Fake News Detection Method Based on Text-Features

Ahlem Drif

Networks and Distributed Systems Laboratory
Faculty of Sciences
University of Sétif 1
Sétif, Algeria
Email: adrif@univ.setif.dz

Zineb Ferhat Hamida

Computer Science Department
University of Sétif 1
Sétif, Algeria
Email: zineb.ferhat@yahoo.com

Silvia Giordano

Networking Lab, SUPSI
University of Applied Sciences of Southern Switzerland
Lugano, Switzerland
Email: silvia.giordano@supsi.ch

***Abstract*—Feature extraction is a critical task in fake news detection. Embedding techniques, such as word embedding and deep neural networks, are attracting much attention for textual feature extraction, and have the potential to learn better representations. In this paper, we propose a joint Convolutional Neural Network model (CNN) and a Long Short Term Memory (LSTM) recurrent neural network architecture, taking advantage of the coarse-grained local features generated by CNN and long-distance dependencies learned via LSTM. An empirical evaluation of our model shows good prediction accuracy of fake news detection, when compared to Support Vector Machine and CNN baselines.**

Keywords–Fake news detection; social networks; deep learning; convolutional neural network; text classification; words embedding technique.

I. INTRODUCTION

Social media have pushed the ability to exchange information at a much greater pace, to a far wider audience than ever before. This information is not always truthful. Because anyone can publish anything on the Internet, the information obtained from this source can be inaccurate or even intentionally false (fake news) [1]. The term ”fake news” became mainstream during the 2016 US presidential election campaign when hundreds of websites published falsified or heavily biased stories - many of them in the pursuit of capitalising on social media advertising revenue. The fake news term, popularised by the US President Donald Trump, is so prevalent now that it is hard to believe that just a few years ago the term was barely used. Besides, there are a variety of reasons for fake news and misinformation growing in levels, and rising in importance. These include how easy it is now to set up a website or even to manipulate a webpage to include the information desired, as well as how suited social media is for fake news broadcasting, combined with the rise of online social media. This work presents a comprehensive study on the features of different fake news datasets. To this extent, we implement methods based on both recent deep learning methods and machine learning methods to effectively detect fake news based on text features.

The rest of the paper is organized as follows: related literature survey is given in Section 2, Section 3 regroups fake news characteristics across different dimensions and summarizes some datasets features, the details of the proposed framework are introduced in Section 4, experimental results are presented in Section 5, and the study is concluded in Section 6.

II. RELATED WORK

The problem of fake news detection is more challenging than detecting deceptive reviews, since the political language on TV interviews, posts on Facebook and Twitters consists mostly short statements. The dissemination of fake news may cause large-scale negative effects, and sometimes can affect or even manipulate important public events. For example, within the final three months of the 2016 US presidential election, the fake news generated to favor either of the two nominees was believed by many people and was shared by more than 37 million times on Facebook [1]. There has been a large body of work surrounding features analysis of fake news. For example, Jin et al. [2] analyzed news articles' images for fake news detection based on multimedia datasets. They explored various visual and statistical image features to predict respective articles' veracity. Moreover, they proposed a fake news detection method utilizing the credibility propagation network built by exploiting conflicting viewpoints extracted from tweets. Yang et al. [3] proposed an efficient model for early detection of fake news through classifying news propagation paths using a multivariate time series. They realized a new deep learning model, which was comprised of four major components, i.e., propagation path construction and transformation, Recurrent Neural Network (RNN) based propagation path representation, CNN-based propagation path representation, and propagation path classification, which were integrated together to detect fake news at the early stage of its propagation.

