Role of Contextual Features in Fake News Detection: A Review

Joma George
Department of Computer Science & Engineering,
Amal Jyothi College Of Engineering, Koovappally, Kanjirappally,
Kerala, India
jomageorge@cs.ajce.in

Shintu Mariam Skariah
Department of Computer Science & Engineering,
Amal Jyothi College of Engineering, Koovappally, Kanjirappally,
Kerala, India
shintumariamskariah@cs.ajce.in

Aleena Xavier T
Department of Computer Science & Engineering,
Amal Jyothi College Of Engineering, Koovappally, Kanjirappally,
Kerala, India
taleenaxavier@amaljyothi.ac.in

***Abstract*—Deceptive news has an intention to defame a person, an institution or an organization. It often has a catastrophic effect on the views of the readers. With the advancement of technology, social media and online sites are made so accessible that, it acts as a catalyst for spreading fake news. Identification of such misinformation is a challenging task even for humans. But with Deep learning and Machine learning techniques, there have been efforts to solve this problem. In this paper the influence of linguistic characteristics and contextual features in fake news detection are analyzed and certain techniques like Naive Bayes, KNN, SVM, Decision tree, Hybrid CNN, CMS etc. are compared.**

Keywords— Fake news detection, Contextual information, Machine Learning, Deep Learning, Multi-head self-attention

I. INTRODUCTION

The evolution of information technology has made a tremendous influence in society. Instant access and information sharing in a blink has made online platform and social media more favorable to people. The lack of supervision in the field of social media and online sites leads to the propagation of fake news. Several studies show that fake news has created a significant influence in the 2016 US presidential election [1]. To mitigate the effects of fake news in online media leads to the development of an automatic fake news detection system based on deep learning techniques.

Fake news detection faces many challenges; with human intellectual quotient, only 50% of deceptive news can be detected. These are often emotionally colored, confusing and also possess many grammatical errors. Spam detection and rumor detection are similar to the deceptive detection system.

Early works on detecting deceptive news is purely based on linguistic characteristics. The linguistic characteristic helps to determine text properties, such as writing styles and language based context. These features are insufficient to discriminate between real and fake news articles. Contextual features like speaker, credit history, home state of the speaker, job title, party affiliation, etc. play a vital role in determining fake news.

II. LITERATURE SURVEY

Based on the detection of fake news in the social media and online platform, different techniques for detecting fake news are rendered:

A. Using Naive Bayes

Artificial Intelligence algorithms like Naive Bayes classifier is used to classify the deceptive news [2]. Facebook news posts are collected to create a dataset for learning and testing the system. The model is implemented as a software system for classifying Facebook news post [3]. Spam messages and deceptive news articles are almost similar i.e. both are emotionally touched, the content may not be correct and sometimes it may have some grammatical issues too [4]. Naive Bayes classifiers can be defined as uncomplicated probabilistic classifiers built on Bayes theorem. Bag of words are used to identify features in deceptive news. It may work by correlating the token of words. By using Bayes theorem conditional probability of the news article is computed.

Probability value depends on the frequent occurrence of words, unknown words and also some specific words in the articles. For frequent words in a document, the conditional probability of 0.5 is assigned. Unknown words are not considered while calculating conditional probability for such words in the news articles. By [2] Naive Bayes have an accuracy of 74%. The main drawbacks of [2] paper are small dataset is used for training and also the ignorance of preprocessing steps (stop removal, stemming).

B. Using KNN

In [5], a dataset with tweets on 'Hilary Clinton' is used to create a dataset, which includes both malicious and credible tweets. The most effective features for classification can be determined, by comparing the sentiment magnitude for credible and malicious tweets and also by considering the relationship among the words. Polarity is plotted based on this sentimental analysis of tweets.

978-1-7281-4210-4/20/$31.00 ©2020 IEEE

To clean the news articles, Text Normalization is performed. Pre-processing can be done by using Parts of Speech (recognize important words) and Named Entity Recognition (prevent the splitting of tokens like proper nouns).

Named Entity Recognition can be executed after POS. The Textual Features can be retrieved by using the Bag of Words method. Finally, we can classify the dataset by using K Nearest Neighbor Algorithm. Euclidean equation is used to calculate the distance between the neighbors. Here, 'k' is selected as 3. The system is classified with an accuracy of 66.6% [5].

