Learning Models for Urdu Fake News Detection

Fazlourrahman Balouchzahi, H L Shashirekha

Department of Computer Science, Mangalore University, Mangalore - 574199, India

Abstract

Detecting fake news from the real news can be modeled as a typical binary text classification problem. Most of the models proposed for fake news detection address the resource rich languages such as English and Spanish but, languages such as Urdu, Persian, Balouchi and many Indian native languages have received very less attention due to unavailability of bench marked corpus. To promote text processing activities on Urdu, which happens to be a resource poor language FIRE 2020 (Forum for Information Retrieval Evaluation) has called for UrduFake, a shared task to detect fake news in Urdu language. High speed of news broadcast and the importance of detecting fake news from the real news made us (team MUCS) to propose three different learning models namely, an ensemble of Machine Learning (ML) models, Transfer Learning (TL) model based on ULMFiT and a hybrid model made up of an ensemble of ML approaches, TL approach and Deep Learning (DL). The proposed methodology utilizes word and character n-grams to train ML model and word embedding vectors to train BiLSTM networks of DL model and for TL model, a pre-trained general domain Urdu Language Model is fine-tuned with the Urdu fake news dataset. Our ML model obtained 5th place among 9 teams that participated in this task.

Keywords

Fake news Detection, Learning Models, BiLSTM, ULMFiT

1. Introduction

Today the speed of broadcasting news is increasing rapidly due to the availability of various online platforms and social media such as Facebook, Twitter, WhatsApp etc. Online platforms serve as a great opportunity for fake news spreaders to manipulate communities' minds and also social trust [1] due to anonymity of users. Fake news can target unity of people in the society and also can impact the society in a negative way. Detecting the ever increasing fake news manually is laborious, time consuming and error prone. Further, as news articles are unstructured text and usually noisy, efficient approaches are required to detect fake news automatically [2]. Most of the proposed fake news detection tasks have addressed resource rich languages such as English and Spanish [3].But, resource poor languages such as Urdu, Persian, Balouchi and many Indian native languages have received less attention due to unavailability or less availability of labeled data. To promote text processing activities on Urdu, FIRE 2020 has called for UrduFake, a shared task to detect fake news in Urdu language [4][5]. Fake news detection can be modeled as a typical binary text classification problem where each news article is classified as either fake or real [6]. In this paper, we, team MUCS, propose three different

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

frs_b@yahoo.com (F. Balouchzahi); hlsrekha@gmail.com (H.L. Shashirekha)

https://mangaloreuniversity.ac.in/dr-h-l-shashirekha (H.L. Shashirekha)

0000-0003-1937-3475 (F. Balouchzahi)

models namely, an ensemble of Machine Learning (ML) models, Transfer Learning (TL) based on ULMFiT and a hybrid model made up of an ensemble of ML models, TL learning model and Deep Learning (DL) model for Urdu fake news detection.

2. Literature Review

Fake news detection is a challenging task particularly for resource poor languages. Due to unavailability or less availability of bench marked corpus, several researchers have created their own datasets and have developed various models to detect fake news. Some of the relevant works are mentioned below: A DL model based on LSTM networks to detect false news from Twitter and news article proposed by Bilal et. al. [1] use emotions to illustrate that false information can be detected based on the combination of different emotional patterns. They have reported an f1 score of 96% on a dataset including trusted news created from English Gig word corpus as real news and collection of news from seven different unreliable news sites as false news. Urdu fake news detection proposed by Ajmad et. al. [3] have used Machine Translation (MT) to translate English fake news dataset consisting of 200 legitimate and 200 fake news [7] to Urdu and combined it with an original Urdu dataset that contains 500 real and 400 fake news [8]. Using character and word n-grams to train Support Vector Machine the authors have reported that the results on original Urdu dataset with f1 score ranging from 0.83 to 0.89 are higher than that of the f1 score obtained for the dataset through MT. Two models based on different learning approaches for English and Spanish languages have been submitted to fake news spreader detection at PAN 2020[1] shared task by Shashirekha et. al. [9] [10] an, i) an ensemble of ML models using majority voting of the three (two Linear SVC classifiers and a Logistic Regression classifier) classifiers built using Unigram TF/IDF, N_gram TF and Doc2Vec feature sets and ii) a TL model based on Universal Language Model Fine-Tuning (ULMFiT) initially trained on a general domain English/Spanish data collected from Wikipedia which is then fine-tuned using target task dataset and used for the fake news spreader detection task as the target model. Trained on the dataset provided by PAN 2020 [11], the ML model obtained 73.50% and 67.50% accuracies and TL model 62% and 64% accuracies on English and Spanish languages respectively.

