

Urdu Fake News Detection using Generalized Autoregressors

Abdullah Faiz Ur Rahman Khilji^a, Sahinur Rahman Laskar^a, Partha Pakray^a and Sivaji Bandyopadhyay^a

^aDepartment of Computer Science and Engineering, National Institute of Technology Silchar, Assam, India

Abstract

Automated fake news detection has become vital in today's digital age. Differentiating legit news from fake ones has become an important classification challenge in natural language processing (NLP). Various transformer-based deep learning approaches have taken widespread adoption from the research community due to its outstanding performance. We have participated in the 2020 Fake News Detection Challenge in the Urdu Language organized by Center for Computing Research (CIC), Instituto Politécnico Nacional (IPN), Mexico and have stood second. In this work, we have implemented a generalized autoregressor based model to classify news into fake or real. We have achieved an overall accuracy of 0.8400 and F1 macro score of 0.8370.

Keywords

Fake News Detection, Classification, Autoregressors, XLNet

1. Introduction

The age of the internet has led to a massive free flow of information. This information flow has an added advantage of being easily accessible through widespread virtualization, leading to cost benefits to humankind. This increased accessibility has led to widespread usage and adaptability of the services it has to offer, to the point that one's principal source of information becomes the internet. Due to its ubiquitous and global nature, it is quite difficult to have a central authority to have control over it, without defeating the true purpose it has to offer. As we are moving more and more of our life to the virtual world, the importance of detecting fake news has increased manifold as it has proven to have a negative impact on our society [1]. There are various instances where fake news has been used intentionally [2] to spread misinformation. This deliberate attempt has amplified in the era of social media where any individual is ready to express their views without considering any facts, one such example is given by [3]. It is also shown that the spread of fake news is exponential [3] and any initial attempts could greatly help in curtailing the issue. Hence, there is a need to automate fake

Forum for Information Retrieval Evaluation 2020, December 16-20, 2020, Hyderabad, India

EMAIL: abdullah_ug@cse.nits.ac.in (A. F. U. R. Khilji); sahinur_ug@cse.nits.ac.in (S. R. Laskar); partha@cse.nits.ac.in (P. Pakray); sivaji.cse.ju@gmail.com (S. Bandyopadhyay)

URL: <https://abdullahkhilji.github.io/> (A. F. U. R. Khilji); <https://sahinurlaskar.github.io/> (S. R. Laskar);

<http://cs.nits.ac.in/partha/> (P. Pakray)

ORCID: 0000-0001-6621-1810 (A. F. U. R. Khilji); 0000-0002-8413-2718 (S. R. Laskar); 0000-0003-3834-5154 (P. Pakray)



© 2020 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

news detection. Fake news detection has also garnered a great deal of attention in the past from both industry and academia. [4].

Our work mainly focuses on the fake news detection task of the Urdu language. In this work, we have considered the task of fake news detection as a binary classification problem. We have used XLNet [5] which is based on generalized autoregressive pre-training for language understanding, to train our models. Also, we have created a monolingual corpus for the Urdu language and developed pre-trained language models for giving an impetus for various downstream NLP tasks in the Urdu language. Our work is based on the dataset and the task provided by the organizers of 2020 Fake News Detection in the Urdu Language [6, 7].

Rest of the paper is organized as follows. In Section 2 we discuss the relevant works in fake news detection. Dataset used in our work is described in Section 3. We describe our system in detail in Section 4. Experimental setup is discussed in brief in Section 5. We discuss the result of our model and its analysis in Section 6. Finally, we conclude along with future works in Section 7.

2. Related Work

Work related to automated fake news detection can be broadly classified into three distinct categories viz. based on content, propagation, and social context [8]. Majorly, content-based approaches have been used in the past that leverages lexical and syntactical features to detect writing styles and words to capture the deceptive nature of the news. One such work by [9], uses satirical cues based on a support vector machine (SVM) [10] algorithm to detect fake news. In the past researchers have also worked upon to compare the linguistic features of fake news with real news [11] and also gives a basis to the role of using stylistic cues to define the truthfulness of a text. Work done by [12] have used computation models and have provided a comparative analysis with human-based fake news detection task. Researchers have also worked with comparing and analyzing the text from left-wing and right-wing news and comparing it with the mainstream [13] and have demonstrated stylometric inquiry into hyper-partisan and fake news.

