

## Research Article

# HMBI: A New Hybrid Deep Model Based on Behavior Information for Fake News Detection

Jian Xing <sup>1,2,3</sup> Shupeng Wang <sup>1,2</sup> Xiaoyu Zhang <sup>1,2</sup> and Yu Ding <sup>1,2</sup>

<sup>1</sup>Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

<sup>2</sup>School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

<sup>3</sup>National Computer Network Emergency Response Technical Team/Coordination Center of China Xinjiang Branch, Urumqi, China

Correspondence should be addressed to Jian Xing; [xingjian@iie.ac.cn](mailto:xingjian@iie.ac.cn) and Yu Ding; [dingyu@iie.ac.cn](mailto:dingyu@iie.ac.cn)

Received 24 May 2021; Accepted 10 November 2021; Published 8 December 2021

Academic Editor: Zhaolong Ning

Copyright © 2021 Jian Xing et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fake news can cause widespread and tremendous political and social influence in the real world. The intentional misleading of fake news makes the automatic detection of fake news an important and challenging problem, which has not been well understood at present. Meanwhile, fake news can contain true evidence imitating the true news and present different degrees of falsity, which further aggravates the difficulty of detection. On the other hand, the fake news speaker himself provides rich social behavior information, which provides unprecedented opportunities for advanced fake news detection. In this study, we propose a new hybrid deep model based on behavior information (HMBI), which uses the social behavior information of the speaker to detect fake news more accurately. Specifically, we model news content and social behavior information simultaneously to detect the degrees of falsity of news. The experimental analysis on real-world data shows that the detection accuracy of HMBI is increased by 10.41% on average, which is the highest of the existing model. The detection accuracy of fake news exceeds 50% for the first time.

## 1. Introduction

Due to the timeliness and convenience of social media, people are more likely to consume news from social media than traditional news organizations. Because of these characteristics, large volumes of fake news are produced on social media to obtain political or economic benefits. It deliberately persuades consumers to accept biased or false beliefs. Because of the phenomenon's ubiquity, "fake news" was named word of the year in 2016 [1]. In order to mitigate the negative effects of fake news, automatic detection of fake news on social media becomes increasingly important. Meanwhile, the development of mobile Internet of Things also brings a lot of fake news, the detection of which is also an aspect of mobile internet security [2–4].

However, fake news detection on social media is a difficult and challenging research problem. First, fake news on social media is short and usually related to new times and key events. Since our understanding of the nature of fake news is not enough, the handmade feature of news content

is usually not sufficient [5]. We need more additional information to improve detection, such as the speaker's social engagements [1]. Second, the popularity of social media allows us to gather social background information from the speaker's perspective and their historical data, which provide us with a wealth of background information beyond the news content. For example, a speaker who creates a lot of fake news is likely to create more fake news [6]. Note that few papers existing in the literatures use social behavior information to detect fake news. Capturing useful speaker behavior patterns and extracting corresponding features from them is an open area of research [1]. Third, fake news often mixes true statements with false ones, especially political statements [7]. For example, fake news may cite true evidence in an incorrect context to support untrue claims [1]. Therefore, news has different degrees of falsity, such as false, barely true, half-true, and mostly true. It is important to note that most existing researches define fake news detection on social media as a simple binary classification problem. Considering the degrees of falsity increases the difficulty of fake

news detection [8]. In particular, political fact-checking is a new challenge as it involves a graded notion of truthfulness [9].

This paper studies the problem of multiple classification fake news detection based on behavior information. In particular, our research mainly answers this question: how to effectively use the social behavior information of the speaker to improve the detection accuracy. Through the solution of this problem, we propose a new hybrid deep model based on behavior information for fake news detection, called HMBI.