3. Methodology

We propose three different learning models for Urdu fake news detection, namely, i) an ensemble of ML models trained with word and character n-grams, ii) TL model based on ULMFiT using a pre-trained Urdu Language Model (LM) fine-tuned with Urdu fake news dataset and iii) HTC - a hybrid model made up of models used in i), ii) and a DL model trained with word embedding vectors. The framework of HTC model is shown in Figure 1.

The base models used for the proposed approaches are described below:

(i) **Ensemble of ML models:** Three ML models, namely, Multinomial Naïve Bayes (MNB), Multilayer Perceptron (MLP), and Logistic Regression (LR) are ensembled using 'hard'

[1]https://pan.webis.de/

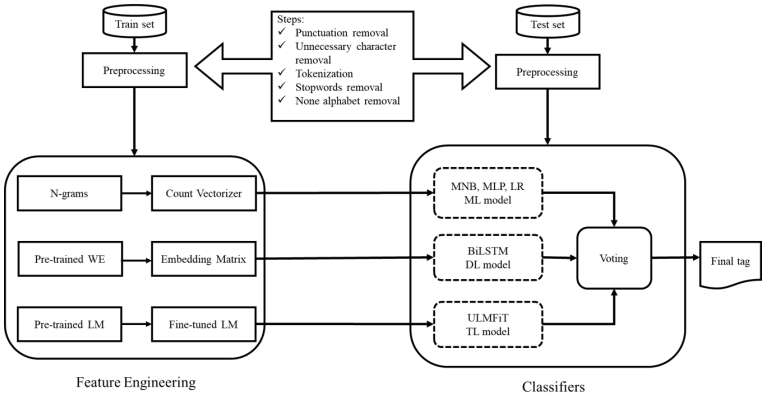

Figure 1: Framework of HTC model.

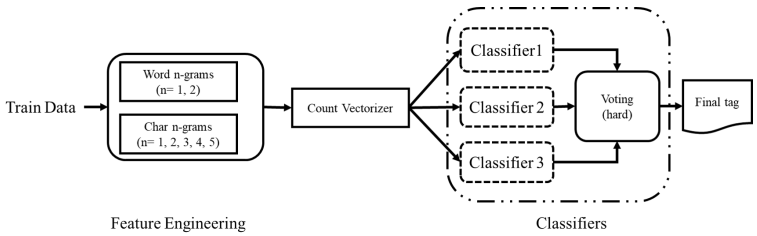

Figure 2: Architecture of ensemble of ML models.

voting. All the three models are trained on vectors obtained using CountVectorizer module from word n-grams (n= 1, 2) and char n-grams (n=1, 2, 3, 4, 5). For MLP, hidden layer sizes are set to (150, 100, 50) and maximum iteration, activation, solver, and random state have been set to 300, Relu, Adam and 1 respectively and for MNB and LR classifiers default parameters are used.Figure 2 gives the architecture of ensemble of ML models.

(ii) **DL model:** It has been implemented using a pre-trained Skipgram word embedding model trained on Wikipedia texts and the parameters used for training are: "alpha: 0.05, "hs": 0, "iter": 15, "max_n": 5, "min_count": 50, "min_n": 2, "negative": 20, "sample": 0.0001, "sg": 1, "size": 300, "window": 10, "word_ngrams": 1[2]. Word embeddings are used to build embedding matrix for the given dataset which is used to train a multi-channel BiLSTM network of three channels with similar configuration as Conv1D (200, 3, activation='relu', padding='same')[3]. The model has been trained in 20 epochs each with a batch of size 256, 128, 64, and 32. Figure 3 shows the architecture of DL model.

