A Simple N-Gram Model for Urdu Fake News Detection

Hamada Nayel[1], Ghada Amer[2]

[1]Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt
[2]Electrical Engineering Department, Faculty of Engineering, Benha University, Benha, Egypt

Abstract

Fake news in social media platforms is a critical issue and it is necessary to detect such news. In this paper, we describe the system submitted to the UrduFake@FIRE2021. The aim of this shared task is to detect fake news in Urdu language. Machine learning approach has been used to build our model. A linear classifier using Stochastic Gradient Descent (SGD) optimization algorithm has been used to develop our system. The proposed model achieved F1-score of 67.9% and secured first rank over all submissions for different teams.

Keywords

Social Media Analysis, Fake News Detection, Linear Classifiers, ML Approach

1. Introduction

News that are intentionally and verifiably false called fake news [1]. In this era, detecting fake news is a critical task due to the tremendous spread of news over the social media platforms. People or organizations with specific background might fabricate and publish fake news for unethical purposes [2]. Fake news can be used to insult and defamation individuals, as well as obstruct social order, incite political unrest, or even undermine the peace and stability of the international community. More interesting and worse, research on the spreading of fake news shows that fake news is significantly faster, deeper, and wider distributed than true news [3].

It has been proven that fake news is spreading exponentially, and any attempts in the first stages would greatly help in reducing the problem [4]. Fake news detection obtained a great deal of interests in the past from both of academic researchers and industry [5].

This paper presents the system submitted to the UrduFake@FIRE2021 shared task [6], held in conjunction with FIRE2021. The rest of the paper is organized as follows: section 2 overviews the related work, section 3 describes in details the structure of our system and section 4 shows the results of our model. In section 5, the results that the proposed model obtained have been discussed and finally in section 6, a sight on the future work has been given.

Forum for Information Retrieval Evaluation, December 13-17, 2021, India
hamada.ali@fci.bu.edu.eg (H. Nayel); ghada.amer@bhit.bu.edu.eg (G. Amer)
https://bu.edu.eg/staff/hamadaali14 (H. Nayel); https://bu.edu.eg/staff/ghadaamer5 (G. Amer)
0000-0002-2768-4639 (H. Nayel); 0000-0001-6083-2376 (G. Amer)

2. Related Work

Researchers have conducted inclusive research in fake news detection. Ensemble approach has been used to develop the model for detecting the fake news in Urdu [7]. Amjad et. al. [8] used Machine Translation (MT) for dataset augmentation. They merged the translated English fake news to Urdu with the original Urdu dataset [9]. Authors used Support Vector Machines (SVM) algorithm with character and word n-grams features to train the model. The model achieved f1-score ranging from 0.83 to 0.89 higher than that of the f1-score obtained for the dataset through MT. Nankai et. al. combined RoBERTa model for word embeddings Convolutional Neural Network (CNN) for character level embeddings along with label smoothing and ensemble learning to develop a deep learning model for Urdu fake news detection [2]. Principally, most of the work in fake news detection focused on English [10, 11]. Research efforts have been done in other languages, such as Arabic [12], Indonesian [13] and Italian [14].

3. Dataset

The dataset used in the shared task, named Bend-The-Truth, was distributed by the organizers, which is divided into training, development and test set. It consists of news articles in six different domains: technology, education, business, sports, politics, and entertainment [9]. The sources of real news are news channels websites, such as BBC Urdu News, CNN Urdu, Express-News, Jung News, Naway Waqat, and many other reliable news websites for the time frame from January 2018 to December 2018. A very rigorous procedure has been followed while collecting the real news. On the other hand, the fake news articles are intentionally written by a group of journalists, each expert in corresponding topics. The fake news articles are in the same domains and almost of the same length as the real news articles. Full details of the dataset are given in [9].

4. Methodology

NEWUrduFake task is modeled as a binary classification task. Given a set of news articles in Urdu, $N = \{n_1, n_2, n_3, ...\}$, the task aims at assigning a label from a predefined set $L = \{F, R\}$ to each news article. The label F refers to the news article which is fake, while R refers to the news article which is true news.

Vector space model has been used to represent the news article and the weighting scores for unique tokens were calculated by Term frequency / Inverse Document Frequency (TF/IDF) [15]. TF/IDF has been used efficiently in native language identification [15], offensive language detection [16, 17], irony detection [18] and author profiling [19, 20]. To evaluate the effect of N-gram models, a wide range of N-gram models have been generated and used along with TF/IDF for building different systems. A set of classification algorithms namely; Multinomial Naive Bayes, SVM, Linear and MLP have been used for training the model.

4.1. Model Structure

The structure of our model is shown in Figure 1. The first phase of our model is feature extraction. In this phase, TF/IDF has been applied to extract the features of the input data. The next phase is training the model, in this phase, we tried different algorithms and evaluated using development set. The final phase is producing the model and applying it to the blind test set to get the final output.

