Automatic Fake News Detection in Urdu Language using Transformers

Iqra Ameer[1], Claudia Porto Capetillo[2], Helena Gómez-Adorno[2] and Grigori Sidorov[1]

[1]Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC), Mexico City, Mexico
[2]Universidad Nacional Autónoma de México (UNAM), Instituto de Investigación en Matemáticas Aplicadas y en Sistemas (IIMAS), Mexico City, Mexico

Abstract

Due to easy access to the internet, the content on social media increased drastically. It is easy to write or spread anything on the web without taking care of the trustfulness of the source. Fake news is now a whole society's problem, sometimes fakes news spread faster than real news. It has adverse effects on people and firms. This makes automatic fake news detection an essential task. Automatic fake news detection has been using in different domains, including social media posts, health, and well-being news, political news, etc. This paper presents the Instituto Politécnico Nacional (Mexico) at FIRE 2021[1] for Urdu language fake news detection shared task [1, 2]. This paper aims to detect fake news on Urdu fake news articles belongs to six different domains, i.e., business, health, showbiz, sports, and technology. In the proposed approach, we applied the state-of-the-art transfer learning algorithm BERT. The best result of 0.91 (see Table 3) is obtained when we trained and validated our model before predictions on the test set. We submitted two different runs of the BERT model in this shared task. Our systems achieved 0.66 accuracy on the unlabeled test dataset provided to evaluate the submitted systems.

Keywords

Fake news, Urdu language, BERT, Classification, Transfer learning

1. Introduction

The universal definition of fake news is: "fictitious articles deliberately fabricated to deceive readers". Fake news became a general public issue, being utilized to spread bogus or rumor information to change individuals' conduct. News websites on the internet and social media spread fake news to increase readers and earn through click-baits. It appeared that the spread of fake news could not be neglected, i.e., the impact of the 2016 US presidential elections [3]. A couple of realities on fake news in the United States:

- Social Media is the source of 62% US citizens for the news [4].
- Fake news had more share on Facebook than mainstream news [5].

[1]http://fire.irsi.res.in/fire/2021/home Last visited: 01-10-2021.

Forum for Information Retrieval Evaluation, December 13-17, 2021, India

iqra@nlp.cic.ipn.mx (I. Ameer); clauporto@comunidad.unam.mx (C. P. Capetillo); helena.gomez@iimas.unam.mx (H. Gómez-Adorno); sidorov@cic.ipn.mx (G. Sidorov)
https://helenagomez-adorno.github.io/ (H. Gómez-Adorno); https://www.cic.ipn.mx/~sidorov/ (G. Sidorov)
0000-0002-1134-9713 (I. Ameer); 0000-0002-9547-924X (C. P. Capetillo); 0000-0002-6966-9912 (H. Gómez-Adorno); 0000-0003-3901-3522 (G. Sidorov)

In this study, we explored the possibility to detect fake textual news based on textual information by applying transfer learning methods [6] on five different domains' news articles in the Urdu language consisting of diverse sorts of information.

1.1. Importance and Applications of Fake News Detection

Nowadays, due to easy access and immense use of the internet, it is easy to spread any news on news websites. Fake news opened up significant issues in our community. The accessibility of data raised difficulties related to checking the credibility/trustworthiness of the information. It is vital to comprehend the effects of sharing possible misinformation. This can take on various forms, and each can adversely affect public communication by misleading and manipulating readers. For example, fake news on Covid-19 is a lot more serious issue as it can impact individuals to take drastic actions by accepting that the news is valid. Surprisingly, a fake statement *"Alcohol is a cure for COVID-19"* prompted numerous deaths and hospitalizations in Iran [7]. This depicts that we are so powerless against fake news in some difficult situations and how extreme the result can be if we overlook them. The initial move towards handling fake news is to recognize it.

Lately, social and political occasions, for example, US presidential election 2016[1], have been set apart by an increasing number of fake news, for example, fabricated news that spread misleading substance, or terribly twist genuine news reports, shared via web-based media platforms. Hence, it is crucial to develop models to detect fake news automatically.