The key contributions of our work are summarized as follows:

- (1) *Model-oriented*: we propose a new HMBI model, which is the first time to apply Bidirectional Encoder Representations from Transformers (BERT) [10], Transformer [11], and Convolutional Neural Networks (CNN) [12] synthetically for fake news detection. Besides news contents, the proposed model can capture multiple aspect representations of the speaker’s social behavior information for detecting the truthfulness ratings of news
- (2) *Feature-oriented*: we find more effective feature information to identify fake news from the speaker’s social behavior information. Through the combination of BERT and CNN, we extract more useful sentence embeddings of news content and word embeddings of textual social behavior information. Meanwhile, we utilize the speaker’s viewpoint or stance for capturing the speaker’s social engagements and other auxiliary information to learn more useful social behavior features on the Transformer model
- (3) The experimental analysis on real-world data shows that HMBI is more accurate than previous work on fake news detection

The rest of this paper is structured as follows. Section 2 introduces the related work. Section 3 gives a formal definition of fake news detection. Section 4 describes the proposed model in details. Section 5 introduces the dataset, presents the experimental settings, and shows the experimental results. Section 6 concludes with discussion.

## 2. Related Work

Fake news detection in social media is a new research field. In Feature-oriented, the existing researches focus on how to design appropriate models and extract effective features from news content and social context. Usually, we focus on news content are mostly based on language or vision [13, 14]. Social context provides lots of useful auxiliary information for judging the authenticity of news and is the key to the success of false news detection [1]. In Model-oriented, knowledge-based advanced natural language processing models and deep network models [15] have been applied to the classification of news authenticity [16] in recent years.

Meanwhile, the nature of social media provides appended information to improve and optimize the detection model, which is based on news content. The analysis uses the user’s point of view, the relationship between relevant social media, and the relevant user social engagements [17–20]. In Data-oriented, news can be easily collected from different social media platforms. The existing publicly available datasets mainly include BuzzFeedNews [21], LIAR [16], PHEME [22], CREDBANK [23], and FakeNewsNet [24].

The existing work has the following limitations: First, existing fake news detection methods rarely use social context information. This is even more important for short statements (less than 20 words on average), because it has little effect on detecting fake news. Next, most of the existing models are based on CNN and LSTM, lack the research and application of the latest natural language processing models, and fail to study fake news from the perspective of the speaker’s social behavior. Finally, a binary classification fake news detection model is not enough to fully reflect the degree of falsity. Wang [16], MMDF [8], and CMS [25] are closely related to our work and use the deep neural network as an automatic feature extractor. Wang combined CNN and LSTM to construct a fake news detection model. MMDF constructed a unified and explicable multisource fake news detection model. CMS used the transformer technology for the first time in fake news detection. However, the accuracy of the three models was only reached 27.4%, 34.77%, and 45.3%, respectively. This shows that detecting the degrees of falsity of news is still an extremely challenging task.

To address the limitations of existing research, we propose a new hybrid deep model based on behavior information for fake news detection. Specifically, we design different modules according to different characteristics of data types in our proposed model, which greatly improves the accuracy of fake news detection. The average detection accuracy is close to 50%, higher than the state-of-the-art approaches.

## 3. Problem Definition

In this section, we detail the mathematical definition of fake news detection in social media. First, we introduce the mathematical definitions of the main components of news. Second, we give a formal definition of fake news detection by referring to the mathematical definitions given by existing researches. We follow the previous work [1, 8] and define the basic notations as follows.

- (i) Let  $w$  denote the *news content*. It only consists of one major component: speaker’s statement. In general, it is short text on social media
- (ii) Let  $a$  denote the *social behavior information of the speaker*. It contains two main components: speaker and context. Speaker  $s_a$  contains a list of key characteristics that describe the news publisher, such as name, party affiliation, job, credit history, and other related attributes. Context  $c_a$  consists of a series of properties that contain the context information

created with news content and includes subject, state, and location

- (iii) Let  $y = \{y_1, y_2, \dots, y_n\}$  be defined as a label set to represent the degrees of falsity, where  $n$  denotes the number of degrees of falsity that can be identified in our proposed model

*Definition 1.* Given the social behavior information of the speaker  $a$  and corresponding news content  $w$ , the target of fake news detection is to automatically predict  $y$  for unlabeled news content  $w$ , i.e.,  $F : a \& w \rightarrow y$  such that,

$$F(a, w) = y, \quad (1)$$

where  $F$  is the target model to be obtained.

