Fake News Detection in the Urdu Language using CharCNN-RoBERTa

Nankai Lin[a], Sihui Fu[a] and Shengyi Jiang[a,b]

[a] *Guangdong University of Foreign Studies, Guangzhou, China*
[b] *Guangzhou Key Laboratory of Multilingual Intelligent Processing, Guangzhou, China*

Abstract
In this article, we report the solution of the team BERT 4EVER for the Fake News Detection in the Urdu Language task in FIRE 2020, which aims to identify deceiving news articles in the Urdu language spread via digital media. We propose the CharCNN-RoBERTa model to tackle the problem. In addition, we adopt label smoothing and ensemble learning to improve the generalization capability. Experimental results as well as the leading position of our team on the task leaderboard demonstrate the effectiveness of our method.

Keywords 1
Fake news detection, Urdu, RoBERTa, CharCNN, Label smoothing

1. Introduction

Fake news refers to the news articles that are intentionally and verifiably false [1]. Some people or organizations might make up and publish fake news for their immoral purposes and interests. In addition to creating public disturbance, fake news could be used to discredit and slander a person, incite social disharmony, foment political unrest, or even undermine peace and stability of the international community. In the Information Age, with a great deal of news springing up all over the Internet every day, it is necessary to develop some methods to automatically determine whether a news source is trustworthy or not. However, existing research mainly focuses on high-resource languages such as Chinese and English. Even though the Urdu language has more than 100 million speakers across the world, it is still a low-resource language in NLP both from the perspective of the inaccessibility of NLP tools as well as the scarcity of the annotated datasets [2].

Luckily, FIFE 2020 proposes the task “Fake News Detection in the Urdu Language" [3][4], which aims at encouraging more work to address the problem of identifying deceiving Urdu news articles spread via digital media. The dataset used in this task comes from [5], whose texts were crawled from 16 news websites including BBC News. The distribution of the dataset is shown in Table 1. Our team, BERT 4EVER, also participates in this task and achieves the first rank in all evaluation metrics. In this report, we will review our solution to this task, namely, the CharCNN-RoBERTa model aided by label smoothing and ensemble learning.

Table 1 The statistics of the Urdu fake news dataset used in this task

Category	Real	Fake
Business	100	50
Health	100	100
Showbiz	100	100
Sports	100	50
Technology	100	100
Total	500	400

Forum for Information Retrieval Evaluation 2020, December 16-20, 2020, Hyderabad, India
EMAIL: neakail@outlook.com (A. 1); sihuifu93@gmail.com (A. 2); jiangshengyi@163.com (A. 3)
ORCID: 0000-0003-2838-8273 (A. 1); 0000-0002-3070-4813 (A. 2); 0000-0002-6753-474X (A. 3)


2. Related work

Shu et al. pointed out two main methods of fake news detection [6]: ***news content models*** and ***social context models***. For **news content models**, a typical work is from Borges et al. [7], which combined string similarity features with a deep neural architecture that leveraged the ideas previously advanced in the context of learning efficient text representations, document classification, and natural language inference. On the other hand, **social context models** try to evaluate the stance which different news sources take towards the assertion, to identify fake news [8], while topic model related methods, such as latent dirichlet allocation (LDA), can be applied to learn such latent stances from topics [9]. Tacchini et al. proposed to construct a bipartite network of users and Facebook posts using the "like" stance information [10]. Jin et al. explored topic models to learn latent viewpoint values and further exploited these viewpoints to learn the credibility of relevant posts and news content [9]. They took advantage of the "wisdom of crowds" information to improve news verification by mining conflicting viewpoints in microblogs.

As regards the Urdu language, Amjad et al. presented a new corpus for fake news detection in their pioneering work [5], i.e. the dataset used in this task. In this work, they also provided their detection method and baseline results. Based on this corpus, they further investigated whether machine translation at its present state could be successfully used as an automated technique for the creation and augmentation of annotated corpora, focusing on the English-Urdu language pairs [11]. The experimental results show that since the current machine translation quality for the English-Urdu translation is not satisfactory, the single data augmentation method by means of machine translation[2] could not provide improvement for fake news detection in Urdu. So, in this task we do not consider data augmentation, although the dataset in question is quite small.