In this article, we worked on fake news detection on news articles in the Urdu language at FIRE 2021[2]. The dataset contains Urdu fake news articles of six different domains, i.e., business, health, showbiz, sports, and technology [8]. For this task, we submitted a system using a transfer learning approach. Specifically, we applied the BERT algorithm to detect fake news in the Urdu language.

2. Related Work

Generally, researchers do not agree when it comes to the definition of fake news. A basic definition of *fake news* is the news that is purposefully fabricated false like news articles to deceive readers. It is adopted in several latest studies [9, 10]. In another definition, deceptive news, for example, news fabrications, hoaxes, satire, etc., are examined as fake news [11, 12]. Although many researchers have been working on fake news detection, automatic fake news detection is an incredibly challenging task for the research community. Therefore, it is required that researchers develop evaluation methods for the fake news detection problem. Some methods use social graph structures to isolate social bots and echo chambers since these are primary fake news sources. This study focused on deep learning approaches.

A convolutional neural network (CNN) is a feed-forward NN comprising hidden convolution and downsampling layers connected with a fully connected output layer. Human visual neurons

---

[1] LevichInstituteandPhysicsDepartment,CityCollegeofNewYork,NewYork,NY10031,USA. Last visited: 27-09-2021

[2] https://www.urdufake2021.cicling.org/home Last visited: 27-09-2021

inspire these networks [13] and represent a variety of Multilayer Perceptron (MLP) networks [14]. The training process of these networks is similar to NNs, but significant differences are in convolution and downsampling layers. Kaliyar et al. [6] proposed a deep convolutional neural network (FNDNet) on Kaggle's fake new corpus to handle fake news detection challenges. Their best-performing model was designed to automatically learn the discriminatory attributes through multiple hidden layers in the deep NN. To extract the various attributes at each layer, they developed a CNN network and achieved promising accuracy of 98.36%.

Ajao et al. [15] applied Long Short-Term Memory (LSTM) and hybrid implementation, specifically CNN-LSTM, on 5,800 tweets related to five rumor stories, including (i) CharlieHebdo, (ii) SydneySiege, (iii) Ottawa Shooting, (iv) Germanwings-Crash, and (v) Ferguson Shooting. They achieved the highest accuracy of 82% on LSTM, and they performed the state-of-the-art performance on the PHEME corpus. Moreover, Mouratidis et al. [16] proposed applied deep neural network SMOTE model and applied it on 2363 tweets corpus of Hong Kong protests in August 2019 [17]. They implemented 18 in total network account features (user id, the Tweet time, Like count, etc.) and linguist features (no. Words, avg. Words in a sentence, no. Long sentences, etc.) [18, 19, 20, 21, 22]. They achieved a great 98% score of F_1 on this small dataset.

Yang and his team [23] worked with Convolutional Neural Networks (CNN) and fed the articles to the network comprising of images to make predictions for fake news. Kaggle's fake news detection corpus[3] was used in this study. Moreover, they manually verified and scrapped real news from genuine official sources like Washington and Post New York Times. The network comprises two sections: (i) text section and (ii) image section. The model's textual section is further divided into the following two subsections: (i) textual explicit—get data from the textual content, for example, length of the news—and the latent text subsection representing the textual content's embedding, restricted to 1000 words. They also divided the image section of the model into two subsections, the one subsection holding information related to characteristics of images, for example, resolution of images or the count of individuals in the images, the other subsection utilized a CNN algorithm on the images. They observed that the model's performance is better when using images (F1-measure = 0.92).

3. Corpus and Task Description

The organizers of fake news detection in the Urdu Language shared task at FIRE 2021 provided one corpus for the Urdu language.

The news articles belong to five different domains: (1) Business, (2) Health, (3) Showbiz, (4) Sports, and (5) Technology. The corpus consists of 1300 labeled Urdu news articles for model training and the 300 unlabeled Urdu news articles for testing the model. The domain distribution are corpus statistics are presented in Table 1 and 2, respectively.

