CNLP-NITS-PP at WANLP 2022 Shared Task: Propaganda Detection in Arabic using Data Augmentation and AraBERT Pre-trained Model

Sahinur Rahman Laskar[1], **Rahul Singh**[1], **Abdullah Faiz Ur Rahman Khilji**[1]
Riyanka Manna[2], **Partha Pakray**[1], **Sivaji Bandyopadhyay**[1]
[1]National Institute of Technology, Silchar, India
[2]Adamas University, Kolkata, India
{sahinurlaskar.nits, rahuljan, abdullahkhilji.nits}@gmail.com
{riyankamanna16,parthapakray,sivaji.cse.ju}@gmail.com

Abstract

In today's time, online users are regularly exposed to media posts that are propagandistic. Several strategies have been developed to promote safer media consumption in Arabic to combat this. However, there is a limited available multilabel annotated social media dataset. In this work, we have used a pre-trained AraBERT twitter-base model on an expanded train data via data augmentation. Our team CNLP-NITS-PP, has achieved the third rank in subtask 1 at WANLP-2022, for propaganda detection in Arabic (shared task) in terms of micro-F1 score of 0.602.

1 Introduction

Communities are significantly impacted by the propagation of rumors, false information, or incomplete information, particularly if the process is spearheaded by the media. In the minds of the target populations, who are consistently the targets of propaganda, the illusion becomes true. Propaganda is regarded as one of the most effective political tools in the modern period and consistently succeeds in drawing sizable populations. Social media is now utilized to distribute propaganda and bogus or misleading news to divert attention away from more pressing problems. Depending on the technology employed, a variety of materials and media are used to spread propaganda. The most modern methods (Vorakitphan et al., 2021) for detecting propaganda are based on language models, which mostly use transformer-based architectures. There are publicly available language models for Arabic, such as AraBERT (Antoun et al., 2020), AraGPT2 (Antoun et al., 2021b) and AraELECTRA (Antoun et al., 2021a). The challenging issue is the requirement of a sufficient amount of annotated multilabel dataset that must cover different varieties or types of propaganda in order to utilize advanced deep learning-based techniques. To encounter this issue, we have increased the train data by augmenting data to the original train set. We have noticed that people often use dictionary/root or stemmed words on social media platforms without adhering to proper grammar. Therefore, we have used Arabic Light Stemmer (Zerrouki, 2012) in the train data and prepared $1,008$ additional synthetic dataset that is directly augmented with the original train data. Moreover, AraBERT (Antoun et al., 2020) twitter-base model is utilized in this work and attains competitive results in terms of standard evaluation metrics (as reported in Section 5).

2 Related Work

In this section, we briefly present the related works of propaganda detection which have been studied recently. By identifying all text fragments that contain propaganda techniques and their type, the authors (Da San Martino et al., 2019) undertake fine-grained analysis of texts. They have contributed a corpus of news articles that are annotated using 18 propaganda techniques at the fragment level and designed a suitable evaluation measure. Also, a multi-granularity neural network is designed and attained better performance than the BERT-based baseline system. In (Dimitrov et al., 2021a), the authors introduce a multi-label multimodal task to detect the different types of propaganda techniques used in memes and release a corpus that includes 950 memes annotated with 22 propaganda techniques. (Dimitrov et al., 2021b) organizes SemEval-2021 task 6 which include subtasks of detecting the persuasion techniques in the text, the text spans where the persuasion techniques are used, and detection of particular technique present in the entire meme (text and image). They explored the benefits of text and image modalities for the detection techniques in the respective shared tasks. Moreover, (Yu et al., 2021) proposed to use of interpretable features with pre-trained language models for detecting deception techniques.

Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP), pages 541 - 544

Train Set	Samples
Before Augmentation	504
After Augmentation	1512

Table 1: Data Statistics of train set (before and after augmentation) in subtask 1.

3 Dataset Description

The dataset[1] used has been provided by the organizers of WANLP 2022 for the shared task on Propaganda Detection in Arabic (Alam et al., 2022). The dataset consists of the text of Arabic tweets and the list of propaganda techniques used in them. There are a total of 21 propaganda techniques, namely, "Appeal to authority", "Appeal to fear/prejudice", "Black-and-white Fallacy/Dictatorship", "Causal Oversimplification", "Doubt", "Exaggeration/Minimisation", "Flag-waving", "Glittering generalities (Virtue)", "Loaded Language", "Misrepresentation of Someone's Position (Straw Man)", "Name calling/Labeling", "Obfuscation", "Intentional vagueness", "Confusion", "Presenting Irrelevant Data (Red Herring)", "Reductio ad hitlerum", "Repetition", "Slogans", "Smears", "Thought-terminating cliché", "Whataboutism", "Bandwagon", and a "no technique" label to indicate no propaganda techniques have been used. The train, validation, and final test set consist of 504, 104, and 440 number of tweets. For data augmentation, for each tweet in the training set, we have used an Arabic Light Stemmer[2] (Zerrouki, 2012) to get the root and stem and obtained synthetic data are added to the training set with the same labels. This brought up the number of training samples to 1512. Table 1 represents augmented train data that is used in this work and Figure 1 presents examples of synthetic data (stem and root).