3. Method

3.1. Overview

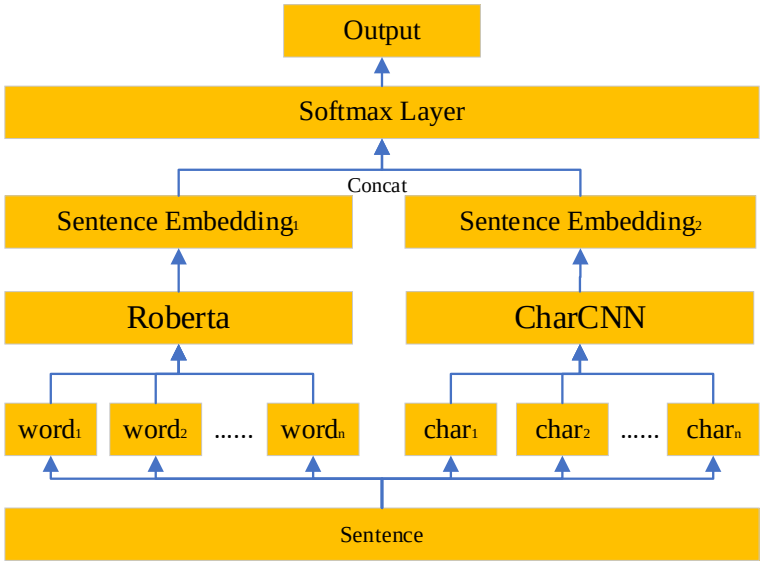

Figure 1: The CharCNN-RoBERTa model we proposed

The model we finally employ in this task is shown in Figure 1. A text is first represented as word embeddings as well as character embeddings. Then they go through the RoBERTa model and the CharCNN model respectively, for obtaining sentence embeddings with respect to both the word level and character level. Next the resulting two sentence vectors are concatenated and fed into the softmax layer for predicting the authenticity of the text. Furthermore, we adopt label smoothing to improve the generalization capability of the model.

[2] They augmented the original Urdu corpus by automatically translating an English fake news dataset into Urdu.

3.2. RoBERTa

Since BERT [12] achieved SOTA performance on multiple tasks of natural language processing, the industry has proposed various pre-trained models (PTMs). In general, during the training phase, given specific training targets, these models will try to learn the general representation of a language from large-scale unlabeled data. When applied in downstream tasks, they are often used as language feature extractors to obtain sentence embeddings. In this task, we use the pre-trained language model RoBERTa [13], also a derivative of BERT. Building upon BERT's language masking strategy, RoBERTa makes some modifications, which include the use of dynamic masking, the removal of the next sentence prediction objective, along with larger mini-batches and more training iterations. All these allow RoBERTa to improve on the masked language modeling objective and lead to better performance for downstream tasks. Luckily, UrduHack has released a pre-trained Urdu RoBERTa model[3]. We download it and further train it only on the real news in the dataset.

3.3. CharCNN

We adopt CharCNN [14] as the character feature extractor. After segmenting the input text by characters, we use the convolutional layer to extract information from the local fields, and then the max pooling layer to perform feature dimensionality reduction, so that the most significant character-level features could be retained.

3.4. Label smoothing

For the cross entropy loss function, we need to use the predicted probabilities to fit the true probabilities. However, fitting the one-hot encoded labels will cause the model to be too confident about its predictions, and hence the generalization capability of the model cannot be guaranteed. The label smoothing technique is thus proposed to mitigate the problem of overconfidence and overfitting.

Assuming that $\boldsymbol{y}$ is a one-hot encoded label, after label smoothing, it can be expressed as:

$$\boldsymbol{y}_i' = (1-\varepsilon) * \boldsymbol{y}_i + \frac{\varepsilon}{K-1} * (1 - \boldsymbol{y}_i)$$

where ε is the smoothing factor, and K refers to the number of categories. In our case, K is 2. In our experiment, ε is set to 0.05.

4. Result

Amjad et al. proposed to use XGBoost with multiple handcrafted features, such as word n-grams, character n-grams and functional word n-grams [5]. Their results as our baseline, we try to evaluate the performance of different strategies described in Section 3. The results on the validation set are shown in Table 2. Further pretraining RoBERTa on the data indeed improves prediction accuracy.

Table 2 The results on the validation set

Model	Real F1	Fake F1
XGBoost with multiple features [8]	87	90
RoBERTa[4]	86	91
RoBERTa+charcnn+pretrain	87	91
RoBERTa +pretrain	**89**	**92**
RoBERTa+pretrain+label smoothing	87	91
RoBERTa+charcnn+pretrain+label smoothing	**89**	**92**

Since the dataset we are dealing with is quite small, to obtain a less biased model, we further conduct five-fold cross-validation for each combined strategy, using the binary F1 score as the evaluation

[3] https://huggingface.co/urduhack/urdu-roberta

[4] The original pre-trained RoBERTa model from UrduHack, without further pre-training on the data in question.

metrics. For each strategy, its mean of the five models is shown in Table 3. This time the further pre-trained RoBERTa along with CharCNN and label smoothing gives the best performance.