 ISBN: 978-1-61208-731-3

Fake news detection based on surface-level linguistic patterns is also a popular trend in this area, such as building classifiers to detect whether tweets are factual or not. Ruchansky et al. [4] proposed an architecture of three components; the first module is a recurrent neural network to capture the temporal pattern of user activity on articles, and, the second module learns the source characteristic based on the behavior of users, and the two were integrated with the third module to classify an article as fake or not. Wang et al. [5] proposed an Event Adversarial Neural Network (EANN), which consists of three main components: the multi-modal feature extractor, the fake news detector, and the event discriminator. The multi-modal feature extractor is responsible for extracting the textual and visual features from posts. It cooperates with the fake news detector to learn the discriminable representation for the detection of fake news. Hardalov et al. [6] used a combination of linguistic, credibility and semantic features to differentiate between real and fake news. In their work, linguistic features include (weighted) n-grams and normalized number of unique words per article. Credibility features include capitalization, punctuation, pronoun usage and sentiment polarity features generated from lexicons. Text semantics were analyzed using embedding vectors method. All feature categories were tested independently and in combination based on self-created datasets. The best performance was achieved using all available features. In addition, Ma et al. [7] observed changes in linguistic properties of messages over the lifetime of a rumor using Support Vector Machine (SVM) based on time series features, then, they showed good results in the early detection of an emerging rumor. Moreover, Conroy et al. [8] illustrated that the best results for fake news detection could be achieved while combining linguistic and network features. Ciampaglia et al. [9] proposed to map the fact-checking task to the well-known task of finding the shortest path in a graph in order to utilize the information provided by knowledge networks. In that case, a shortest path indicates a higher probability of a truthful statement. Wang [10] [11] designed a hybrid Convolutional Neural Network (CNN or ConvNet) to integrate metadata with text. The best performance was achieved when incorporating different metadata features. Lendavi and Reichel [12] investigated contradictions in rumors sequences of micro-posts by analyzing posts at the text similarity level. The authors argue that vocabulary and token sequence overlap scores can be used to generate cues to veracity assessment, even for short and noisy texts. Joulin et al. [13] proposed a text classification model based on n-gram features, dimensionality reduction, and a fast approximation of the softmax classifier. This fast text classifier is built upon a product quantization method in order to minimizes the softmax loss l over N documents, therfore, it gives accurate results with less training and evaluation time [14]. For a full review of the state of the art in fake news detection in social media, see Zhou et al. [15].

In this work, we aim at building a new solution for addressing the detection of fake news based on the textual content of the news. For this reason, we realize a joint CNN-LSTM model, which can be defined by adding CNN layers in the front, followed by Long Short Term Memory (LSTM) layers with a dense layer on the output. Indeed, when analyzing fake news with such combination, the CNN acts like a trainable feature detector for the fake news content. It learns powerful convolutional features, which operate on a static spatial input. The LSTM, instead, receives a sequence of such high-level representations and generates a description of the content or maps it to some static class of outputs. We show that this combined approach works better than baselines approaches.

III. EXPLORING FEATURES EXTRACTION

The propagation of false information on social media is related to several factors, such as the information content and the users' behaviors. In this Section, in order to build a deep learning model that extracts discriminative characteristics of fake news, we study the most relevant attributes at the content level, user level, and social level [16]. Below, we elaborate on each level.

A. Content level

In order to capture the different aspects of fake news and real news, existing work relies on news content. Basically, the useful features that mostly are extracted from news content are linguistic-based and visual-based. Different kinds of linguistic features can be built: (i) lexical features, including character level and word-level features, such as total words, characters per word, frequency of large words, and unique words; (ii) syntactic features, including sentence level features, such as frequency of function words and phrases, i.e., n-grams and bag of words approaches [17], or Punctuation and Parts of Speech (POS) tagging. Also, visual-based characteristics have been shown to be an important manipulator for fake news propaganda [18]. As we have characterized, fake news exploits the individual vulnerabilities of people and thus often relies on sensational or even fake images (or fake videos) to provoke anger or other emotional response in the consumers.

B. User level

User based features represent the characteristics of those users who have interactions with the news on social media. These user level features are extracted to infer the credibility and reliability of each user using various aspects of user demographics, such as: registration age, number of followers and followees, number of tweets the user has authored, etc. [19]. Further, the users engagement in news dissemination process ranges from users response to a post up to spreading news pieces. Several works have observed that there are major psychological and cognitive factors that heavily increase the user engagement to fake news spreading:

- naive realism: consumers tend to believe that their perception of reality is the only accurate view, while others who disagree are regarded as uninformed, irrational, or biased [20].
- confirmation bias: consumers prefer to receive information that confirms their existing views [21].