In this paper, fake news detection in social media is defined as a multiple classification problem. The reasons are as follows: fake news usually mixes true statements and it is more practical to predict the degrees of falsity of news than to generate a binary judgment [1].

#### 4. The Proposed Model

In this section, we fully illustrate the proposed Hybrid Deep Model based on Behavior Information (HMBI) for fake news detection. The model mainly consists of two parts, a module for extracting textual data features of news content and textual social behavior information of the speaker and another module for encoding digital social behavior information of the speaker. The architecture of HMBI is shown in Figure 1. Specifically, HMBI consists of the following three parts:

- (1) *Textual Data Feature Extracting:* BERT is used to extract sentence embeddings of news content and word embeddings of textual social behavior information. The above sentence embeddings and word embeddings are concatenated, and then, we use CNN to learn a representation vector  $t$
- (2) *Digital Data Encoding:* we use a fully connected layer and apply Transformer to encode digital social behavior information. Finally, Max-Pooling is applied to output a high-level representation vector  $d$
- (3) *Integration:* we concatenate the output vectors of the two modules to form the final vector  $v$  and then carry out the final classification through a fully connected layer

Next, we introduce the details of each module and then introduce the training procedure of the proposed model.

*4.1. Textual Data Feature Extracting.* In the first module, we take news content and textual social behavior information as the input and then extract their precomputed feature vectors through BERT. We get the feature vectors of dimension 768 and use them as the input of CNN, which contains three

two-dimensional convolutional layers and a maximum pooling layer. The detailed process is

$$P = \Gamma(f \odot \text{Concat}(W, T_a)), \quad (2)$$

$$t = \text{Max - pooling}(P), \quad (3)$$

where  $W$  is a matrix of sentence embeddings representing news content and  $T_a$  is a matrix of word embeddings representing textual social behavior information.  $\Gamma$  is an activation function,  $f$  is a filter, and  $P$  is the corresponding feature map after convolution.

*4.2. Digital Data Encoding.* In the second module, we use linear projection to map digital social behavior information to 300 dimensions and then follow the overall architecture of the Transformer, using a stack encoder that consists of two identical layers. Each layer has a multihead self-attention mechanism and a position-type fully connected feedforward network. At each layer, there is a multihead self-attention mechanism followed by a position-wise fully connected feedforward network. The process of multihead self-attention mechanism with digital social behavior information is [11]

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (4)$$

$$\text{MultiHead}(D_a) = \text{Concat}(\text{Attention}_1, \dots, \text{Attention}_h). \quad (5)$$

$D_a$  is a matrix representing digital social behavior information and is given by linear projection.  $Q$ ,  $K$ , and  $V$  are its different subspace matrices.  $d_k$  represents the  $Q$  vector dimension, and  $h$  represents the number of heads in the multihead self-attention mechanism. Afterwards, we take the output vector of the last encoder and perform a max-pooling operation on it. Finally, we get the final representation of digital social behavior information. The process of max-pooling operation is

$$d = \text{Max - pooling}(r_1, r_2, \dots, r_j), \quad (6)$$

where  $r_i$  is the  $i$ -th column of output matrix of the last encoder.

*4.3. Integration.* Finally, we propose a module that concatenates  $t$  and  $d$  to get the final feature representation  $v$ . Then,  $v$  is sent to a fully connected layer to better differentiate the degrees of falsity of news.

$$Y = \text{softmax}(W^T v + b), \quad (7)$$

where  $Y \in \mathbb{R}^n$  and  $n$  is the number of classes.