Table 3 The results of the five-fold cross-validation

Num	Model	F1
1	RoBERTa+pretrain	90.99% (0.000052)
2	RoBERTa+pretrain+label smoothing	91.18% (0.000575)
3	RoBERTa+charcnn+pretrain	91.25% (0.000215)
4	**RoBERTa+charcnn+pretrain+label smoothing**	**91.41% (0.000194)**

* Each value in the bracket is the variance of the five models.

To further promote the generalization capability, from the perspective of ensemble learning, we also attempt to combine the decisions of the strategies mentioned in Table 3. In our case, we employ the averaging technique, namely, taking an average of the predictions from the models in question and making it as the final prediction. Since a team is allowed to submit three different runs in this task, we ultimately provide three predictions by three solutions for the final testing. For the first model, we ensemble the predictions given by all four strategies. As for the second one, we only ensemble those strategies without label smoothing. The third model only adopts the single strategy which gives the best result in the five-fold cross-validation. The results of these three submissions, released by the task committee, are shown in Table 4. They show that the ensemble technique can indeed improve overall performance. Moreover, label smoothing can also lend a helping hand.

Table 4 The final result of our submissions

Ensemble models	F1	Accuracy
1+2+3+4	**0.9071**	**0.9075**
1+3	0.8995	0.9000
4	0.8969	0.8975

5. Conclusion

In this report, we present the solution of BERT 4EVER to the Fake News Detection in the Urdu Language task in FIRE 2020. We propose to use the CharCNN-RoBERTa model, along with label smoothing and ensemble learning, to identify deceiving news articles. The results of our approach achieve the top rank on the task leaderboard, with regard to all evaluation metrics. In the future, we will further explore effective strategies for dealing with low-resource languages. At the same time, we will also study how to use a small number of data to achieve competitive results in deep learning.

6. Acknowledgements

This work was supported by the Key Field Project for Universities of Guangdong Province (No. 2019KZDZX1016), the National Natural Science Foundation of China (No. 61572145) and the National Social Science Foundation of China (No. 17CTQ045). The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

7. References

[1] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake News Detection on Social Media: A Data Mining Perspective, ACM SIGKDD Explorations Newsletter 19(1) (2017), pp. 22–36.

[2] C. Cieri, M. Maxwell, S.M. Strassel, J. Tracey, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, Selection Criteria for Low Resource Language Programs, in: Proceedings of the 10th. International Conference on Language Resources and Evaluation, LREC'2016, European Language Resources Association, (2016).

[3] M. Amjad, G. Sidorov, A. Zhila, A. Gelbukh, P. Rosso, UrduFake@FIRE2020: Shared Track on Fake News Detection in Urdu, in: Proceedings of the 12th Forum for Information Retrieval Evaluation, (2020).
[4] 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).
[5] M. Amjad, G. Sidorov, A. Zhila, H. G. Adorno, "Bend the Truth": Benchmark Dataset for Fake News Detection in Urdu Language and Its Evaluation. Journal of Intelligent & Fuzzy Systems Preprint, (2020), pp. 2457 – 2469.
[6] K. Shu, S. Wang, H. Liu, Beyond News Contents: The Role of Social Context for Fake News Detection, Association for Computing Machinery, (2019), pp. 312-320.
[7] L. Borges, B. Martins, P. Calado, Combining Similarity Features and Deep Representation Learning for Stance Detection in the Context of Checking Fake News, Journal of Data and Information Quality, 11(3) (2019), pp. 1936-1955.
[8] B. Riedel, I. Augenstein, G. P. Spithourakis. A simple but tough-to-beat baseline for the Fake News Challenge stance detection task, (2017).
[9] Z. Jin, J. Cao, Y. Zhang, J. Luo. News verification by exploiting conflicting social viewpoints in microblogs. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI Press, Phoenix, Arizona, (2016), pp. 2972-2978.
[10] E. Tacchini, G. Ballarin, M. L D. Vedova, S. Moret, L. de Alfaro, Some like it hoax: Automated fake news detection in social networks, CoRR, (2017).
[11] 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, European Language Resources Association, Marseille, France, (2020), pp. 2537–2542.
[12] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, (2019).
[13] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov. RoBERTa: A Robustly Optimized BERT Pretraining Approach, (2020).
[14] X. Zhang, J. Zhao, Y. Lecun, Character-level Convolutional Networks for Text Classification, Neural Information Processing Systems, (2015), pp. 649–657.