4 System Description

The AraBERT (Antoun et al., 2020) twitter-base model is utilized for the task of multilabel propaganda classification and used example source code (Antoun et al., 2020)[3] for Text Classification. However, the example code (Antoun et al., 2020) is restricted for single-label classification. To prepare it for multi-label classification, we have changed the input labels to the model to be one hot encoded to indicate multiple labels and modify the macro-F1 scorer to give a score for multiple labels. We used data augmentation; in particular, generated synthetic training data using root and stem substitution from the original train samples and prepared additional synthetic examples. For preprocessing, the default ArabertPreprocessor[4] has been used. During training to get the predicted labels for one tweet, we selected the number of predicted labels corresponding to the number of true labels for that tweet. For training, we have used 0.1 drop-out, Adam optimizer with a default learning rate, and a batch size of 16. The model is trained on a single NVIDIA Quadro P2000 GPU for 5 epochs based on early stopping criteria, i.e, the model training is halted if it does not converge on the validation set for more than 5 epochs. The training process took less than 5 minutes. To make predictions with the model, the sentiment analysis pipeline is used from HuggingFace transformers[5], which returns scores corresponding to each of the labels for a given input. Then we selected all the labels that provide a score greater than or equal to 0.32 as the predicted labels. We observed multiple scores for predictions on the validation test set and found that most correct labels have a score greater than 0.30 and there was a large gap in the score for the labels that have scored less than 0.30.

5 Results

The WANLP 2022 shared task organizer (Alam et al., 2022) published the evaluation result[6] of the propaganda detection in Arabic. The shared task includes two subtasks, namely, Subtask 1: A multilabel classification problem (Given the text of a tweet, identify the propaganda techniques used in it). Subtask 2: A sequence tagging task (Given the text of a tweet, identify the propaganda techniques used in it together with the span(s) of text in which each propaganda technique appears). Herein, we have participated in Subtask 1 with a team named CNLP-NITS-PP and achieved the third ($3rd$) position where a total of fifteen (15) teams participated and four (4) teams participated in Subtask 2. The

[1] https://gitlab.com/arabic-nlp/propaganda-detection/

[2] https://github.com/linuxscout/tashaphyne

[3] https://github.com/aub-mind/arabert/tree/master/examples

[4] https://huggingface.co/aubmindlab/bert-base-arabertv02-twitter

[5] https://colab.research.google.com/drive/19zAYftPaXcNDZ6N6Pyj8K8BJXtkEgglx?usp=sharing

[6] https://sites.google.com/view/propaganda-detection-in-arabic/results?authuser=0

ID	Original Text	Stem	Root
1391667689656102914	RT @AJABreaking: | عاجل حركة حماس: ما يجري في المسجد الأقصى مجزرة حقيقية ستدفع سلطات الاحتلال الإسرائيلية ثمنها	RT @AJABreaking: | عاجل حركة حماس: ما يجري في المسجد الأقصى مجزرة حقيقية ستدفع سلطات الاحتلال الإسرائيلية ثمنها	RT @AJABreaking: | عجل حرك حمس: م جر ف لمسجد لءقصي مجزر حقق ستدفع سلطت لحتلل لءسرءل ثمنه
139257525971126682821	رؤساء البرلمانات العربية يطالبون بتدخل دولي عاجل لوضع حد نهائي لممارسات إسرائيل "الإجرامية غير الإنسانية	رؤساء البرلمانات العربية يطالبون بتدخل دولي عاجل لوضع حد نهائي لممارسات إسرائيل "الإجرامية غير الإنسانية	رءسء لبرلمنت لعرب طلبن بتدخل دل عجل لضع حد نهء لممرست ءسرءل "لءجرم غر لءنسن
138621674411773542 5	▪ الهيئة القيادية العليا لأسرى حماس: تأجيل أو إلغاء الانتخابات سيكون له أثر خطير على شعبنا الفلسطيني وإيمانه بالاحتكام للعملية الديموقراطية.	الهيئة القيادية العليا لأسرى حماس: تأجيل أو إلغاء الانتخابات سيكون له أثر خطير على شعبنا الفلسطيني وإيمانه بالاحتكام للعملية الديموقراطية	لهء لقد لعل لءسري حمس: تءجل ء علغء لنتخبت سكن له ءثر خطر علي شعبن لفلسطن ءمنه بلحتكم للعمل لدمقرط

Figure 1: Examples of synthetic data (stem and root).