3.1. Task Description

UrduFake @ FIRE 2021: Fake news detection in the Urdu language is a binary classification problem, where it is asked to classify if a particular news article is fake or real. The submitted

[3]https://www.kaggle.com/mrisdal/fake-news Last visited: 03-07-2021

Table 1
Corpus Statistics

Domains	Real	Fake
Business	150	80
Health	150	130
Showbiz	150	130
Sports	150	80
Technology	150	130
Total	**750**	**550**

Table 2
Corpus Statistics

Training Dataset		Testing Dataset	
Real	**Fake**	**Real**	**Fake**
600	438	150	112

models in the shared task are evaluated and ranked by using standard evaluation metrics such as accuracy and macro F_1 score.

4. Methodology

This section describes our approach applied to handle the Urdu language fake news detection shared task.

4.1. Preprocessing

Preprocessing benefits in classification tasks to increase the accuracy of the model [8, 24, 25, 26]. For our approach, we approaches, we performed the following preprocessing on raw Urdu articles:

- Normalized the text using normalize_whitespace,
- Striped the punctuation using remove_punctuation,
- Converted all accent characters into ASCII characters using remove_accents function,
- Lowercased the text.

4.2. Proposed Model

Bidirectional Encoder Representations from Transformers (BERT) is a mostly used transfer learning-based model in NLP-based tasks. The architecture of BERT is highly deep and multi-layered bidirectional, which consists of forwarding and backward layers. The model is to learn the context and structure of the term or word based on the nearby text on both sides. It's an encoder that is based on the transformer blocks. The first step involves in the process of BERT is the training of a large corpus that is unlabeled. It is also called the pre-training of data. The trained model is then used for specific problems in NLP by using its knowledge. It requires a lot of parameter tuning (Fine Tuning) for specific problems to achieve good results.

There are two kinds of BERT models, including (1) BERT Base and (2) BERT Large. These models are developed based on transformers. The Base model consists of 12 layers, while the Large contains 24 layers. The layers are also called transformers. The Base model contains 12 attention heads, and the Large consists of 16. The 110 million parameters were used in the Base model and 340 million in the Large model. In this study, we used bert-base-multilingual-cased variant of the BERT model. This model is pre-trained on the top 104 languages with the largest Wikipedias[4]. This model is pre-trained on two NLP tasks, including (1) Masked Language Modeling (MLM) and (2) Next Sentence Prediction (NSP). The input is provided to the model in the form of embeddings representation. Before providing the input sequences to the model, 15% of the terms were replaced with Mask token. For the computations, the input is provided to the next layer for intermediate representations by using a transformer and generating the next layer's output. The output is converted into the vocabulary dimensions. In the topmost layer, the probabilities are calculated for each word, and classification is performed [27].

4.3. Hyperparameter Tuning

We performed following parameter tuning to choose a set of optimal hyperparameters for our BERT model:

- **Batch size**: 16 and 32,
- **Learning rate**: 5e-5, 3e-5, 2e-5,
- **Epsilon**: 1e-8,
- **No. of Epochs**: 4.

5. Results and Analysis

Table 3 presents the Accuracy, Macro Precision (Ma_P), Macro Recall (Ma_R), and F_1 scores obtained by applying the transfer learning method on the dataset provided by the organizers of the Urdu fake news detection shared task at FIRE'2021. In this Table, "Experiments" refers to the number of experiments we performed using the BERT model. The"Model" refers to the transfer learning-based models applied in this study. We trained and tested two models to analyze the performance of models on the Urdu fake news dataset. Please note that both BERT-1 and BERT-2 models are the same. The only difference between these models is the training process. We trained the BERT-1 by using only training data (see table 2) and tested on the test dataset. In contrast, the BERT-2 was trained on training and validated on a validation set.