Prospect theory describes decision making as a process by which people make choices based on the relative gains and losses as compared to their current state. This desire of maximizing the reward of a decision, to have social gains, can be modeled from an economic game theoretical perspective [22] by formulating the news generation and consumption cycle as a two-player strategy game. In fact, there are two kinds of key players in the information ecosystem: publisher and consumer. The utility for the publisher stems from two perspectives:

- short-term utility: the publisher's profit which is positively correlated with the number of consumers reached.
- long-term utility: the publisher's reputation in terms of news authenticity.

The utility for consumers consists of two parts:

- information utility: obtaining true and unbiased information.
- psychology utility: receiving news that satisfies their prior opinions and social needs, e.g., confirmation bias and prospect theory.

Both publisher and consumer try to maximize their overall utilities in the strategy game that is the news consumption process.

C. Social level

Social dimensions refer to the heterogeneity and weak dependency of social connections within different social communities. Users' perceptions of fake news pieces are highly affected by their like minded friends on social media, while the degree differs along different social dimensions. Thus, it is worth exploring why and how different social dimensions play a role in spreading fake news. Recent findings [23] showed that users on Facebook tend to select information that adhere to their system of beliefs and to form polarized groups, i.e., echo chambers. For example, users on Facebook always follow like-minded people and thus receive news that promote their favored existing narratives. The echo chamber effect facilitates the process by which people consume and believe fake news due to the following psychological factors:

- Social credibility, which means that people are more likely to perceive a source as credible if others perceive the source as credible, especially when there is not enough information available to access the truthfulness of the source.
- Frequency heuristic, which means that consumers may naturally favor information they hear frequently, even if it is fake news. Del Vicario et al. [24] showed that social homogeneity is the primary driver of content diffusion, and one frequent result is the formation of homogeneous, polarized clusters. Most of the time, the information is taken by a friend having the same profile (polarization), i.e., belonging to the same echo-chamber.

In Table I, we categorize the methods discussed in Section II, based on the features of the fake information analyzed. The majority of fake news detection algorithms are content feature based, in that they rely on developing efficient features of news content that individually or jointly are able to distinguish between real and fake information.

IV. Model Construction

In this work, we combined a CNN and a LSTM, which is a type of Recurrent Neural Network. Figure 1 shows the overview of our combined CNN-LSTM neural network for fake news detection. In fact, there are many interesting properties that one can get from combining convolutional neural networks and LSTM network, as we will discuss later in this work.

We build our CNN-LSTM deep neural networks model as follows: the embedding layer is the first layer in the model and it represents each statement (text) as a row of vectors. Each vector represents a token based on the word-level used. Each word in the statement, which is one token in the word level, is embedded into a vector with length of 300. This layer is a matrix of size $w \times v$, where v is the length of the vector and w is the number of tokens in the statements. The value of w is the maximum length of a statement. Any statement that contains less than the maximum number of tokens in the statement will be padded with $<$ Pad$>$ to have the same length as the maximum statement length. Each matrix in the word level has the size of 50 $\times$ 300. The vocabulary size is 10,000. We used a pre-trained 300-dimensional Google News Vectors method (GloVe) [25] of learning word embeddings from text. After that, we added a drop-out layers [26] to reduce overfitting and set the drop-out probability to 0.2 when training. The output is fed to the next layer. Then, we added a Convolutional Neural Networks layer that extract features from local input patches. We used 10 filters with size 3 to extract features of words from statement. Each filter detects multiple features in the text using ReLu [27] activation function in order to represent them in the feature map. Then, a standard max pooling operation is performed on the latent space, followed by a LSTM layer. The forward one-dimensional (1D) max pooling layer is a form of non-linear down sampling of an input tensor $X \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_p}$. 1D max pooling partitions the input tensor data into 1D subtensors along the dimension k, selects an element with the maximal numeric value in each subtensor, and transforms the input tensor to the output tensor Y by replacing each subtensor with its maximum element. The max operation or function is the most commonly used technique for this layer and it is used in our experiments. The reason for selecting the highest value is to capture the most important feature and reduce the computation in the advanced layers. The LSTM layer has a set number of units and the input of each cell is the output from the previous max pooling layer. In fact, the output vectors of the max pooling layer become inputs to the LSTM networks to measure the long-term dependencies of feature sequences. The outputs from LSTMs are merged and then passed to a fully connected layer. We need the expressive power of two fully connected layers. The last dense layer converts the array into a single output in the range $\{0, 1\}$. Thus, the sigmoid function is used [28].