C. *Using SVM and Decision Tree*

Supervised machine learning techniques are used to classify the multi-level multi-class deceptive news [6-9]. In this paper [7], the LIAR dataset was used for classification which includes multi-labels like pants-fire, false, barely true, half-true, mostly-true, and true [6].

The system has mainly three steps: Feature Extraction, Relabeling and Learning. Fig.1 depicts the systematic flow of the model. Manual pre-processing is used to clean the textual data. The textual data should be tokenized and then convert the data to numeric values. Manual codification is used to remove spelling mistakes and abbreviations (example: Ph.D. - Doctor of Philosophy) in the dataset. Features are extracted by using the Weka tool and relabel the multi-class dataset to binary labels (Fake/Real). Then learns the training data and map the textual data to new classes.SVM and Decision Tree algorithms are used to classify the relabelled dataset. In [7], these classification algorithms have attained an accuracy of 66.3 % and 60.4% respectively [10-12].

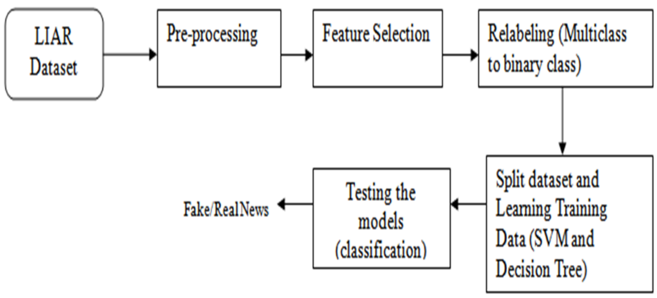

Fig. 1 Multi-label fake news detection using multi-layered supervised learning [7].

D. *Random Forest and Naive Bayes*

Sentimental analysis is incorporated with machine learning algorithms to classify the deceptive news [13]. Three datasets are used to assess the performance of the model. And they are, merged dataset (it is a combination of Kaggle, Politifact and Emergent datasets), George McIntire dataset, and LIAR dataset. The merged dataset was constructed by taking the median of all rows present in three datasets and also some uncommon columns are deleted. The text pre-processing can be done by various methods like bigrams, trigrams, count vectorizer, tf-idf vectorizer, tf-idf vectorizer with cosine similarity etc. These different pre-processing methods are applied to the merged dataset and analyzed it by using Naïve Bayes. The maximum accuracy of 81.6% is attained by using tf-idf vectorizer with cosine similarity. And by using tf-idf vectorizer, it has an accuracy of 80.1%. Additional columns like sentiments, cosine similarity and tf-idf scores are added with merged datasets. Then new merged dataset was trained by using Naive Bayes and Random Forest classifier. Then the model was evaluated and then analyzed by using other two datasets. All datasets are assessed by considering tf-idf vectorizer with or without cosine similarity.

George McIntire dataset was preprocessed with tf-idf cosine vectorizer, and evaluated by Naive Bayes and Random Forest algorithm then obtained an accuracy of 84.3% and 83.9% respectively. Accuracy for George McIntire dataset without cosine similarity, for Naïve Bayes 80.7% and Random forest 79.1%. While applying the Random Forest algorithm in the merged dataset, the model obtained an AUC of 0.60.

For evaluating LIAR dataset, Multinomial Naive Bayes and Random Forest algorithms are used. Several combinations (Text + Content) are considered for this dataset.

While applying different combinations on LIAR dataset with cosine similarity, the maximum accuracy obtained by Naive Bayes Classifier is 18.5% and Random Forest is 35.6%. While using tf-idf without cosine similarity, the accuracy obtained for Naive Bayes is 17.2% and 34.8% for Random Forest [13]. While comparing all accuracies of this model, Random Forest performed better than Naïve Bayes.

E. *Rumour Classification Algorithm*

The news is sometimes related to the writer's views or his/her writing styles. So, fake news also needs to be detected on the basis of above. Rumours/Fake news is usually published with an intention to defame or for any negative impacts. Rumours and deceptive news are related to each other.