Overall, the best results are achieved on the BERT-2 model. In Exp1, we trained the BERT-1 model on the training dataset and evaluated the model on the testing dataset. We achieved 0.89 accuracy. In Exp2, we divided the training dataset into train and validation sets (10% of training data) to validate the BERT-2 and tested it on the test dataset. We achieved 0.90 accuracy, which is slightly more than Exp1.

[4]https://huggingface.co/bert-base-multilingual-cased Last visited: 20-09-2021.

Table 3
Obtained results using BERT on training data

Experiments	Model	Accuracy	Ma_P	Ma_R	F_1
Exp1	BERT-1	0.89	0.90	0.90	0.89
Exp2	BERT-2	0.90	0.91	0.90	0.91

On the unlabeled test set, our model obtained 0.66 accuracy[5], which is relatively low. Usually, BERT based systems get higher results, so maybe the results were not evaluated correctly. This highlights that fake news detection in the Urdu language is a challenging task. Moreover, the complex transfer learning models need a lot of training data for better training.

6. Conclusion and Future Work

Automatic fake news detection is a classification task aiming to develop a reliable model classifying a given text as either fake or real. The fake news proliferation on news websites, social media posts, blogs, etc., and the Internet is misleading people to the extent that it needs to be stopped. In this study, we described our approach to detecting fake news on Urdu news articles belonging to six different domains: business, health, showbiz, sports, and technology.

The best result of 0.91 on the labelled dataset (see Table 3) is obtained using state-of-the-art transfer learning BERT algorithm when trained and validated our model before predictions on the test set. We submitted two different runs of the BERT model in this shared task. On an unlabeled dataset, our systems achieved 0.66 accuracy.

In the future, we plan to apply other transfer learning-based models such as RoBERTa, XLNet, etc., to classify Urdu fake news articles.

7. Acknowledgments

The work was done with partial support from the Mexican Government through the grant A1-S-47854 of the CONACYT, Mexico, grants 20211784, 20211884, and 20211178 of the Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional, Mexico and grants DGAPA-UNAM PAPIIT project number TA100520 of the Instituto de Investigación en Matemáticas Aplicadas y en Sistemas of Universidad Nacional Autónoma de México, Mexico. The authors thank the CONACYT for the computing resources brought to them through the Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje of the Laboratorio de Supercómputo of the INAOE, Mexico and the support of Microsoft through the Microsoft Latin America PhD Award.

References

[1] M. Amjad, S. But, H. I. Amjad, A. Zhila, G. Sidorov, Urdufake@ fire2021: Shared track on fake news identification in urdu, In Forum for Information Retrieval Evaluation, 2021.

[5]https://www.urdufake2021.cicling.org/results-and-rankings Last visited: 01-10-2021.