Very few works have been done on fake news detection on the Urdu language in the past, due to very less availability of the requisite dataset for this task. For the given Urdu fake news detection task [14] is the only available dataset. Work done by [15] uses machine translation (MT) approaches to detect fake news in Urdu language. Various machine learning (ML) approaches like random forest (RF), logistic regression (LR), and boosting (BO) have been utilized by [16] to detect fake news in the Spanish language. Works like [17] and [18] uses emotional cues in the text to detect fake news.

3. Dataset

3.1. Description

For our work, the organizers of 2020 Fake News Detection in the Urdu Language Task provided us with the dataset for fake news detection [14]. The authors of [14] manually collected and

annotated for this binary classification task. The train data included a total of 638 samples, 350 from real and 288 from fake. The evaluation set included a total of 262 samples, out of which 150 are real and 112 fake. The test set comprised of 400 news samples. Since, a large Urdu monolingual corpus is not available for pre-training our model (as discussed in Section 2). We have created our own Urdu monolingual corpus for pre-training the deep learning (DL) models. We have prepared the dataset by extracting Urdu sentences from various sources including web pages, blogs, books, and government websites. The resultant monolingual data obtained contains about 11 million sentences.

3.2. Preprocessing

The data provided by the organizers was already preprocessed as discussed by the authors of [14]. For the monolingual corpus, we have used UrduHack¹ to preprocess the collected corpus. The preprocessing step included the removal of URLs, email IDs, and mobile numbers. The text was also normalized, space was added after punctuation marks and extra spaces were removed. Diacritics were also removed along with normalizing Urdu characters to obtain the text in the proper Urdu range.

4. System Description

For pretraining as well as fine-tuning we have followed the architecture of [5]. The pretraining and fine-tuning procedure are described in Section 4.1 and Section 4.2 respectively. We have used the XLNet model instead of bidirectional encoder representations from transformers (BERT) [19] as, the later model has some drawbacks to offer. The BERT based architecture tends to corrupt the input with masks as it does not consider the dependency between the masked positions. Also, the BERT model has a discrepancy between fine-tuning and pretraining the model [5] as some symbols like [MASK] are absent during the fine-tuning step.

4.1. Pretraining

Leveraging unlabeled corpora has been very beneficial in various NLP tasks. Moreover, data for unsupervised tasks are easily available in large quantities compared to their supervised counterparts. Principally, there are two distinct pre-training objective's viz. autoencoding (AE) and autoregressive (AR) language modeling [5]. Hence, following the work of [5], we have pre-trained our model based on the XLNet architecture, which employs a generalized autoregressive method that takes the advantage of both pre-training objectives AE and the AR language modeling. Apart from the architectural improvements in BERT, this XLNet model based on [20], integrates the recurrence mechanism and relative encoding scheme that improves performance in tasks involving longer sentences, like the given task of fake news classification.

¹<https://github.com/urduhack>

Table 1

Test Results of our Team CNLP-NITS on the Fake News Classification Task

	Precision	Recall	F1 Score
Fake Class	0.8359	0.7133	0.8235
Real Class	0.8419	0.9160	0.8235

4.2. Fine Tuning

For fine-tuning the pre-trained model for the main binary classification task i.e. to detect whether the news is legit or fake we have followed the fine-tuning procedure of BERT [19] as given by [5]. Following [5], a span-based prediction is employed, wherein a fixed length is sampled and consecutive span of tokens are randomly selected as prediction targets. The problem of pretrain finetune discrepancy is also solved here.

5. Experimental Setup

We have trained the XLNet model that uses the AR pre-training method and employs the use of language modeling objectives based on permutation. Since, each individual Urdu document/news provided by the organizers was to be predicted as fake or real, following [5] we used a longer sequence length of 512. Due to a smaller classification dataset and limited computation power, we trained a much smaller model as originally trained by the authors of [5]. We have used 6 layers with 4 attention heads, and an embedding dimension of 256. Inherently, larger parameters were scaled down by four. Here, the XLNet based model is pre-trained on the same domain of news dataset. Feature weighting scheme's are not used. The embedding features from the pre-training step (obtained after Section 4.1) is used in the fine-tuning step (as discussed in Section 4.2) for classification. The XLNet classification task can be termed as the classification algorithm used.

