Detection of Propaganda Techniques in News Articles

Ankita Dinesh Fatale
ankitadi@buffalo.edu

Nikhil Pentapalli
spentapa@buffalo.edu

1 Abstract

This paper represents a model that can detect a specific text fragment containing propaganda techniques in given news articles.This task is featured into two sub-tasks, Span Identification(SI) and Technique Classification(TC).For SI,we used a pre-trained BERT language model as base architecture and enhance it with tagging schemes established for development of Named Entity Recognition (NER) model.This constructed a system to identify propaganda spans in the text/news articles.For TC, we used contextual feature to pre-train RoBERTa model, that classified the propaganda technique of the input spans choosing from the list of 14 propaganda techniques which might contain multiple propaganda's for single fragment.

2 Introduction

Propaganda is a misleading information used to promote or publicize a particular political cause or point of view.The purpose of propaganda is to influence people's opinion or behavior actively.These techniques are intended to go unnoticed to achieve maximum effect on audience. Our goal is to detect a specific text fragment containing propaganda techniques in given news articles.It is featured in following sub-tasks :

1.Span Identification : Given a plain text document , our model detects those specify fragments which contains at least one propaganda technique and generate spans for these fragments(i.e start and end indexes of the fragments containing propaganda).

2.Technique Classification : Given a text fragment identified as propaganda and its document context, our model identifies the applied propaganda technique in the fragment.

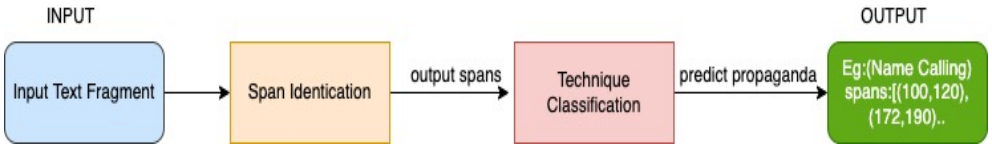

Figure 1: High Level Task

For baseline, we implemented span identification task with BILSTM-CRF along with Propaganda/Non-Propaganda and BIOE tagging schemes, and for the second task which is technique classification we used LSTM but as it is a multi-class classification problem we use categorical cross entropy as our loss function.

As an improvement to our baseline model, for span identification task we made use of state-of-the-art transformer model architecture like BERT along with Propaganda/Non-Propaganda and BIOE tagging scheme to enhance NER model. For the second sub-task,to obtain the contextual sequence representation for the propaganda snippet and perform classification, we employ the RoBERTa transformer architecture.

3 Related Work

There was related previous task which take non-contextualized word embeddings, e.g., based on FastText and GloVe (Gupta et al., 2019; Al-Omari et al., 2019), or handcrafted features such as LIWC, quotes and

questions (Alhindi et al., 2019). For the technique-classification subtask the LSTM-CRF (Gupta et al., 2019) or biLSTM-CRF (Alhindi et al., 2019) models were applied besides BERT (Yu et al., 2019). In some previous works some efforts have been made to balance and increase the dataset to achieve better results.

BERT architecture with hyperparameters tuning without activation function (Mapes et al., 2019). (Yoosuf and Yang, 2019) focused first on the pre-processing steps to provide more information regarding the language model along with existing propaganda techniques, then they employ the BERT architecture casting the task as a sequence labeling problem.

4 Model Architecture

4.1 Span Identification:

The task of span identification is identified as a binary sequence tagging task.For this, each token needs to be classified either as propaganda or not propaganda.To do that, we feed the input sequence to pre-trained transformer architecture like BERT which has inbuilt tokenizers and it generates the embeddings for each token in the input sequence.The generated embeddings are forwarded to the transformer's top layer and the binary classification of whether the token belongs to P/NP is performed .

