A GRU-based Fake News Prediction System: Working Notes for UrduFake-FIRE 2020

Saichethan Miriyala Reddy[a], Chanchal Suman[b], Sriparna Saha[b] and Pushpak Bhattacharyya[b]

[a] IIIT Bhagalpur,India
[b] Indian Institute of Technology Patna,India

Abstract
With the escalating use of the Internet worldwide and substantially increasing impact produced by the availability of ambiguous information, the challenge to quickly identify fake news in digital media in various languages becomes more acute. In this work, we have worked on a poor resource language i.e. Urdu for detecting fake news. The latent features of the news articles are extracted from Bidirectional GRU, followed by the concatenation of average and max-pooling layers. Finally, the class label is predicted from the softmax layer. This work is a part of UrduFake, a shared task of FIRE-2020. An average f1 of 80.78% is achieved on the test data. The developed system achieved the fourth position in the competition.

Keywords
Fake news, Urdu, Gated Recurrent unit, Transformers

1. Introduction

The easy access, low cost, and rapid spread of information lead users to digest social media news. With these advantages, the widespread of fake news also occurs with social media. Thus the usage of social media for news can be considered as a double-edged sword. Fake news is low-quality news, having the wrong information. The wrong information is put intentionally, with aim of spreading. propagandists usually manipulate the news for conveying the political influence or message. For example, in order to spread false stories, Russia created fake accounts and social bots, as per reports. [1]. Fake news grows significantly when a change in public opinion is demanded during any event. This is more common in digital media. After the 2016 U.S. presidential elections, fake news and its impacts were widely discussed. The study conducted by Silverman shows that there are 8,711,000 shares, reactions, and comments on the top 20 fake news stories of the election on Facebook [2]. Thus, assessing the credibility of news and detecting it before its dissemination becomes an important task. This can be done through different natural language processing algorithms, such that one can determine whether a source is trustworthy or not [3]. The availability of information through different social media platforms has raised the challenges associated with testing the trustworthiness of the data

Forum for Information Retrieval Evaluation 2020, December 16-20, 2020, Hyderabad, India
miriyala.cse.1725@iiitbh.ac.in (S. M. Reddy); 1821cs11@iitp.ac.in (C. Suman); sriparna@iitp.ac.in (S. Saha); pb@iitp.ac.in (P. Bhattacharyya)
https://saichethan.github.io/ (S. M. Reddy); https://github.com/chanchalIITP (C. Suman)

automatically [4]. For this reason, it is necessary to build systems for controlling the amount of factually incorrect and misleading data on the Web [5].

Motivated by the success of deep learning-based techniques, we have developed a fake news detection system for Urdu tweets in this work. We used a bi-directional gated recurrent unit for finding the feature representation of the dataset samples. Finally, the learned representation is fed to a combination of the pooling layer, followed by the softmax layer for the classification. Experiments are performed on the dataset released by the organizers of UrduFake- a shared task of FIRE-2020 [6, 7]. The problem objective of the task is to address the problem of detecting deceiving information in the Urdu language using digital media text. An average F1 of 80.78% with 81.75% accuracy is achieved from our developed system on the test dataset. Our proposed system got the fourth rank in the competition.

2. Previous Works

Fake news detection is divided into three categories i) serious fabrication, ii)large-scale hoaxes, and iii) humorous fake news [8]. Most of the recently developed model are based on deep learning [9, 10, 11, 12, 13, 14, 15]. In [9], a competitive model is proposed for finding the relationship between original false information and the updated information. It reduces the impact of false information. Trust of users is also considered as an important feature for detecting fake news [10]. Different kinds of social networks are examined in [11]. Detection and mitigation of fake news is also examined through these networks. A geometric based deep learning is adapted, for detecting the fake news [12]. In [13], the task-generic features are applied for solving the deception task. An emotion-based fake news detection method is proposed in [14], which combines the social emotion and the publisher emotion.

3. Dataset

The Urdu fake news dataset is composed of news articles in six different domains: education, technology, sports, business, entertainment, and politics. The news included in this dataset is intentionally written by a group of professional journalists, each proficient in corresponding topics [5]. There are 900 news articles, of which 500 belong to the real class, and 400 articles belong to the fake class. The samples of the real class are taken from legitimate news sources. The dataset developers have manually verified the authenticity of the articles. For the generation of the fake subset, professional journalists were hired. The hired journalists were native in Urdu and were instructed to intentionally write deceptive news articles [5].

4. Methodology Used

Motivated by the recent works for detecting fake news [9, 10, 11, 12, 13, 14, 15], we have used three different methodology for developing the system. They are discussed below.