Team	Macro-F1	Micro-F1	Rank
mgamal88	0.185	0.649	1
Team_IITD	0.183	0.609	2
CNLP-NITS-PP	**0.068**	**0.602**	**3**
basem	0.068	0.602	3
josephattieh	0.177	0.602	3
gauravsingh	0.105	0.600	4
Team_iCompass	0.191	0.597	5
ArabicProcessors	0.137	0.585	6
mostafa-samir	0.186	0.580	7
SirenAI	0.153	0.578	8
earendil	0.111	0.565	9
mhmud.fwzi	0.087	0.552	10
Mohtaj	0.076	0.494	11
tesla	0.120	0.355	12
Baseline (Random)	0.043	0.079	13

Table 2: Our system's results (marked as bold) and other participants results on subtask 1 propaganda detection in Arabic.

automatic evaluation metric micro-F1 is mainly considered to evaluate the results of different submission teams. However, the task organizer also reports macro-F1. Table 2 presents the results of our system (marked as bold).

6 Discussion

In this work, we have presented preliminary experimental work in subtask 1 only at WANLP 2022 shared task. In future work, we need to explore and examine different deep-learning-based models such as AraGPT2, AraELECTRA, AraXLNet on the same benchmark data set released in this shared task and utilize another dataset, namely, PTC corpus[7] for both tasks, i.e., multilabel classification and sequence tagging tasks. Moreover, we will manually observe the benchmark dataset to identify the clue for the expansion of train data, the novelty in the multilabel annotation, preprocessing, and model training that could be increased accuracy in multilabel classification and sequence tagging tasks.

7 Conclusion

This paper demonstrates our work in subtask 1, propaganda detection in Arabic shared task at WANLP-2022. To handle the data scarcity problem in this shared task, we have proposed to use a data augmentation strategy and utilization of a domain-specific pre-trained language model (AraBERT twitter-base model) that shows remarkable results. This work motivates us to explore propaganda detection in Indian languages which will be beneficial for a multilingual country like India.

Acknowledgements

The authors would like to thank the Center for Natural Language Processing (CNLP), Artificial Intelligence (AI), and the Department of Computer Science and Engineering at the National Institute of Technology, Silchar for providing the requisite support and infrastructure to execute this work.

[7] https://propaganda.qcri.org/semeval2020-task11/

References

Firoj Alam, Hamdy Mubarak, Wajdi Zaghouani, Preslav Nakov, and Giovanni Da San Martino. 2022. Overview of the WANLP 2022 shared task on propaganda detection in Arabic. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop*, Abu Dhabi, UAE. Association for Computational Linguistics.

Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. Arabert: Transformer-based model for arabic language understanding. In *LREC 2020 Workshop Language Resources and Evaluation Conference 11–16 May 2020*, page 9.

Wissam Antoun, Fady Baly, and Hazem Hajj. 2021a. AraELECTRA: Pre-training text discriminators for Arabic language understanding. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 191–195, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Wissam Antoun, Fady Baly, and Hazem Hajj. 2021b. AraGPT2: Pre-trained transformer for Arabic language generation. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 196–207, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeño, Rostislav Petrov, and Preslav Nakov. 2019. Fine-grained analysis of propaganda in news article. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5636–5646, Hong Kong, China. Association for Computational Linguistics.

Dimitar Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. 2021a. Detecting propaganda techniques in memes. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6603–6617, Online. Association for Computational Linguistics.

Dimitar Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. 2021b. SemEval-2021 task 6: Detection of persuasion techniques in texts and images. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 70–98, Online. Association for Computational Linguistics.

Vorakit Vorakitphan, Elena Cabrio, and Serena Villata. 2021. PROTECT - A pipeline for propaganda detection and classification. In *Proceedings of the Eighth Italian Conference on Computational Linguistics, CLiC-it 2021, Milan, Italy, January 26-28, 2022*, volume 3033 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Seunghak Yu, Giovanni Da San Martino, Mitra Mohtarami, James Glass, and Preslav Nakov. 2021. Interpretable propaganda detection in news articles. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1597–1605, Held Online. INCOMA Ltd.

Taha Zerrouki. 2012. Tashaphyne, arabic light stemmer.