In this model, Twitter Fake News Dataset is used to extract the emotional score [14]. The score is found out by using the count of emotional scores of negative words to count of emotional scores of positive words. Then t- statistic value and null hypothesis is determined. No pre-processing methods are used in this model. Most relevant words can be extracted by using any two methods: Latent Semantic Analysis (LSA) [15] and Latent Dirichlet Allocation (LDA) [16].

Using the rumour classifier algorithm, predict the tweets to rumour or non-rumour labels. In rumour classification algorithm LDA is used. LDA is a collection of topics, which splits the topics based on certain probabilities and helps to extract the most relevant words. Append all the extracted word features and create a new feature vector. Then parse all feature vectors to classifier.

For deceptive news or rumour classification, Machine Learning algorithms like SVM, logistic regression, decision trees, random forest, extreme gradient boosting (XG-Boost) and Deep Learning algorithm like Long Short Term Memory (LSTM) recurrent network integrated with Hierarchical

Attention Networks (HAN) are used. HAN contemplates the hierarchical structure of documents (document - sentences - words). It also incorporates an attention mechanism that enables to ascertain the most relevant words and sentences in a corpus while considering the context. Accuracies attained by SVM (86%), Logistic regression (84%), Decision trees (77%), Random Forest (85%), XG-Boost (84%), LSTM-HAN (86%) [14].

F. Using LSTM & GRU

A dataset is created by collecting news articles from different authentic sites. Here, we used (Long Short Term Memory) LSTM and Grated Recurrent units (GRU) [17].

Recurrent neural network (RNN) is a variety of Artificial Neural Network (ANN) in which all nodes are connected sequential manner. In RNN, when the output from each input layer is passed through a huge number of hidden layers it causes a gradient descent problem. The backpropagation in RNN is the main issue for this problem. So it loses its memory values. LSTM and GRU are the units of RNN and these are built to overcome the gradient vanishing or exploding due to long-time RNN. These have cells, input gate, output gate and forget gate.

Before applying to the model, the dataset should be cleaned. For pre-processing the dataset, Natural Language Tool Kit (NLTK) is used. Stop words Removal, Lemmatization and stemming are used for pre-processing. After Pre-processing, text statements are tokenized to word tokens using Keras pre-processing tokenizer. Then the dataset was padded to make each news article of equal length. And then the padded sequences are fed to the neural networks input layer.

To make word embeddings of the datasets, here used FastText. It breaks the words to n-grams and fed it to the neural networks embedding layer. The loss depiction is based on Binary Cross Entropy Loss and it can be optimized by using Adam optimizer. LSTM model has an accuracy of 94.3% and GRU has an accuracy of 91.9% [17].

G. Using Hybrid CNN

A hybrid CNN (convolution neural network (CNN) + Bidirectional LSTM) model is implemented to integrate the metadata with text. The speaker related meta-data are integrated with text to build a deep learning architecture for fake news detection. The training experiments are done by using the LIAR dataset and it has six fine-grained labels [6]. Initialize an embedding matrix, to encode the metadata. To capture the dependencies between the meta-data characteristics, CNN is implemented [6]. A max-pooling operation is performed on the statement embedding matrix, and then a bi-directional LSTM (BiLSTM) layer is used to encode the metadata features. Then concatenate the results from max-pooled text representations of TextCNN with the meta-data representation of the bi-directional LSTM [18-21]. Finally, the results are fed to a fully connected layer with a softmax activation function to generate the final prediction. While combining text with all metadata, the model attains the maximum accuracy of 27.4%.

H. Using Contextual Features with Multi-head Self Attention Mechanism (CMS)

In this paper, bogus news is detected by considering both linguistic and contextual features like job title, historical background of the speaker, etc. The different sequence order of contextual features can make vital influence in system accuracy. This paper ignores the sequential influences of contextual features by implementing the transformer technique like Multi-head Self Attention Mechanism and it can automatically learn features for fake news detection [22].

To implement contextual features with Multi-head Self Attention (CMS), it includes two main components:

- Linguistic Feature Extractor: TextCNN module is used to extract the linguistic characteristics from the LIAR Dataset's Statement [6].

- Context Encoding: To record contextual characteristics (features like speaker, speaker history, speaker job title, party, etc.) Multi-head Self Attention Mechanism is adopted.

Then integrate outputs from both extractors by a fully connected layer with a softmax function to classify news to deceptive or not [22].