Table 1
Class-wise Dataset statistics

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

4.1. Model-I

Firstly, the news samples are broken into tokens via the Keras tokenizer. After that, the tokenized samples are fed to a bi-directional Gated Recurrent Unit (GRU), for generating the feature representation. A pre-trained Urdu word embedding [16], developed from the skip-gram model is used for representing the word into its semantic space. To represent the spatial relationship of features generated from Bi-GRU, the max pooling, and average pooling is calculated. Finally, the two pooling representations are concatenated together to form the final feature representation. The generated representation vector is fed to a sigmoid layer for classifying into a fake and real class.

4.2. Model-II

Model-II is similar to Model-I, except for the pooling layer. In this model, we have used only average pooling for representing the spatial connections. Finally, the representation vector generated from the pooling layer is fed to the sigmoid layer for classification.

4.3. Model-III

The multi-head self-attention based transformer is used for generating the vector representation of news samples. The generated feature is then passed through a global max pooling layer with a dropout of 10%. The pooled vector is then passed through a hidden layer followed by the 10% dropout. Finally, softmax is used for class label prediction.

4.4. Implementation Details

In table 2, the different hyper-parameters along with the values are shown. 250 is taken as the maximum sequence length for the samples. Padding and truncating are used for making all the samples, equal size. 10000 features are taken in the generated vocabulary. 300 size of Urdu word embedding[16] is taken for representing the words, and 64 hidden units are defined in the bi-GRU network.

Table 2
Hyper-parameters description

Hyper-parameter	Unit
Max. sequence length	250
Vocabulary Size	10000
Embedding dimension	300
No. of nodes in GRU	64

Table 3
Performance of our developed models on test data

	Fake			Real				
Model	Precision	Recall	F1-Mac	Precision	Recall	F1-Mac	Avg F1	Avg Accuracy
Model-I	88.11	59.33	78.80	79.59	95.20	78.80	80.78	81.75
Model-II	83.96	59.33	77.59	79.25	93.20	77.59	79.61	80.50
Model-III	82.75	16.00	52.86	66.03	98.00	52.86	59.37	67.25

5. Results

In this section, we have discussed the performance of our proposed models on the given Urdu dataset. The average accuracy achieved on test data is 81.75%, 79.61%, and 59.37%. The average F1 achieved for the best model i.e Model-I is 80.78%. The detailed results for the three models are shown in table 3. The performance of the transformer based model is poor in comparison to the GRU-one. One of the reason can be overfitting, due to more number of parameters and less number of samples.

6. Conclusion

Automatic detection of fake news is a promising area of research. In recent years, accessing reliable and accurate information has become difficult for users, due to the abundant information on social media. Thus, automatic detection of fake-news predictions on social media has become a promising area of research. In this work, a bi-GRU-based fake news prediction system is developed, which uses Urdu embedding for word representation. The experiments are performed on the dataset shared by the UrduFake FIRE-2020 organizers. Our developed system stood the fourth rank in the competition.

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 (2017) 22–36.
[2] N. Sitaula, C. K. Mohan, J. Grygiel, X. Zhou, R. Zafarani, Credibility-based fake news detection, in: Disinformation, Misinformation, and Fake News in Social Media, Springer, 2020, pp. 163–182.

[3] B. Ghanem, P. Rosso, F. Rangel, An emotional analysis of false information in social media and news articles, ACM Transactions on Internet Technology (TOIT) 20 (2020) 1–18.
[4] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, B. Stein, A stylometric inquiry into hyperpartisan and fake news, arXiv preprint arXiv:1702.05638 (2017).
[5] 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 (2020) 1–13.
[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] V. L. Rubin, Y. Chen, N. K. Conroy, Deception detection for news: three types of fakes, Proceedings of the Association for Information Science and Technology 52 (2015) 1–4.
[9] H. Zhu, H. Wu, J. Cao, G. Fu, H. Li, Information dissemination model for social media with constant updates, Physica A: Statistical Mechanics And Its Applications 502 (2018) 469–482.
[10] K. Shu, S. Wang, H. Liu, Understanding user profiles on social media for fake news detection, in: 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), IEEE, 2018, pp. 430–435.
[11] K. Shu, H. R. Bernard, H. Liu, Studying fake news via network analysis: detection and mitigation, in: Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining, Springer, 2019, pp. 43–65.
[12] F. Monti, F. Frasca, D. Eynard, D. Mannion, M. M. Bronstein, Fake news detection on social media using geometric deep learning, arXiv preprint arXiv:1902.06673 (2019).
[13] A. Olivieri, S. Shabani, M. Sokhn, P. Cudré-Mauroux, Creating task-generic features for fake news detection, in: Proceedings of the 52nd Hawaii International Conference on System Sciences, 2019.
[14] C. Guo, J. Cao, X. Zhang, K. Shu, M. Yu, Exploiting emotions for fake news detection on social media, arXiv preprint arXiv:1903.01728 (2019).
[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, 2020, pp. 2537–2542.
[16] S. Haider, Urdu word embeddings, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018.

A. Online Resources

The sources for the implementation are available via

- Source code link.