For comparison, we used two baselines: a Support Vector Machine classifier (SVM) [29], and a Convolutional Neural Network model [30]. For SVM, we used Scikit-learn library which provides very strong performances on short text classification problems. For CNN, we used TensorFlow [31] for the implementation. The CNN baseline model is obtained as follows: we performed unsupervised learning of word-level embeddings. Then, we added a drop-out layer with probability equal to 0.2. The output of the drop-out layer is fed to a convolution layer ConvNet1D with 10 filters with size 3 and the activation function ReLu. Then, a standard max pooling operation is performed followed by a 1D global average pooling layer. The average pooling layer output is passed to a fully connected layer followed by a drop-out layer with 0.6 drop-out probability. We added a fully connected layer to trade network-depth for increasing the chances to get a

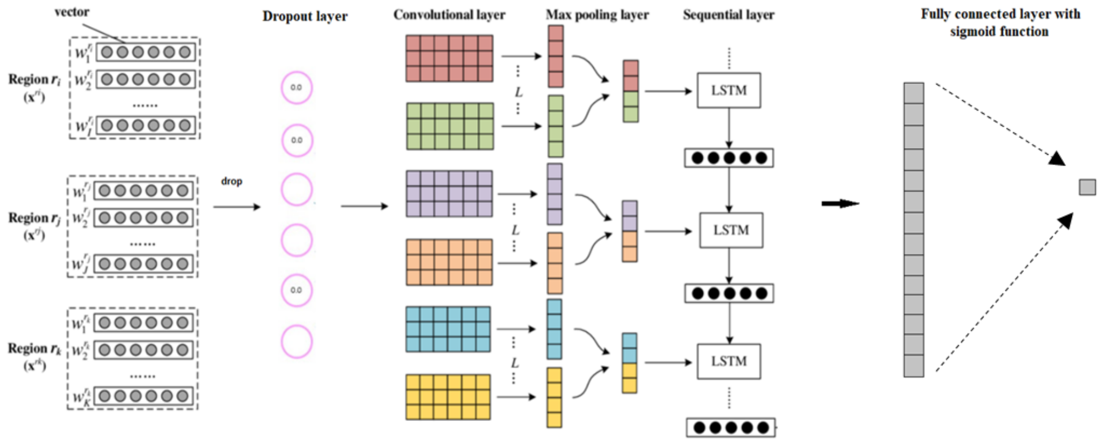


Figure 1. The architecture of the proposed fake news detection model

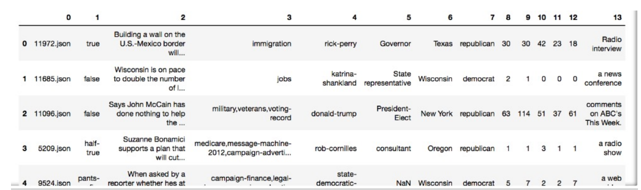
better learned layer. The sigmoid activation function is used to generate the final classification. For this CNN baseline, we have added two drop-out layer in order to improve the accuracy. This choice was made empirically.

V. EXPERIMENTAL SETTINGS AND RESULTS

A. Dataset pre-processing

To fairly evaluate the performance of the proposed model, we conduct the experiments on two real social media datasets: Liar dataset [10] and News Articles dataset [32]. These two datasets contain a rich metadata that would help to descriminate text-features.

The Liar dataset [10] is collected from the fact-checking website PolitiFact through its API [33]. The website PolitiFact.COM focused on looking at specific statements made by politicians and rating them for accuracy. The Liar dataset includes a rich set of metadata for each speaker: statement, party affiliations, current job, home state, as well as historical counts of inaccurate statements. These various metadata can be granular enough to define features at the content level. The Liar dataset comprises 12,836 short statements labeled for truthfulness, subject, context/venue, speaker, state, party, and prior history, as illustrated in Figure 2. This dataset considers six fine-grained labels for the truthfulness ratings: pants-fire, false, barely-true, half-true, mostly-true, and true. In our work, we analyze the correlation between these labels. Figure 3 shows that, from the perspective of the description space, some labels might well be just correlated noise. For this reason, we merge the mostly-true and the half-true labels into the true label, and merge the barely-true and the pants-fire labels into the false label. The association between labels may give birth to a better classification standard.