(iii) **TL model:** It consists of three stages namely, Language Model (LM) training, LM fine-tuning, and target task classifier. LM is a probability distribution over word sequences in a language. In TL model, the knowledge obtained in solving one task called source task is used to develop another task, called the target task [12] [13].In the proposed TL

[2]https://github.com/urduhack/urdu-word-vectors

[3]A 1D Convolutional Neural Networks CNN is very effective for deriving features from a fixed-length segment of the overall dataset, where it is not so important where the feature is located in the segment.

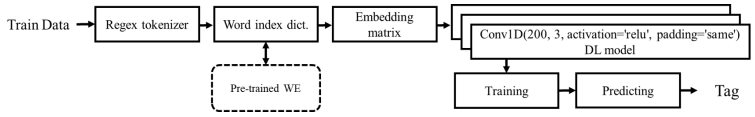

Figure 3: Architecture of DL model.

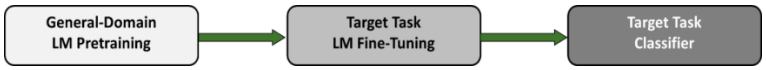

Figure 4: Transfer Learning model frame work.

model based on ULMFiT, source model is a pre-trained general domain Urdu LM[4] that represents the general features of Urdu language and target model is a fake news detection model. The pre-trained LM is fine-tuned with the target task dataset for Urdu fake news detection. TL model is implemented based on ULMFiT architecture introduced by Howard et. al. [14] and target classifier using text.models module from fastai library. Inspired by Stephen et. al. [15], an encoder for an ASGD Weight-Dropped LSTM (AWD-LSTM) is implemented which can be plugged in with a decoder and classifying layers to create a text classifier. AWD-LSTM has shown noticeable results on word-level models consisting of a word embedding of size 400, 3 hidden layers and 1150 hidden activations per layer [14]. A framework of TL model is shown in Figure 4.

4. Experimental Results

Train and test data are pre-processed by removing punctuation, stopwords, numbers and unnecessary characters such as @, , $, %. Classifier models are constructed using the respective features extracted by the feature engineering module. Test data is classified based on the majority voting of the predicted labels in case of ensemble of ML models and HTC model.

4.1. Dataset

The training and development corpus called Bend-The-Truth data consisting of Fake and Real news provided by UrduFake[5] task organizers are shown in Table 1. Dataset consists of Urdu news articles collected from various channels such as BBC Urdu News, CNN Urdu, Express-News, Jung News, Naway Waqat, and some other news websites [3]. Further, 400 news articles are provided by the organizers as private test set for evaluating the learning models.

4.2. Results

The labels for the test data predicted by the three proposed models are submitted to UrduFake shared task organizers and the results reported by organizers are shown in Table 2. Among

[4]https://github.com/anuragshas/nlp-for-urdu
[5]https://www.urdufake2020.cicling.org/home

Table 1
Statistics of the corpus used for training and development set

Category	Business	Health	Showbiz	Sports	Technology	Total
Real	100	100	100	100	100	500
Fake	50	10	100	50	100	400

Table 2
Results of our proposed models

Models	Fake			Real			F1 avg.	Accuracy
	P	R	F1 mac.	P	R	F1 mac.		
Ensemble of ML	0.7833	0.6266	0.7707	0.8000	0.8960	0.7707	0.7894	0.7950
TL	0.5918	0.3866	0.6143	0.6953	0.8400	0.6143	0.6509	0.6700
HTC	0.7956	0.4933	0.7192	0.7524	0.9240	0.7192	0.7467	0.7625