Fig. 2 depicts the architectural view of the CMS.

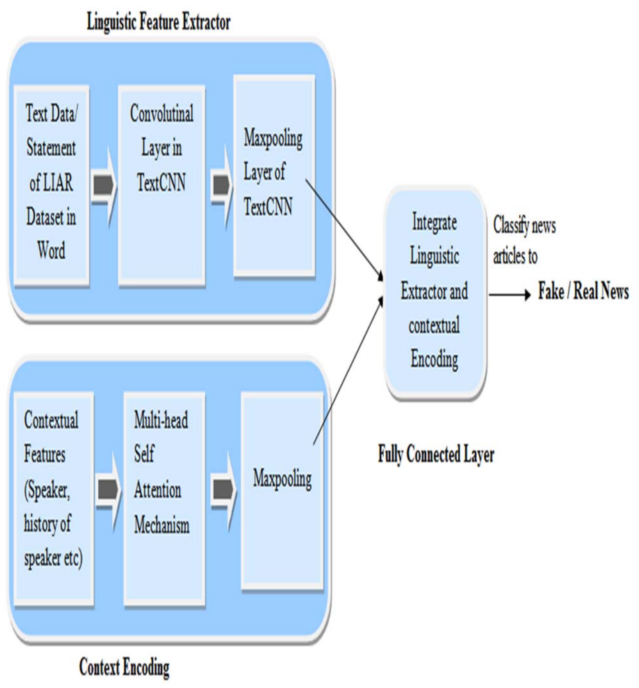

Fig. 2. CMS Architecture: Linguistic Feature Extractor –To extract linguistic characteristics; Context Encoding – To extract contextual features; finally integrate to classify [17].

Clean the text data by pre-processing method. To initialize the embedding layer, 300-dimensions pre-trained word2vec embeddings method is used. The batch size of 64 is used. To optimize the learning rate of the model choose Adam optimizer. For Linguistic Feature Extractor, kernel size of (2, 3,4) and each kernel is adopted with 50 filters. For

Contextual Encoding component, the number of layers of three and heads of two is used.

In [22], linguistics characteristics and contextual features are considered to implement a fake news detection system. In addition to this, CMS ignore the sequential order of these contextual features by using a multi-head self attention mechanism. In comparison to other deep learning models like Hybrid CNN, Text CNN, BiLSTM; the contextual features with multi-head self attention mechanism (CMS) have improved performance. It has attained an accuracy of 45.3%.

III. Analysis

Diverse machine learning and deep learning techniques are used to compare and examine the fake news in online media. In this survey, we have also considered both linguistic characteristics and contextual features. In addition to that we have analyzed the role of contextual features in fake news detection. Table.1. depicts the detailed summary report of the survey.

TABLE I. SURVEY SUMMARY

Method	Dataset Used	Accuracy	Advantages	Disadvantages
Naive Bayes	Facebook news post (BuzzFeed News)	74%	This method is useful for Spam Filtering	Pre-processing is not done.
K-Nearest Neighbor	Tweets on Hilary Clinton	66.6%	Sentimental analysis is used to identify credible or fake tweets.	This paper adopted Meagre datasets and accuracy is not enough.
SVM	LIAR Dataset	66.29%	Feature Selection is done by Weka.	Manual pre-processing is used.
Decision Tree		60.4%		
Random Forest	• George McIntire dataset • LIAR dataset • Merged Dataset	35.6% (LIAR dataset tf-idf with cosine similarity)	Analyses different text pre-processing methods.	Some false statements will negatively affect accuracy while considering the sentiments.
Naive Bayes		18.5% (LIAR dataset tf-idf with cosine similarity)		
SVM	Twitter Fake News Dataset	86%	Sentiment based fake news detection Important features are extracted using LDA and LSA	No Pre-processing Methods Purely Based on sentimental features, it may not always give correct results because sometimes positive statements may have any negative intention too.
Logistic Regression		84%		
Decision Tree		77%		
Random Forest		85%		
XG-Boost		84%		
LSTM-HAN		86%		
Hybrid CNN	LIAR Dataset	27.2%	Used speaker-related metadata with text	Due to overfitting, BiLSTM doesn't perform well.
Multi-head Self Attention Mechanism	LIAR Dataset	45.3%	Considered contextual features and ignored its sequential order.	The accuracy of the model is not up to mark.