	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	11972.json	true	Building a wall on the U.S.-Mexico border will...	immigration	rick-perry	Governor	Texas	republican	30	30	42	23	18	Radio interview
1	11685.json	false	Wisconsin is on pace to double the number of l...	jobs	katrina-shankland	State representative	Wisconsin	democrat	2	1	0	0	0	a news conference
2	11096.json	false	Says John McCain has done nothing to help the ...	military,veterans,voting-record	donald-trump	President-Elect	New York	republican	63	114	51	37	61	comments on ABC's This Week.
3	5209.json	half-true	Suzanne Bonamici supports a plan that will cut...	medicare,message-machine-2012,campaign-adverti...	rob-cornilles	consultant	Oregon	republican	1	1	3	1	1	a radio show
4	9524.json	pants-	When asked by a reporter whether hes at	campaign-finance,legal-	state-democratic-	NaN	Wisconsin	democrat	5	7	2	2	7	a web

Figure 2. Liar dataset attributes

The News Articles dataset [32] comprises 20,800 stories labeled as unreliable or reliable, as shown in Figure 4. The News Articles dataset contains text, author, and title.

In the pre-processing phase, we have dropped the rows with missing values. Also, we have removed from each text the punctuations marks and the stop-words, which represent the most common words in a language, such as ”are”, ”as”, ”the”, etc. In addition, we have applied a stemming process to cut off the end or the beginning of the word, taking into account a list of common prefixes and suffixes that can be found in an inflected word. For example, for News Articles dataset , after

TABLE I. COMPARISON OF FEATURES-BASED FAKE NEWS DETECTION METHODS.

Methods	Content level		User level		Social level	
	Linguistic	Visual	User profile	Credibility features	Diffusion network	Freindship network
Ma et al. (2015) [7]	✓					
Conory et al. (2015) [8]	✓				✓	
Ciampaglia et al. (2015)[9]	✓				✓	
Lendavi et al. (2016)[12]	✓					
Hardalov et al. (2016)[6]	✓		✓			
Julian et al. (2016)[13]	✓					
Wang 2016 [5]	✓					
Jin et al. (2017)[2]		✓		✓		
Ruchansky et al. (2017)	✓		✓			
Wang et al. (2018)	✓	✓				
Yang et al. (2018)			✓	✓	✓	

 ISBN: 978-1-61208-731-3

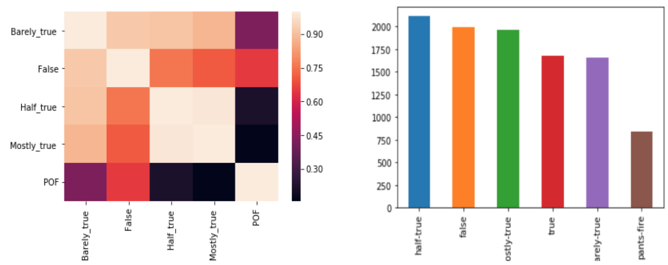


Figure 3. a) Correlation betwen labels of Liar dataset b) The Liar dataset labels

dropping the rows with missing labels or with an empty text, we have obtained 7,924 real labels and 10,361 fake labels. After that, we have applied a tokenization technique which is the process of splitting the given text into smaller pieces called tokens (words, numbers and others can be considered as tokens). Finally, we have created sequences with a vocabulary size of 10,000 for the Liar datset and 50,000 for the News Articles dataset. We have used a padding to obtain equally sized sequences.

	id	title	author	text	label
0	0	House Dem Aide: We Didn't Even See Comey's Let...	Darrell Lucus	House Dem Aide: We Didn't Even See Comey's Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...	1
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1

Figure 4. News Articles dataset attributes

B. *CNN-LSTM Implementation*

For the experiment, we needed to separate training and testing sets. We have randomly split the dataset into approximately 80% training set and 20% testing set. In order to fine-tune the model hyperparameters, we needed a validation dataset; therefore, we split again the training dataset into 70% training set and 10% validation set. Table II shows the corpus statistics.