*4.4. Training Process.* In this subsection, we describe the training process of HMBI (see Algorithm 1 for the details). In each iteration of the algorithm, the sentence embeddings  $W$  and word embeddings  $T_a$  are extracted by the pretrained

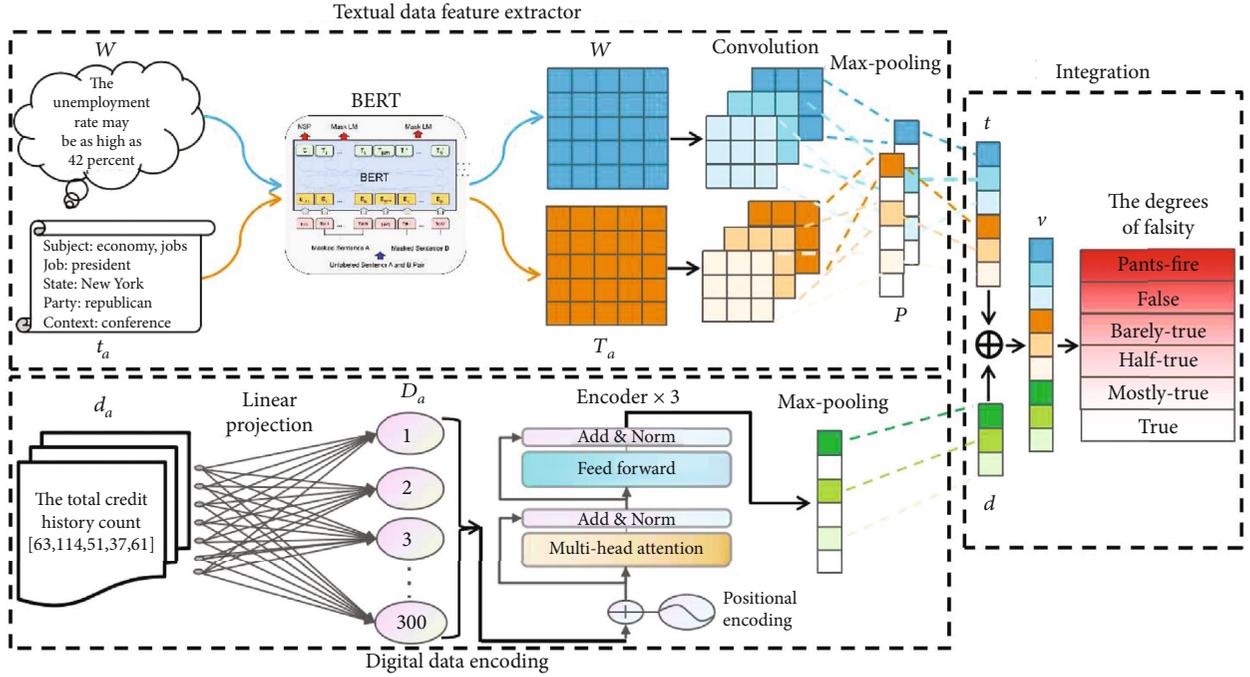


FIGURE 1: An illustration of HMBI.

**Input:** news content  $w$  and social behavior information  $a$ , and labels  $y = \{y_1, y_2, \dots, y_n\}$ .

**Output:** target model  $F$ .

```

1 for number of epoch do
2  $w \rightarrow W$  through BERT pretrained model;
3  $t_a \rightarrow T_a$  through BERT pretrained model;
4 compute  $P$  according to Equation (2):  $P = \Gamma(f \odot \text{Concat}(W, T_a))$ ;
5 compute  $t$  according to Equation (3):  $t = \text{Max-pooling}(P)$ ;
6  $d_a \rightarrow D_a$  linear projection;
7 for each encoder do
8 compute  $Q, K, V$  according to Equations (4) and (5):
9  $\text{Attention}(Q, K, V) = \text{softmax}(QK^T / \sqrt{d_k})V$ ;
10  $\text{MultiHead}(D_a) = \text{Concat}(\text{Attention}_1, \dots, \text{Attention}_h)$ ;
11 end
12 compute  $d$  according to Equation (6):  $d = \text{Max-pooling}(r_1, r_2, \dots, r_j)$ ;
13 integrate  $t$  and  $d$  into  $v$ :  $v = t \oplus d$ ;
14 compute  $Y$  according to Equation (7):  $Y = \text{softmax}(W^T v + b)$ ;
15 end
16 if the training converges then
17 Calculate - Centers( $a \& w : F$ );
18 end
19 return  $F$ ;

```

ALGORITHM 1: Training procedure of HMBI.

model of BERT (lines 2 and 3). We concatenate them and compute representation vector  $t$  through convolutional layers and max-pooling layer (Equations (2) and (3)). After that,  $D_a$  is given by linear projection (line 6). We compute representation vector  $d$  through encoders (Equations (4), (5), and (6)). Representation vectors  $t$  and  $d$  are concatenated to build the final feature representation  $v$  (line 13). We compute the final prediction  $Y$  through the fully

connected layer (Equation (7)). Finally, once the training converges, the prediction function  $F$  is returned, which can be used for prediction (line 17).