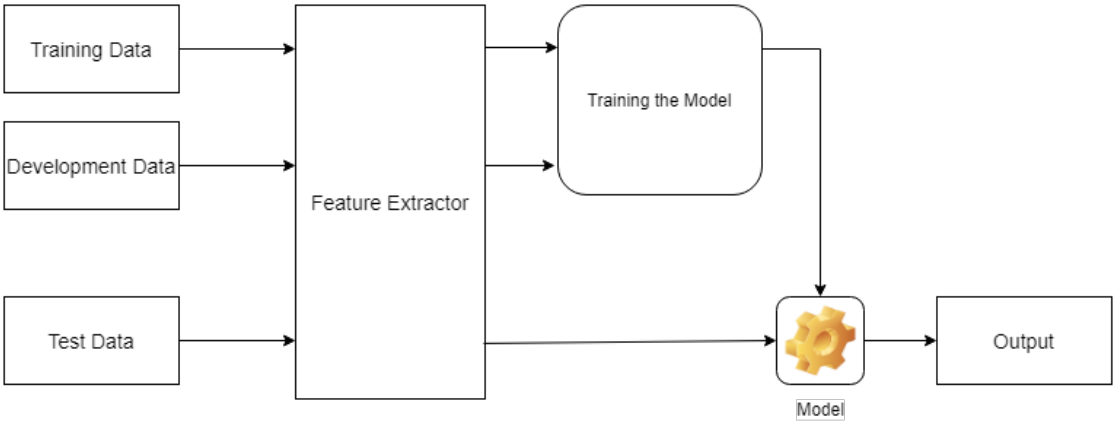

Figure 1: Model Structure

4.2. Experimental Setup

A simple tokenization technique has been used based on white space character. In feature extraction, tokens have been used without any preprocessing, which increases the number of features. TF/IDF has been extracted for uni-gram, bi-gram and tri-gram models. SVM with linear kernel, linear classifier with SGD training and MLP algorithms with different nodes have been implemented and evaluated on the development set.

5. Results and Discussion

In the model development phase, we applied the different models on the training dataset, and the results are given in Table 1. Result shows that SGD classifier outperformes MLP and SVM for bi-gram and tri-gram models. We decided to submit the output of SGD algorithm. Table 2 shows the results of applying different algorithms with ranges of n-gram models. It s clear that the best performed classifier is SGD with bi-gram model. Also, it secured the first rank among all the participants.

The proposed model used different machine learning classification algorithms, and the best performed algorithm has been used for output submission. TF/IDF with a wide range of n-gram

models have been used to extract the feature for training the model, and a set of rich features has been produced.

Table 1
Results of development set

		F1-Macro	Accuracy
Uni-gram	SVM	0.688	0.737
	SGD	0.741	0.756
	MLP	0.766	0.786
Bi-gram	SVM	0.626	0.695
	SGD	0.752	0.767
	MLP	0.752	0.756
Tri-gram	SVM	0.611	0.683
	SGD	0.717	0.737
	MLP	0.699	0.729

Table 2
Results of test set

		F1-Macro	Accuracy
Uni-gram	SVM	0.538	0.693
	SGD	0.677	0.737
	MLP	0.644	0.737
Bi-gram	SVM	0.549	0.710
	SGD	0.679	0.757
	MLP	0.630	0.737
Tri-gram	SVM	0.542	0.707
	SGD	0.677	0.757
	MLP	0.643	0.743

6. Conclusion

Our system uses TF/IDF as a language model, it is very basic and simple. Using more accurate language model such as word embeddings may improve the performance of the model. On the other hand, preprocessing step, if added, may enhance the accuracy of the system.

References

[1] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: A data mining perspective, SIGKDD Explor. Newsl. 19 (2017) 22–36. URL: https://doi.org/10.1145/3137597.3137600. doi:10.1145/3137597.3137600.

[2] N. Lin, S. Fu, S. Jiang, Fake news detection in the urdu language using charcnn-roberta, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2020 - Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020, volume 2826 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 447–451. URL: http://ceur-ws.org/Vol-2826/T3-2.pdf.

[3] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, Science 359 (2018) 1146–1151. URL: https://www.science.org/doi/abs/10.1126/science.aap9559. doi:10.1126/science.aap9559. arXiv:https://www.science.org/doi/pdf/10.1126/science.aap9559.

[4] A. Peck, A problem of amplification: Folklore and fake news in the age of social media, The Journal of American Folklore 133 (2020) 329–351. doi:10.5406/jamerfolk.133.529.0329.

[5] 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, Science 359 (2018) 1094–1096. URL: https://www.science.org/doi/abs/10.1126/science.aao2998. doi:10.1126/science.aao2998.