The most prevalent dataset in our survey is LIAR Dataset, so we made a comparative study by considering the LIAR dataset alone. The machine learning algorithms used for classifying LIAR dataset in this paper are SVM, Decision Tree, Random Forest and Naïve Bayes. Deep learning models used in this paper are Hybrid CNN and CMS. Only linguistic characteristics are used in these machine learning models but both linguistic and contextual features are considered for classifying the dataset. The accuracies attained by machine learning models by considering only the linguistic features are SVM-66.29%, Decision Tree-60.4%, Random Forest-35.6%, Naïve Bayes- 17.2%. Hybrid CNN and CMS have an accuracy of 27.2% and 45.3% respectively by considering both contextual and linguistic characteristics. By using self attention mechanism in CMS, it ignores the influence of the relative positions of each element in the sequence.

Fig. 3 shows the performance evaluation bar chart; for different machine learning and deep learning techniques with LIAR dataset.

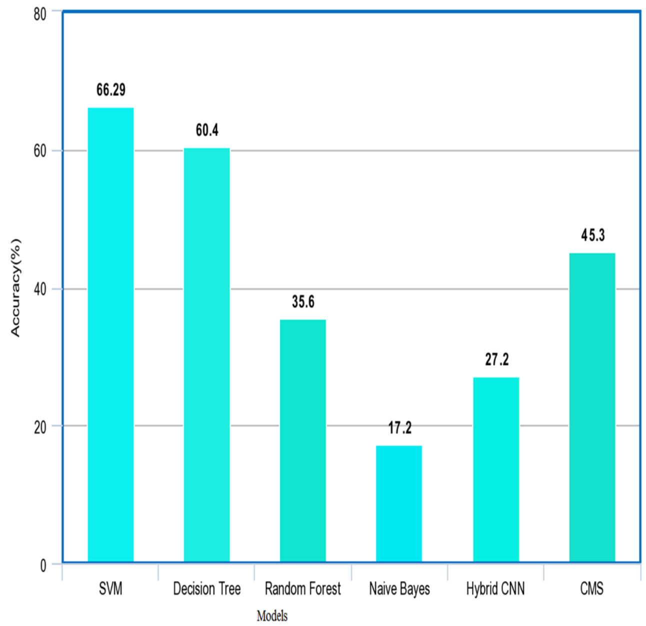

Fig. 3 Performance Evaluation Graph: For different classification algorithm with LIAR Dataset

IV. Conclusion

In this paper, we have correlated and analyzed various machine learning and deep learning models for detecting fake news on social media and other online sites. The Machine Learning algorithms like Naïve Bayes, KNN, SVM, Decision Tree, Random Forest, Logistic Regression and XG-boost are analyzed. Deep learning models like Hybrid CNN, LSTM-HAN and Contextual features with Multi-head Self Attention Mechanism are examined. Different datasets like Facebook news posts, Twitter dataset, LIAR dataset, etc. are used to analyze various models. Most commonly used dataset in our survey was LIAR dataset and we have drawn an evaluation graph with LIAR dataset based on different machine learning and deep learning techniques.

Naive Bayes classifies the Facebook news post with an accuracy of 74%, While KNN classifies the tweets of 'Hillary Clinton' with accuracy of 66.6%, SVM and Decision Tree classifies with an accuracy of 66.3% and 60.4% respectively. Random Forest and Naive Bayes classifies three datasets like George McIntire dataset, LIAR dataset and merged dataset. Different pre-processing techniques are used to clean these datasets. Rumour classification algorithm classifies the Twitter news dataset; and it is based on sentiment score. Then the model classifies the dataset using different machine and deep learning classifiers. LSTM and GRU model eliminates the gradient vanishing problem of RNN. RNN-LSTM and RNN-GRU model attains an accuracy of 94.3% and 91.9% respectively.

The above mentioned models are strictly based on the linguistics characteristics of the news articles. But in Hybrid CNN model, the contextual characteristics are also considered. But the sequential order of contextual features badly affects its accuracy. An accuracy of 27% is attained by the Hybrid CNN model. By applying different sequences of contextual features to the model, the classification accuracy of the system varies. To ignore the order of contextual features, a Multi-head self attention mechanism is implemented. This model includes both linguistic characteristics and contextual features. The system classified the news articles with an accuracy of 45.3%.