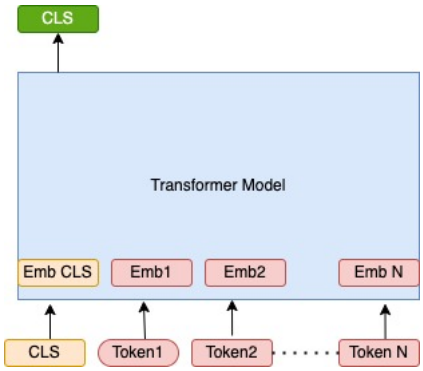

Figure 2: SI Architecture

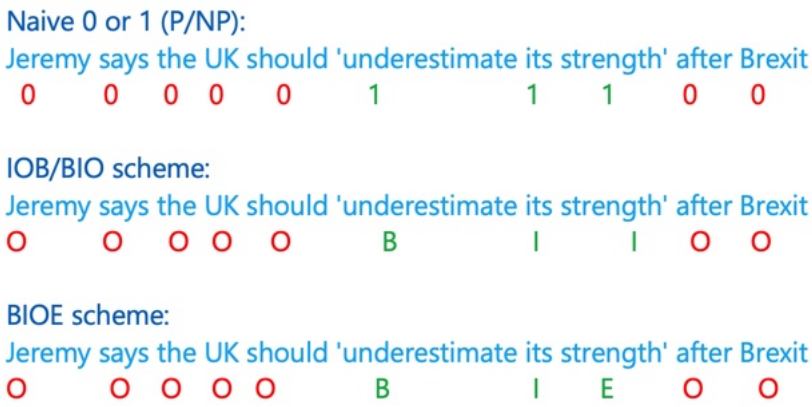

Figure 3: Tagging schemes example

For this approach, we used P/NP and BIOE tagging schemes to implement NER model. P/NP performed token level classification as it identified each token either as propaganda(P) or not propaganda(NP).Whereas BIOE provides proper span identification of propaganda text fragment.For BIOE,each token is either tagged B or I, where B denotes the beginning of propaganda and I, the in-

side of the propaganda. An O indicates all other words which are not propaganda and E indicated end of propaganda text.

4.2 Technique Classification:

Technique Classification is a multi-class sequence classification task where for a given propaganda fragment we need to identify the propaganda or propaganda's that it belongs to. We use the input sequence along with its context and pass both inputs to a pre-trained Transformer model like RoBERTa. The Transformer output sequence representation S and context representation C. Combining both the outputs we get a contextual vector representation V which is then passed to the classifier layer on top which is our final classification.

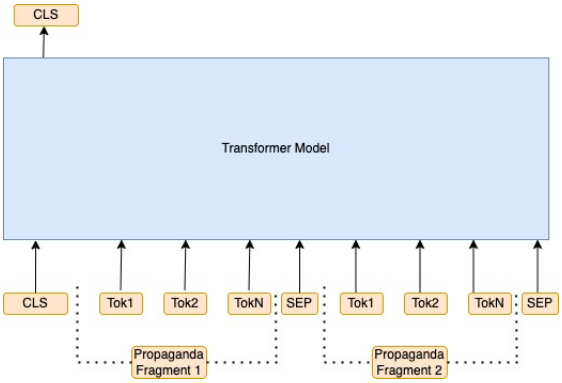

Figure 4: TC Model Architecture

The context around the span can be categorized into multiple levels like sentence level, paragraph level and article level. To get the context of whole paragraph or article would be difficult as they can be long. So we can use title/headline which is the first sentence in the given dataset as our context which can be obtained from CLS presentation which is suggested by Hou and Chen (2019). Another effective approach would be of sentence level. In order to capture sentence context , we capture all the sentences that the current propaganda fragment is spread across which might lead to multiple sentences to be considered. Also we limit the sentence context length to 150 words due to the above mentioned problem of length. We can make use of this length which would be more than sufficient to infer the meaning.