TABLE II. DATASETS STATISTICS.

Dataset Statistics	Liar	News articles
Training set size	10,240	11,703
Validation set size	1,284	2,925
Testing set size	1,267	3,657
Real label	7,134	7,924
Fake label	5,657	10,361

We implement our CNN-LSTM framework in Keras [34],following a pattern composed of 7 layers as described in Section IV. We train the network for 400 epochs with a batch size equal to 64 (the number of training examples utilized in one iteration) using Stochastic Gradient Descent (SGD) as optimization for loss function, employing the ReLu as activation function at convolution layer and the sigmoid as activation function at the output layer. We tune these hyperparameters on a validation set (10 % of the data). Table III shows a summury of the proposed CNN-LSTM model.

TABLE III. MODEL SUMMARY

Layer	Input shape	Output shape
Embedding	(None, 50)	(None, 50, 300)
drop-out	(None, 50, 300)	(None, 50, 300)
Conv1D	(None, 50, 300)	(None, 48, 10)
Max Pooling	(None, 48, 10)	(None, 24, 10)
LSTM	(None, 24, 10)	(None, 30)
Dense	(None, 30)	(None, 64)
Output layer: Dense	(None, 64)	(None, 1)

TABLE IV. THE RESULTS OF DIFFERENT METHODS ON TWO DATASETS.

Dataset	Method	Accuracy	Precision	Recall
Liar dataset	SVM	0.608	0.603	0.608
	CNN	0.614	0.611	0.614
	CNN-LSTM	**0.623**	**0.620**	**0.623**
News Articles dataset	SVM	0.683	0.680	0.683
	CNN	0.708	0.701	0.708
	CNN-LSTM	**0.725**	**0.721**	**0.725**

Table IV shows the experimental results of baselines and the proposed approaches on two datasets. We can observe that the overall performance of the proposed CNN-LSTM is much better than the baselines in terms of accuracy, precision and recall on both datasets. On the Liar dataset, the CNN-LSTM outperformed all models, resulting in an accuracy of 62.34%. On News Articles dataset, the highest value 72.50% of accuracy shows that we can well describe fake news content using such CNN-LSTM pair. Therefore, it is more efficient to apply our model on a large dataset in order to improve the fake news detection as opposed to a small datasets. Furthermore, since we have found that the CNN-LSTM model based on text-features discriminates the truthfulness of fake news , we are going to incorporate various metadata in our framework deep learning model. This could help to improve the accuracy of the fake news detection results

VI. Conclusion

In this work, we study the problem of fake news detection. We focus on fake news detection methods based on text-features. We propose a hybrid CNN-LSTM model as a combination of a convolution layer, used to extract unlabeled features, and a LSTM layer used to capture long-term dependencies between the sequences in order to learn a regulatory grammar to improve predictions. Experiments on two real-world datasets demonstrate the heigh accuracy of the CNN-LSTM model in classifying fake news.

The achieved results open several interesting directions for future work. First of all, we believe that fake news detection performance can be further improved. For this reason, we are studying the advantage of using all the other metadata (statement, author, title, and subject) for fake news detection. Second, to better understand the fake news detection characteristics and how to better use deep learning for that, more thorough experiments are required in the future and will be conducted on different datasets. Finally, we aim to understand the correlation between data diffusion, influence [35] and fake news, and we started designing a scenario for studying this aspect.

REFERENCES

[1] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," Journal of Economic Perspectives, vol. 31, no. 2, 2017, pp. 211–36.

[2] 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, 2016, pp. 598–608.

[3] Y. Liu and Y.-F. B. Wu, "Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks," in Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[4] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. ACM, 2017, pp. 797–806.

[5] W. Yaqing et al, "Eann: Event adversarial neural networks for multi-modal fake news detection," in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018, pp. 849–857.

[6] M. Hardalov, I. Koychev, and P. Nakov, "In search of credible news," in International Conference on Artificial Intelligence: Methodology, Systems, and Applications. Springer, 2016, pp. 172–180.

[7] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.-F. Wong, "Detect rumors using time series of social context information on microblogging websites," in Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. ACM, 2015, pp. 1751–1754.

[8] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, 2015, pp. 1–4.