## 5. Experiments

In this section, we evaluate HMBI on a real-world fake news dataset. We compare the detection accuracy produced by

HMBI with a set of representative baselines. First, we introduce the real-world dataset LIAR. Second, we present the experimental settings. Finally, we present experimental results to demonstrate the effectiveness of HMBI.

*5.1. Dataset.* To make a fair comparison, we run a series of comparative experiments on the real-world dataset LIAR, which was published by Wang. It is one of the existing publicly available benchmark datasets used in previous work. It is from politifact.com. It consists of 12,836 manually marked short political statements. These statements come from various contexts, which include TV interviews, campaign speeches, and news releases. There are six types of news labels: pants-fire, false, barely-true, half-true, mostly-true, and true. The dataset contains three subsets: train, valid, and test. They account for 80%, 10%, and 10% of the entire dataset, respectively [8]. LIAR contains two parts of data. The data are described in detail as follows.

- (i) *News content.* It includes only one data field: statement. The statement is mainly short sentences from American politicians, covering a variety of different topics, including military, medical, and biological. We model the statement using the architecture described in Section 4.1
- (ii) *Social behavior information of the speaker.* It consists of two components: speaker and context. Speaker includes four data fields: name, party affiliation, job, and credit history. Context includes three data fields: subject, state, and location. It is the speaker's personal information and textual context information about the statement. We combine all these data fields (except credit history) and model it using the architecture described in Section 4.1. The credit history is a five-dimensional digital vector. It represents the number of statements the speaker has on the five classes: pants-fire, false, barely-true, half-true, and mostly-true. We use the architecture presented in Section 4.2 to model it

*5.2. Experimental Settings.* We use BERT pretrained model uncased\_L-12\_H-768\_A-12 (<https://github.com/google-research/bert>) to initialize sentence embeddings and word embeddings and take the second last layers as the final output. The batch size is 8. The dropout rate is 50%. The learning rate is 0.0001. The optimizer selects the Adam optimizer [26]. The epoch parameter is set to 16. In Textual Data Feature Extracting, the kernel size is (2, 3, 4). In Digital Data Encoding, the number of layers is 3 and the number of heads is 2. We use the average accuracy of 10 trials as the measurement accuracy.

*5.3. Experimental Results.* We experimentally compared HMBI with five representative baselines:

*Wang [16].* It is a fake news detection model based on CNN and bi-directional LSTM. It combines metadata with text and is a new hybrid convolutional neural network. This hybrid approach can improve a text-only deep learning model.

*MMFD [8].* It is a unified and explicable fake news detection model. It has the characteristics of automatic feature extraction, multisource fusion, and automatic degrees of falsity detection.

*Long [27].* It is a fake news detection model based on LSTM. It proves that the speaker profiles provide valuable information to validate the credibility of news articles.

*BERT [10].* It is a new language representation model. The full name is Bidirectional Encoder Representations from Transformers. The pretrained BERT model can create the most advanced model for eleven tasks.

*CMS [25].* It is a fake news detection model based on multihead self-attention for extracting features from context information. It can automatically capture dependencies between context information and learn global representation from context information.

The model deployment of the five representative baselines is detailed in the original. The experimental parameters are all in accordance with the original.