[2] M. Amjad, S. But, H. I. Amjad, A. Zhila, G. Sidorov, A. Gelbukh, Overview of the shared task on fake news detection in urdu at fire 2021, In CEUR Workshop Proceedings, 2021.
[3] H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election, Journal of economic perspectives 31 (2017) 211–36.
[4] J. Gottfried, E. Shearer, News use across social media platforms 2016, http://www.journalism.org/2016/05/26/news-use-across-social-me dia-platforms-2016/ (2016).
[5] C. Silverman, L. Alexander, How teens in the balkans are duping trump supporters with fake news. buzzfeed, 14 november, 2016.
[6] R. K. Kaliyar, A. Goswami, P. Narang, S. Sinha, Fndnet–a deep convolutional neural network for fake news detection, Cognitive Systems Research 61 (2020) 32–44.
[7] N. Karimi, J. Gambrell, Hundreds die of poisoning in iran as fake news suggests methanol cure for virus, 2020.
[8] M. Amjad, G. Sidorov, A. Zhila, H. Gómez-Adorno, I. Voronkov, A. Gelbukh, "bend the truth": Benchmark dataset for fake news detection in urdu language and its evaluation, Journal of Intelligent & Fuzzy Systems 39 (2020) 2457–2469.
[9] E. Mustafaraj, P. T. Metaxas, The fake news spreading plague: was it preventable?, in: Proceedings of the 2017 ACM on web science conference, 2017, pp. 235–239.
[10] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, B. Stein, A stylometric inquiry into hyperpartisan and fake news, arXiv preprint arXiv:1702.05638 (2017).
[11] M. Balmas, When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism, Communication research 41 (2014) 430–454.
[12] V. L. Rubin, N. Conroy, Y. Chen, S. Cornwell, Fake news or truth? using satirical cues to detect potentially misleading news, in: Proceedings of the second workshop on computational approaches to deception detection, 2016, pp. 7–17.
[13] M. Matsugu, K. Mori, Y. Mitari, Y. Kaneda, Subject independent facial expression recognition with robust face detection using a convolutional neural network, Neural Networks 16 (2003) 555–559.
[14] Y. LeCun, et al., Lenet-5, convolutional neural networks, URL: http://yann. lecun. com/exdb/lenet 20 (2015) 14.
[15] O. Ajao, D. Bhowmik, S. Zargari, Fake news identification on twitter with hybrid cnn and rnn models, in: Proceedings of the 9th international conference on social media and society, 2018, pp. 226–230.
[16] D. Mouratidis, M. N. Nikiforos, K. L. Kermanidis, Deep learning for fake news detection in a pairwise textual input schema, Computation 9 (2021) 20.
[17] M. N. Nikiforos, S. Vergis, A. Stylidou, N. Augoustis, K. L. Kermanidis, M. Maragoudakis, Fake news detection regarding the hong kong events from tweets, in: IFIP international Conference on Artificial Intelligence Applications and Innovations, Springer, 2020, pp. 177–186.
[18] A. Sittar, I. Ameer, Multi-lingual author profiling using stylistic features., in: Working Notes for MAPonSMS at FIRE'18 - Workshop Proceedings of the 10th International Forum for Information Retrieval Evaluation (FIRE'18). CEUR-WS.org, CEUR, DAIICT, Gujarat, India (2018), 2018, pp. 240–246.
[19] I. Pervaz, I. Ameer, A. Sittar, R. M. A. Nawab, Identification of author personality traits

using stylistic features, in: CLEF 2015 Evaluation Labs and Workshop – Working Notes Papers, 8-11 September, Toulouse, France, 2015.

[20] I. Pervaz, I. Ameer, A. Sittar, R. M. A. Nawab, Identification of author personality traits using stylistic features: Notebook for pan at clef 2015., in: CLEF (Working Notes), Citeseer, 2015.

[21] I. Ameer, G. Sidorov, R. M. A. Nawab, Author profiling for age and gender using combinations of features of various types, Journal of Intelligent & Fuzzy Systems 36 (2019) 4833–4843.

[22] M. H. F. Siddiqui, I. Ameer, A. F. Gelbukh, G. Sidorov, Bots and gender profiling on twitter., in: Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, Lugano, Switzerland, September 9-12., 2019.

[23] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li, P. S. Yu, Ti-cnn: Convolutional neural networks for fake news detection, arXiv preprint arXiv:1806.00749 (2018).

[24] N. Ashraf, S. Butt, G. Sidorov, A. Gelbukh, Cic at checkthat! 2021: Fake news detection using machine learning and data augmentation, CLEF, 2021.

[25] I. Ameer, G. Sidorov, Author profiling using texts in social networks, in: Handbook of Research on Natural Language Processing and Smart Service Systems, IGI Global, 2021, pp. 245–265.

[26] I. Ameer, M. H. F. Siddiqui, G. Sidorov, A. Gelbukh, Cic at semeval-2019 task 5: Simple yet very efficient approach to hate speech detection, aggressive behavior detection, and target classification in twitter, in: Proceedings of the 13th International Workshop on Semantic Evaluation, 2019, pp. 382–386.

[27] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).