[9] G. L. Ciampaglia et al, "Computational fact checking from knowledge networks," PloS one, vol. 10, no. 6, 2015, p. e0128193.

[10] W. Y. Wang, LIAR, 2017, https://www.cs.ucsb.edu/william/data/liardataset.zip, Last access: June 12, 2019.

[11] W. Ferreira and A. Vlachos, "Emergent: a novel data-set for stance classification," in Proceedings of the conference of the North American chapter of the association for computational linguistics: Human language technologies, 2016, pp. 1163–1168.

[12] P. Lendvai and U. D. Reichel, "Contradiction detection for rumorous claims," arXiv preprint arXiv:1611.02588, 2016.

[13] A. Joulin et al, "Fasttext. zip: Compressing text classification models," arXiv preprint arXiv:1612.03651, 2016.

[14] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," arXiv preprint arXiv:1607.01759, 2016.

[15] X. Zhou and R. Zafarani, "Fake news: A survey of research, detection methods, and opportunities," arXiv preprint arXiv:1812.00315, 2018.

[16] K. Garg, V. Arnaboldi, and S. Giordano, "A novel approach to predict retweets and replies based on privacy and complexity-aware feature planes," in Proceedings of the 5th International Workshop on Complex Networks and their Applications, 2016., 2016, pp. 459–471.

[17] J. Fürnkranz, "A study using n-gram features for text categorization," Austrian Research Institute for Artifical Intelligence, vol. 3, 1998, pp. 1–10.

[18] A. Gupta et al, "Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy," in Proceedings of the 22nd international conference on World Wide Web. ACM, 2013, pp. 729–736.

[19] C. Castillo, M. Mendoza, and B. Poblete, "Predicting information credibility in time-sensitive social media," Internet Research, vol. 23, no. 5, 2013, pp. 560–588.

[20] E. S. Reed, E. Turiel, and T. Brown, "Naive realism in everyday life: Implications for social conflict and misunderstanding," in Values and knowledge. Psychology Press, 2013, pp. 113–146.

[21] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises." Review of general psychology, vol. 2, no. 2, 1998, p. 175.

[22] M. Gentzkow, J. M. Shapiro, and D. F. Stone, "Media bias in the marketplace: Theory," in Handbook of media economics. Elsevier, 2015, vol. 1, pp. 623–645.

[23] W. Quattrociocchi, A. Scala, and C. R. Sunstein, "Echo chambers on facebook," Available at SSRN 2795110, 2016.

[24] M. Del Vicario et al, "The spreading of misinformation online," Proceedings of the National Academy of Sciences, vol. 113, no. 3, 2016, pp. 554–559.

[25] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.

[26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," The Journal of Machine Learning Research, vol. 15, no. 1, 2014, pp. 1929–1958.

[27] A. Kulkarni and A. Shivananda, "Natural language processing recipes: Unlocking text data with machine learning and deep learning using python," 2019.

[28] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in International Workshop on Artificial Neural Networks. Springer, 1995, pp. 195–201.

[29] F. Pedregosa et al, "Scikit-learn: Machine learning in python," Journal of machine learning research, vol. 12, no. Oct, 2011, pp. 2825–2830.

[30] Y. Kim, "Convolutional neural networks for sentence classification," arXiv preprint arXiv:1408.5882, 2014.

[31] M. Abadi, M. Isard, and D. G. Murray, "A computational model for tensorflow: an introduction," in Proceedings of the 1st ACM SIGPLAN International Workshop on Machine Learning and Programming Languages. ACM, 2017, pp. 1–7.

[32] Fake-news, 2015, https://www.kaggle.com/c/fake-news/data, Last access: June 12, 2019.

[33] W. Y. Wang, "" liar, liar pants on fire": A new benchmark dataset for fake news detection," arXiv preprint arXiv:1705.00648, 2017.

[34] F. Chollet, "Keras: Theano-based deep learning library, 2015," URL http://keras. io, Last access: June 12, 2019.

[35] L. Luceri, A. Vancheri, T. Braun, and S. Giordano, "On the social influence in human behavior: Physical, homophily, and social communities," in Proceedings of the Sixth International Conference on Complex Networks and Their Applications., 2017, pp. 856–868.