The comparison results on LIAR are shown in Table 1. Compared with five representative baselines, the accuracy of HMBI is increased by 10.41% on average. Compared with CMS, which performs best in five representative baselines, the accuracy of HMBI is increased by 3.38%. Both HMBI and the other five representative baselines use social behavior information as a feature, but obviously, HMBI performs better. The normalized confusion matrix is shown in Figure 2. It shows the results of HMBI. The detection accuracies of "mostly-true," "barely-true," "false," and "pants-fire" are all more than 50%. The detection accuracy of "pants-fire" is the highest, which is 66%. However, the detection accuracy of "true" was relatively low, only 22%. The reasons are as follows: First, the number of "true" samples in the dataset is relatively small, which is not conducive to feature learning. Second, the difference between "true" and "mostly-true" is very small and more likely to be misjudged by each other. The matrix also shows that 29% of "true" is misjudged as "mostly-true." When only using a single statement for detection, the detection accuracy of all models is less than 30%. This shows that mixed speaker behavior features are more effective than single news content features. In brief, HMBI outperforms five representative baselines, and it is better at detecting fake news.

The reasons for the better performance of HMBI are as follows: First, the BERT pretrained model can better extract sentence embeddings of news content and word embeddings of textual social behavior information. Second, using Transformer, we can make better use of digital social behavior information. Third, beyond news content, social behavior information of the speaker can also be used to improve the accuracy of fake news detection [28].

From the above experiments, we can see that the detection accuracy of multiple classification fake news is still relatively low, less than 50% on average. Incompleteness of LIAR is one of the reasons. For example, the missing rate of data field "job" and "state" are 27.89% and 21.62%, respectively, and there are only five categories in the data field "credit history," which does not cover all the categories of news. But, quite apart from that, there is still much room

TABLE 1: Performance comparison.

Method	Accuracy (%)	Accuracy (%) (statement only)
Wang	27.40	27.00
MMFD	34.77	29.06
Long	41.50	25.50
BERT	42.71	27.89
CMS	45.30	22.40
HBMI	48.68	27.59

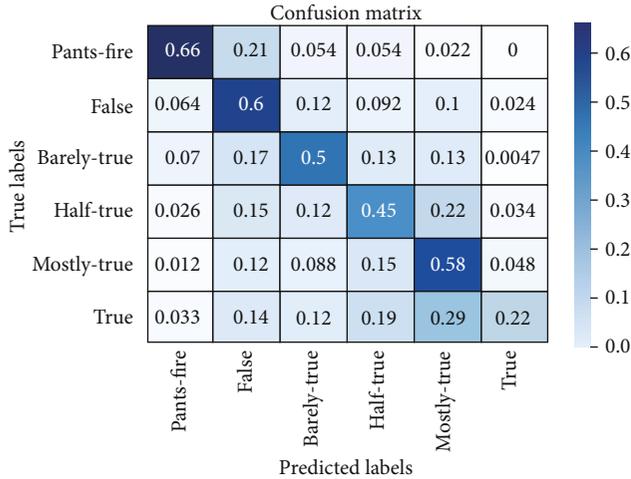


FIGURE 2: The confusion matrix of HBMI on LIAR.

for improvement in the feature extraction and optimization model.

**5.4. Repeatability.** The experimental equipment is configured as 128GB memory and a GeForce RTX 2080 GPU. The source code is available at <https://github.com/xingjian215/HBMI>.

## 6. Conclusion

In this paper, we present a new hybrid deep model to automatically learn more useful features from news content and social behavior information. Inspired by the latest natural language processing technology, we apply BERT, Transformer, and CNN synthetically to improve the detection accuracy of fake news. Experimental results further demonstrate that HBMI can more accurately detect the degrees of falsity of news.

## Data Availability

Our source code is available at <https://github.com/xingjian215/HBMI>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61931019 and grant no. U1803263).