[6] A. Maaz, S. Butt, H. I. Amjad, A. Zhila, G. Sidorov, A. Gelbukh, Overview of the shared task on fake news detection in urdu at fire 2021, in: Working Notes of FIRE 2021 - Forum for Information Retrieval Evaluation, CEUR Workshop Proceedings, CEUR-WS.org, 2021.

[7] F. Balouchzahi, H. L. Shashirekha, Learning models for urdu fake news detection, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2020 - Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020, volume 2826 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 474–479. URL: http://ceur-ws.org/Vol-2826/T3-7.pdf.

[8] 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. URL: https://aclanthology.org/2020.lrec-1.309.

[9] 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, Journal of Intelligent Fuzzy Systems 39 (2020) 2457–2469. URL: https://doi.org/10.3233/JIFS-179905. doi:10.3233/JIFS-179905.

[10] F. M. R. Pardo, A. Giachanou, B. Ghanem, P. Rosso, Overview of the 8th author profiling task at PAN 2020: Profiling fake news spreaders on twitter, in: L. Cappellato, C. Eickhoff, N. Ferro, A. Névéol (Eds.), Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, September 22-25, 2020, volume 2696 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020. URL: http://ceur-ws.org/Vol-2696/paper_267.pdf.

[11] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, Automatic detection of fake news, in: Proceedings of the 27th International Conference on Computational Linguistics, Association for Computational Linguistics, Santa Fe, New Mexico, USA, 2018, pp. 3391–3401. URL: https://aclanthology.org/C18-1287.

[12] M. Alkhair, K. Meftouh, K. Smaïli, N. Othman, An arabic corpus of fake news: Collection, analysis and classification, in: K. Smaïli (Ed.), Arabic Language Processing: From Theory to Practice, Springer International Publishing, Cham, 2019, pp. 292–302.

[13] I. Y. R. Pratiwi, R. A. Asmara, F. Rahutomo, Study of hoax news detection using naïve bayes classifier in indonesian language, in: 2017 11th International Conference on Information Communication Technology and System (ICTS), 2017, pp. 73–78. doi:10.1109/ICTS.2017.8265649.

[14] F. Pierri, A. Artoni, S. Ceri, Investigating italian disinformation spreading on twitter in the context of 2019 european elections, PloS one 15 (2020). URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0227821.

[15] H. A. Nayel, H. L. Shashirekha, Mangalore-University@INLI-FIRE-2017: Indian Native Language Identification using Support Vector Machines and Ensemble Approach, in: P. Majumder, M. Mitra, P. Mehta, J. Sankhavara (Eds.), Working notes of FIRE 2017 - Forum for Information Retrieval Evaluation, Bangalore, India, December 8-10, 2017., volume 2036 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2017, pp. 106–109. URL: http://ceur-ws.org/Vol-2036/T4-2.pdf.

[16] H. Nayel, NAYEL at SemEval-2020 task 12: TF/IDF-based approach for automatic offensive language detection in Arabic tweets, in: Proceedings of the Fourteenth Workshop on Semantic Evaluation, International Committee for Computational Linguistics, Barcelona (online), 2020, pp. 2086–2089. URL: https://aclanthology.org/2020.semeval-1.276.

[17] A. Allam, H. Abdallah, E. Amer, H. Nayel, Machine learning-based model for sentiment and sarcasm detection, in: Proceedings of the Sixth Arabic Natural Language Processing Workshop, Association for Computational Linguistics, Kyiv, Ukraine (Virtual), 2021, pp. 386–389. URL: https://aclanthology.org/2021.wanlp-1.51.

[18] H. A. Nayel, W. Medhat, M. Rashad, BENHA@IDAT: Improving Irony Detection in Arabic Tweets using Ensemble Approach, in: P. Mehta, P. Rosso, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2019 - Forum for Information Retrieval Evaluation, Kolkata, India, December 12-15, 2019, volume 2517 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 401–408. URL: http://ceur-ws.org/Vol-2517/T4-3.pdf.

[19] H. A. Nayel, NAYEL@APDA: Machine Learning Approach for Author Profiling and Deception Detection in Arabic Texts, in: P. Mehta, P. Rosso, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2019 - Forum for Information Retrieval Evaluation, Kolkata, India, December 12-15, 2019, volume 2517 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 92–99. URL: http://ceur-ws.org/Vol-2517/T2-3.pdf.

[20] M. Sobhi, A. Hassan, A. El-Sawy, H. Nayel, Machine learning-based approach for Arabic dialect identification, in: Proceedings of the Sixth Arabic Natural Language Processing Workshop, Association for Computational Linguistics, Kyiv, Ukraine (Virtual), 2021, pp. 287–290. URL: https://aclanthology.org/2021.wanlp-1.34.