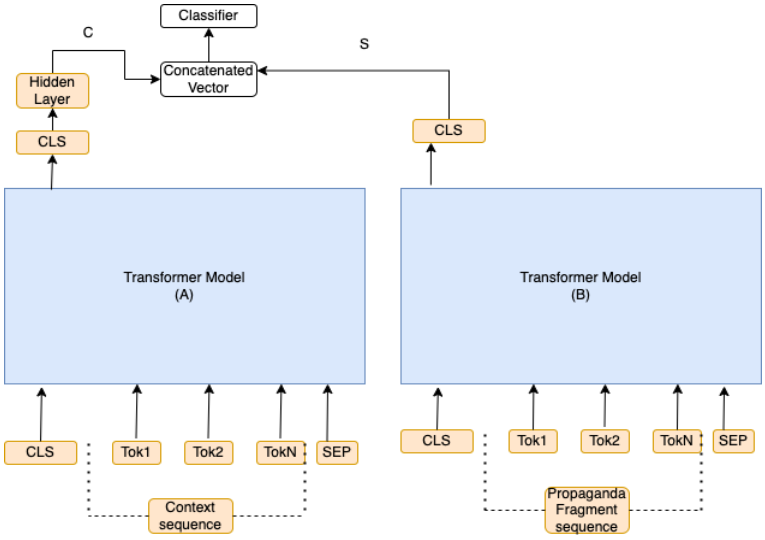

Figure 5: TC Sentence and Context Architecture

Initially, we experimented by taking both sentence and context vectors and contacting them and passing to the transformer model. The problem with this approach was that the fragments that consists of

less words (up to 3 words) takes the longer context text which influences the final representation rather than the these fragments.

So, in order to overcome this problem both contextual and sequence representations are passed independently to different Transformers. We add an additional hidden layer which helps reduce the dimension of the context vector as well as give more attention to the propaganda sequence.

4.3 Data Preprocessing:

From the given dataset we split the articles to sentences which resulted in 17800 sequences with max sequence length of 179 words. Using the characters level spans labels we convert them to token labelled as propaganda or not propaganda. In our training dataset there are consecutive tokens with same labels.so they are merged to get final character level spans.

In case of technique classification, using the provided characters spans and the labels we try to get the surrounding sentence context. We limit the maximum words to 150 i.e to select words from both sides of the propaganda until the sentence reaches end or exhausts the defined word limit of 150. Our training data consists of (S,C) sentence and context pair and label as propaganda technique/techniques.

5 Results

5.1 Results on Development Set

5.1.1 SI Task

Model	LR-scheduler	Epochs	F1 score	Precision	Recall
BILSTM CRF(Baseline)	None	10	0.31	0.22	0.52
BERT P/NP(Baseline BERT)	None	4	0.33	0.30	0.36
BERT BIO	None	5	0.38	0.36	0.41
BERT BIO	Yes	5	0.42	0.38	0.46
BERT BIOE	None	4	0.44	0.40	0.48
BERT BIOE	Yes	5	0.49	0.42	0.59

For SI task, we achieved a F1 score of 0.49 for development dataset using the BERT BIOE model. For SI baseline, we used BILSTM CRF model to train a custom NER model, and we achieved a F1 score of 0.31 which was quite low.Hence , BERT BIOE model shows a lot of improvement.

5.1.2 TC Task

Model	LR-scheduler	Epochs	F1 score
LSTM(Baseline)	Yes	10	0.44
BERT(Baseline)	None	4	0.52
BERT Exp1	None	5	0.54
BERT Exp2	Yes	5	0.55
RoBERTa Exp1	None	4	0.58
RoBERTa Exp2	Yes	5	0.61

In Exp1, the input sentence and context are feed to single transformer sequentially whereas in Exp2-they are passed to separate pre-trained transformers and the output from both is concatenated at the end.

For baseline ,we have taken only sequence of text as input and use LSTM to do the multi-class classification and we have achieved a F1 score of 0.44. Our state-of-the-art model RoBERTa with Exp2 achieved F1 score of 0.61 on development dataset.

5.2 Results on Test Set

5.2.1 SI Task

For the SI Task, the BERT BIOE performed the best in terms of F1 score and prediction.txt is generated and F1 score of 0.451 achieved. The results are submitted to the leaderboard of propaganda semeval task11

5.2.2 TC Task

For the TC Task, the RoBERTa Exp2 performed the best in terms of F1 score and prediction.txt is generated and submitted where we achieved F1 score of 0.574.The results are submitted to the leaderboard of propaganda semeval task11.

6 Discussion and Error Analysis

The main idea of SI Task is to detect propaganda spans for which we use the architecture shown above in Figure 2 . The input to the transformer would be Bert Tokens after sentence is tokenized. We use pre-trained transformer models provided by hugging face which uses pytorch framework. We use bert-base-uncased along with different tagging schemes like BIO,BIOE and naive P/NP tagging schemes for the span identification task. For the technique classification, we use Roberta-uncased along with different tagging schemes. For the above task, we fine tuned the pre-trained transformer models with the following parameters : Epochs=5 and learning rate of 2*10-5 and a batch size of 8.