So conclude our discussion by recommending that Multi-head self attention mechanism with contextual features enacted well than any other techniques which we have discussed here. Future research can integrate features like contextual characteristics of news articles using transformer techniques, which also incorporate images and videos to complement the performance of fake news detection systems.

References

[1] Allcott H,Gentzkow: Social media and fake news in the 2016 election. J .Econ. Perspect.31(2),211–236(2017).

[2] Mykhailo Granik,Volodymyr Mesyura :Fake News Detection Using Naive Bayes Classifier,2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON).

[3] Craig Silverman, Lauren Strapagiel, Hamza Shaban, Ellie Hall, Jeremy Singer-Vine. (2016, Oct.20).Hyperpartisan Facebook pages are publishing false and misleading information at an alarming rate. [Online]. Available: https://www.buzzfeed.com/craigsilverman

[4] Markines B ,Cattuto C & Menczer F.(2009, April).Social spam detection.In Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web (pp. 41-48).

[5] Amitabha Dey, Rafsan Zani Rafi, Shahriar Hasan Parash, Sauvik Kundu Arko and Amitabha, Fake News Pattern Recognition using Linguistic Analysis, 2018 Joint 7th International Conference on Informatics, Electronics & Vision; 2018 2nd International Conference on Imaging, Vision & Pattern Recognition.

[6] Wang, W.Y: Liar, liar pants on fire: a new benchmark dataset for fake news detection. In the Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Page. 422–426 (2017).

[7] Tayyaba Rasool, Wasi Haider Butt, Arslan Shaukat, M. Usman Akram; Multi-Label Fake News Detection using Multi-layered Supervised Learning, National University of Sciences and Technology Islamabad, Pakistan, ICCAE 2019, February 2325, 2019, Perth, WN, Australia.

[8] Long , Lu Q, Xiang R, Li and Huang; C.R. Fake news detection through multi-perspective speaker profiles; 2017.

[9] Antoniadis S, Litou and Kalogeraki; A model for identifying misinformation in online social networks.Springer,2015.

[10] Kim, J Tabibian B, Oh, A, Sch B, No.246,Gomez-Rodriguez, M. Leveraging the Crowd to Detect and Reduce the Spread of Deceptive News and Misinformation. In Proceedings of the Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (Marina Del Rey, CA, USA, 2018). ACM.

[11] E M Hashem , M S Mabrouk , A study of support vector machine algorithm for liver disease diagnosis, Am. J. Intel. System (2014).

[12] M I Jordan, M Mitchell, Machine learning ; Trends, Perspectives, and prospects Science.

[13] Fake News Detection Using Sentiment Analysis, Bhavika Bhutani ,Neha Rastogi ,Priyanshu Sehgal ,Archana Purwar. IEEE 2019.

[14] Sentiment aware fake news detection on online social networks; Oluwaseun Ajao, Deepayan Bhowmik and Shahrzad Zargari.IEEE 2019

[15] T. Hofmann, “Probabilistic latent semantic analysis,” in Proceedings of the 15th conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 1999, pp. 289–296.

[16] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” Journal for machine Learning research, vol. 3, no. Jan, pp. 993–1022, 2003.

[17] FIND: Fake Information and News Detections using Deep Learning, Abhishek Verma, Vanshika Mittal, Suma Dawn, IEEE 2019.

[18] Ruchansky, N., Seo, S. and Liu, Y. CSIS: A hybrid deep model for fake news detection. ACM, 2017.

[19] Liu, Y., Wu, Y.F.B.: Early detection of fake news on social media through propagation path classification with recurrent and CNN. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)

[20] Xiang Zhang, Junbo Zhao, and Yann LeCun; 2015 Character level sconvolutional networks for text classification. In Advances in neural information processing systems. Page: 649–657.

[21] Sepp Hochreiter and Jurgen Schmidhuber. 1997.LSTM. Neural computation 9(8):1735–1780.

[22] Learning Contextual Features with Multi-head Self-attention for Fake News Detection Yangqian Wang, Hao Han, Ye Ding, Xuan Wang, and Qing Liao International Conference on Cognitive Computing ICCC 2019,Springer.