## References

- [1] K. Shu, S. Amy, W. Suhang, T. Jiliang, and L. Huan, "Fake news detection on social media: a data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017.
- [2] S. Li, D. Zhao, X. Wu, Z. Tian, A. Li, and Z. Wang, "Functional immunization of networks based on message passing," *Applied Mathematics and Computation*, vol. 366, article 124728, 2020.
- [3] S. Li, L. Jiang, X. Wu, W. Han, D. Zhao, and Z. Wang, "A weighted network community detection algorithm based on deep learning," *Applied Mathematics and Computation*, vol. 401, article 126012, 2021.
- [4] P. Zhu, X. Wang, D. Jia, Y. Guo, S. Li, and C. Chu, "Investigating the co-evolution of node reputation and edge-strategy in prisoner's dilemma game," *Applied Mathematics and Computation*, vol. 386, article 125474, 2020.
- [5] N. Ruchansky, S. Seo, and L. Yan, "CSI: a hybrid deep model for fake news detection," in *CIKM '17: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Singapore Singapore, 2017.
- [6] M. del Vicario, A. Bessi, F. Zollo et al., "The spreading of misinformation online," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 3, pp. 554–559, 2016.
- [7] R. Oshikawa, J. Qian, and W. Y. Wang, "A survey on natural language processing for fake news detection," 2018, <https://arxiv.org/abs/1811.00770>.
- [8] H. Karimi, P. Roy, S. Saba-Sadiya, and J. Tang, "Multi-source multi-class fake news detection," in *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA, 2018.
- [9] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: analyzing language in fake news and political fact-checking," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark, 2017.
- [10] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," 2018, <https://arxiv.org/abs/1810.04805>.
- [11] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," 2017, <https://arxiv.org/abs/1706.03762>.
- [12] C. Szegedy, W. Liu, Y. Jia et al., *Going deeper with convolutions*, IEEE Computer Society, 2014.
- [13] S. Afroz, M. Brennan, and R. Greenstadt, *Detecting hoaxes, frauds, and deception in writing style online*, IEEE Computer Society, 2012.
- [14] 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*, vol. 19, no. 3, pp. 598–608, 2017.
- [15] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju Island, Korea, 2012.

- [16] W. Y. Wang, "“Liar, liar pants on fire”: a new benchmark dataset for fake news detection," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, 2017.
- [17] S. M. Mohammad, P. Sobhani, and S. Kiritchenko, "Stance and sentiment in tweets," *ACM Transactions on Internet Technology*, vol. 17, no. 3, pp. 1–23, 2017.
- [18] Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News verification by exploiting conflicting social viewpoints in microblogs," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Hoenix, Arizona, USA, 2016.
- [19] M. Gupta, P. Zhao, and J. Han, "Evaluating event credibility on twitter," in *Proceedings of the 2012 SIAM International Conference on Data Mining (SDM)*, Anaheim, CA, 2012.
- [20] Z. Jin, J. Cao, Y.-G. Jiang, and Y. Zhang, "News credibility evaluation on microblog with a hierarchical propagation model," in *2014 IEEE International Conference on Data Mining*, Shenzhen, China, 2014.
- [21] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," 2017, <https://arxiv.org/abs/1702.05638>.
- [22] A. Zubiaga, M. Liakata, R. Procter, G. Wong Sak Hoi, and P. Tolmie, "Analysing how people orient to and spread rumours in social media by looking at conversational threads," *PLoS One*, vol. 11, no. 3, p. e0150989, 2016.
- [23] T. Mitra and E. Gilbert, "Credbank: a large-scale social media corpus with associated credibility annotations," in *Ninth international AAAI conference on web and social media*, Oxford, UK, 2015.
- [24] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fake-NewsNet: a data repository with news content, social context and dynamic information for studying fake news on social media," 2018, <https://arxiv.org/abs/1809.01286>.
- [25] Y. Wang, H. Han, Y. Ding, X. Wang, and Q. Liao, "Learning contextual features with multi-head self-attention for fake news detection," in *Cognitive Computing – ICC3 2019*, Springer, Cham, 2019.
- [26] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *3rd International Conference on Learning Representations*, San Diego, 2015.
- [27] Y. Long, "Fake news detection through multi-perspective speaker profiles," in *Proceedings of the Eighth International Joint Conference on Natural Language Processing*, Taipei, Taiwan, 2017.
- [28] K. Shu, S. Wang, and H. Liu, "Beyond news contents: the role of social context for fake news detection," in *Proceedings of 12th ACM International Conference on Web Search and Data Mining*, Melbourne VIC Australia, 2017.