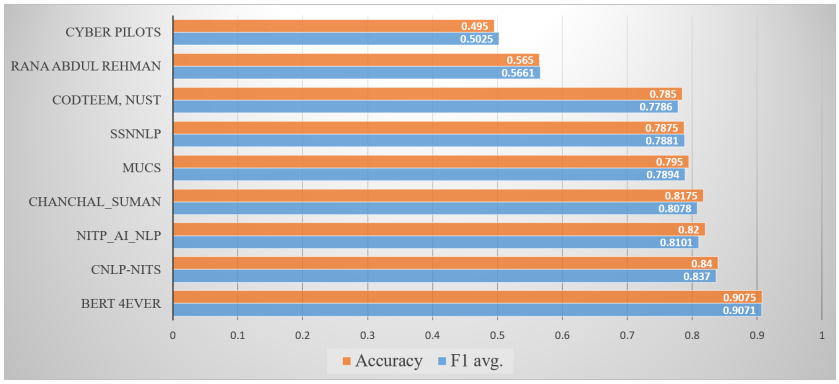

Figure 5: Comparison of accuracy and F1 average of the models submitted by 9 teams

the three proposed models, ensemble of ML models obtained higher results compared to other two models with an average f1 score of 0.7894. Also, our team, MUCS, obtained 5th rank in UrduFake challenge among the 9 participating teams. The higher performance for ensemble of ML models is due to n-grams features that have already proved their effectiveness in many works in NLP. TL model has obtained less performance from what was expected because of a general domain LM used as pre-trained LM instead of domain specific pre-trained LM. Further, the lower performance of DL model is may be because only word embeddings are used as features. The lower performances of DL and TL models have resulted in lower performance of HTC model. A comparison of accuracy and F1 average of the models submitted by the 9 teams is shown in Figure 5.

5. Conclusion and Future work

We, team MUCS, proposed three different learning models namely, an ensemble of ML models, TL model based on ULMFiT and HTC - a hybrid model made up of an ensemble of ML models, TL model based on ULMFiT and DL model for the detection of UrduFake news task in FIRE 2020.

Our team, obtained 5th rank for ensemble of ML models among the 9 participating teams. We would like to explore different features and improve learning models and perform experiments on native and low resource languages such as Urdu, Persian and other Indian languages.

References

[1] 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.

[2] J. Tang, Y. Chang, H. Liu, Mining social media with social theories: a survey, ACM Sigkdd Explorations Newsletter 15 (2014) 20–29.

[3] 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.

[4] M. Amjad, G. Sidorov, A. Zhila, P. Rosso, A. Gelbukh, Urdufake@fire2020: Overview of the track on fake news detection in urdu, In Proceedings of the 12th Forum for Information Retrieval Evaluation. (2020).

[5] 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.

[6] C. Aggarwal, C. Zhai, A survey of text classification algorithms in mining text data (2012) 163–222.

[7] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, Automatic detection of fake news, arXiv preprint arXiv:1708.07104 (2017).

[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 (2020) 1–13.

[9] M. D. Anusha, H. L. Shashirekha, N. S. Prakash, Ensemble model for profiling fake news spreaders on twitter - notebook for pan at clef 2020, In Linda Cappellato, CarstenEickhoff, Nicola Ferro, and AurélieNévéol, editors, CLEF 2020 Labs and Workshops, Notebook Papers, CEUR-WS.org (2020).

[10] F. Balouchzahi, H. L. Shashirekha, Ulmfit for twitter fake news spreader profiling - notebook for pan at clef 2020, In Linda Cappellato, CarstenEickhoff, Nicola Ferro, and AurélieNévéol, editors, CLEF 2020 Labs and Workshops, Notebook Papers, CEUR-WS.org (2020).

[11] F. Rangel, A. Giachanou, B. Ghanem, P. Rosso, Overview of the 8th author profiling task at pan 2020: profiling fake news spreaders on twitter, in: CLEF, 2020.

[12] F. Balouchzahi, H. L. Shashirekha, PUNER-Parsi ULMFiT for Named-Entity Recognition in Persian Texts, Technical Report, EasyChair, 2020.

[13] S. Faltl, M. Schimpke, C. Hackober, Ulmfit: State-of-the-art in text analysis (2019).

[14] J. Howard, S. Ruder, Universal language model fine-tuning for text classification, arXiv preprint arXiv:1801.06146 (2018).

[15] S. Merity, N. S. Keskar, R. Socher, Regularizing and optimizing lstm language models, arXiv preprint arXiv:1708.02182 (2017).