multi-modal feature extractor, the fake news detector, and the event discriminator. To validate the effectiveness of the proposed model, they [12] chosen baselines from the three categories: single modality models, multi-modal models, and the variant of the proposed model.

Abedalla et al. [13] developed different models to detect fake news and assembled the models from Convolutional Neural Network (CNN), Long Short-Term Memory network (LSTM) and Bidirectional LSTM (Bi-LSTM) and achieved 71.2% accuracy as the best one.

Shabani et al. [14] proposed a method that uses a hybrid machine-crowd approach for detection of potentially deceptive news using TF-IDF, sentiment related features and LIWC paralinguistic feature extractor,for evaluation they used five machine learning classification models: Logistic Regression, SVM, Random Forest, Neural Networks, and Gradient Boosting Classifier where the Neural Network Model gives the highest accuracy of 81.64%.

Shu et al. [16] proposed a tri-relationship embedding framework TriFN, which modelled publisher-news relations and user-news interactions simultaneously for fake news classification and tested the baseline features on different learning algorithms thus choosing the one that achieves the best performance, the algorithms include Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, XGBoost, AdaBoost, and Gradient Boosting.

References

1. Granik, Mykhailo, and Volodymyr Mesyura. 'Fake news detection using naive Bayes classifier.' 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON). IEEE, 2017.
2. Ruchansky, Natali, Sungyong Seo, and Yan Liu. 'Csi: A hybrid deep model for fake news detection.' Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017.
3. Abu Salem, F.K., Al Feel, R., Elbassuoni, S., Jaber, M. and Farah, M. 2019. FA-KES: A Fake News Dataset around the Syrian War. Proceedings of the International AAAI Conference on Web and Social Media. 13, 01 (Jul. 2019), 573-582.
4. Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., & Liu, H. (2019). Unsupervised Fake News Detection on Social Media: A Generative Approach. Proceedings of the AAAI Conference on Artificial Intelligence, 33, 5644–5651. doi:10.1609/aaai.v33i01.33015644
5. Traylor, T., Straub, J., Gurmeet, & Snell, N. (2019). Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as a Supervised Learning Estimator. 2019 IEEE 13th International Conference on Semantic Computing (ICSC). doi:10.1109/icosc.2019.8665593
6. Kai Shu, Suhang Wang, and Huan Liu. 2018. Understanding User Profiles on Social Media for Fake News Detection. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE.
7. K. Shu, S. Wang, J. Tang, R. Zafarani, and H. Liu, “User identity linkage across online social networks: A review,” ACM SIGKDD Explorations Newsletter, vol. 18, no. 2, pp. 5–17, 2017.
8. Ajao, Oluwaseun, Deepayan Bhowmik, and Shahrzad Zargari. 'Fake news identification on twitter with hybrid cnn and rnn models.' Proceedings of the 9th International Conference on Social Media and Society. 2018.
9. Zubiaga, Arkaitz, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. ”Detection and resolution of rumours in social media: A survey.” ACM Computing Surveys (CSUR) 51, no. 2 (2018): 32.
10. Buntain, Cody, and Jennifer Golbeck. 'Automatically Identifying Fake News in Popular Twitter Threads.' Smart Cloud (Smart Cloud), 2017 IEEE International Conference on. IEEE, 2017.
11. Ahmed, H., Traore, I., & Saad, S. (2017). Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, 127–138. doi:10.1007/978-3-319-69155-8_9
12. Wang, Yaqing, et al. 'Eann: Event adversarial neural networks for multi-modal fake news detection.' Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining. 2018.
13. Abedalla, Ayat, Aisha Al-Sadi, and Malak Abdullah. 'A Closer Look at Fake News Detection: A Deep Learning Perspective.' Proceedings of the 2019 3rd International Conference on Advances in Artificial Intelligence. 2019.
14. Shabani, S., & Sokhn, M. (2018). Hybrid Machine-Crowd Approach for Fake News Detection. 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC). doi:10.1109/cic.2018.00048
15. 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.

doi:10.1145/3137597.3137600
16. Shu, Kai, Suhang Wang, and Huan Liu. 'Beyond news contents: The role of social context for fake news detection.' Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. 2019.