6. Results

The results for the fake news detection task was declared by the organizers of 2020 Fake News Detection in the Urdu Language. The authors of this work participated by the team name CNLP-NITS and as shown by the overall results on the website² our team stood second out of the 39 teams participated. The participated systems were evaluated based on precision, recall, and F1 score. Our system reported an overall accuracy of 0.8400 and an overall F1 macro score of 0.8370. Precision, Recall and F1 score for the fake class and the real class are as shown in Table 1.

²<https://www.urdufake2020.cicling.org/results-and-rankings>

7. Conclusion and Future Work

Our automated fake news detection system adopts the generalized autoregressors technique for the binary classification task. Even though our XLNet based model shows good results, there are avenues for improvements. The monolingual corpus required for the pre-training step can be increased further. Several architectural improvements could also be made to incorporate the challenges faced in the Urdu language.

Acknowledgments

Authors would like to thank the organizers of the 2020 Fake News Detection in the Urdu Language Task from Center for Computing Research (CIC), Instituto Politécnico Nacional (IPN), Mexico for organizing this competition. The authors are also grateful to the Center for Natural Language Processing (CNLP) and Department of Computer Science and Engineering at National Institute of Technology, Silchar for providing the requisite support and infrastructure to execute this work.

References

- [1] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: A data mining perspective, in: SIGKDD Explor., 2017, pp. 22–36.
- [2] B. Nyhan, J. Reifler, When corrections fail: The persistence of political misperceptions, in: Political Behavior, 2010, pp. 303–330.
- [3] A. Peck, A problem of amplification: Folklore and fake news in the age of social media, in: The booktitle of American Folklore, 2020, pp. 329–351.
- [4] D. M. J. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, J. L. Zittrain, The science of fake news, in: Science, 2018, pp. 1094–1096.
- [5] Z. Yang, Z. Dai, Y. Yang, J. G. Carbonell, R. Salakhutdinov, Q. V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, in: Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems, 2019, pp. 5754–5764.
- [6] M. Amjad, G. Sidorov, A. Zhila, A. Gelbukh, P. Rosso, Urdufake@fire2020: Shared track on fake news detection in urdu (2020). Proceedings of the 12th Forum for Information Retrieval Evaluation (FIRE 2020), Hyderabad, India.
- [7] M. Amjad, G. Sidorov, A. Zhila, A. Gelbukh, P. Rosso, Overview of the shared task on fake news detection in urdu at fire 2020, CEUR Workshop Proceedings (2020). Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2020), Hyderabad, India.
- [8] X. Zhou, R. Zafarani, Fake news: A survey of research, detection methods, and opportunities, in: CoRR, 2018, p. 1.
- [9] V. 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.
- [10] W. S. Noble, What is a support vector machine?, in: Nature Biotechnology, 2006, pp. 1565–1567.
 - [11] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, 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, 2017, pp. 2931–2937.
 - [12] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, Automatic detection of fake news, in: Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018, 2018, pp. 3391–3401.
 - [13] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, B. Stein, A stylometric inquiry into hyper-partisan and fake news, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers, 2018, pp. 231–240.
 - [14] M. Amjad, G. Sidorov, A. Zhila, H. Gómez-Adorno, I. Voronkov, A. F. Gelbukh, "bend the truth": Benchmark dataset for fake news detection in urdu language and its evaluation, in: J. Intell. Fuzzy Syst., 2020, pp. 2457–2469.
 - [15] M. Amjad, G. Sidorov, A. Zhila, Data augmentation using machine translation for fake news detection in the urdu language, in: Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020, 2020, pp. 2537–2542.
 - [16] J. P. Posadas-Durán, H. Gómez-Adorno, G. Sidorov, J. J. M. Escobar, Detection of fake news in a new corpus for the spanish language, in: J. Intell. Fuzzy Syst., 2019, pp. 4869–4876.
 - [17] A. Giachanou, P. Rosso, F. Crestani, Leveraging emotional signals for credibility detection, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019, 2019, pp. 877–880.
 - [18] B. Ghanem, P. Rosso, F. M. R. Pardo, An emotional analysis of false information in social media and news articles, in: ACM Trans. Internet Techn., 2020, pp. 19:1–19:18.
 - [19] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, pp. 4171–4186.
 - [20] Z. Dai, Z. Yang, Y. Yang, J. G. Carbonell, Q. V. Le, R. Salakhutdinov, Transformer-xl: Attentive language models beyond a fixed-length context, in: Proceedings of the 57th Conference of the Association for Computational Linguistics, 2019, pp. 2978–2988.