6.1 Error analysis

Initially we for SI task we experimented with BILSTM and then to improve the F1 score we move to BERT architecture where we tried with different epochs and used LR scheduler to improve the model F1 score. After multiple iterations and hyperparameter tuning we achieved our best model as shown in the results.

For TC task we started out with LSTM without taking context separately.As context is important we then utilized Bert pretrained model for sentence as well as context. we have made two different experiments as described above in results and we achieved the best F1 score of 0.61 with epochs=5 and using learning rate scheduler.

6.2 Area's of Improvement

Instead of treating the tasks separately we can use the output spans generated from the Task1(span identification) and pass it to Task2(Technique classification) to predict end to end which gives the spans as well as the propaganda.

For the first task, we can introduce RoBERTa with 5 epochs and also take a propaganda vocabulary to use that as advantage since the dataset size is not huge and is imbalanced this would help us give more weightage to the spans containing the propaganda.

For the second task, we can experiment with the size of the hidden layers for context vector and use different dropouts as to see model not over-fitting over the training data.

6.3 Data Imbalance

As we can see there is an imbalance in the dataset which is show in Figure 6. For techniques like Loaded Language and Name Calling , Labelling which has many examples in the dataset the F1 score obtained was around 0.65 to 0.7. There are some minority classes (Whataboutism,Straw Men,Red Herrin) and other combined techniques where the F1 scores were quite low which are around 0.1 or less.

7 Conclusion

We proposed two models, one for detecting the fragments of propaganda and another one for identifying the technique of the propaganda. We used the pre-trained Transformers language models for both tasks and modified them according to the task. We also used different tagging schemes like BIO ,BIOE which helps model learn more context and improve the F1 score.

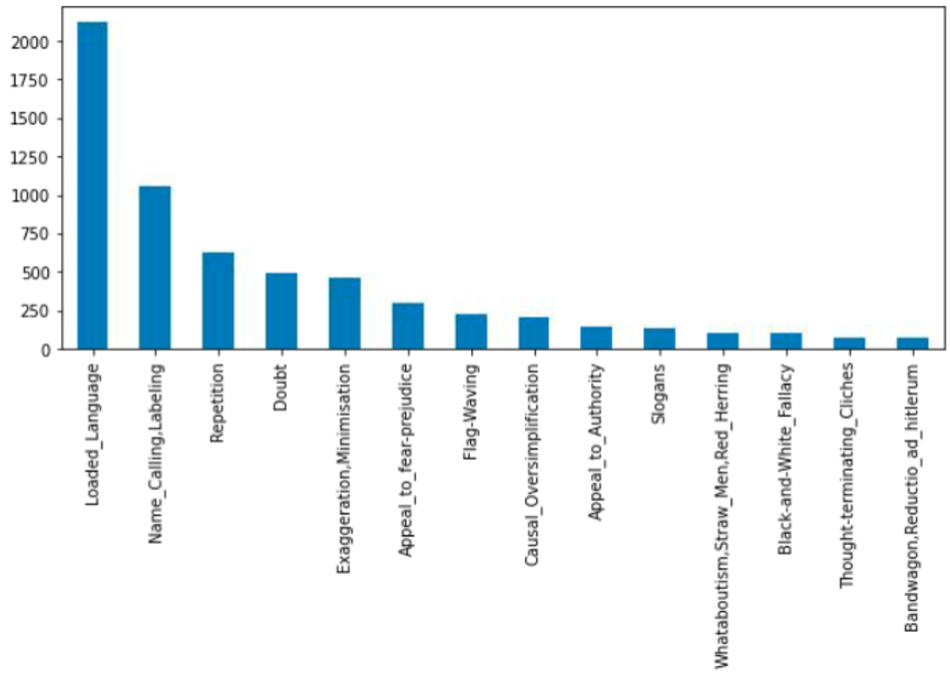

Figure 6: Dataset Imbalance

8 Work Distribution

Tasks	Ankita	Nikhil
SI Baseline	40	60
TC Baseline	60	40
SI Improvement	50	50
TC Improvement	50	50
Report	50	50

We have made equal contribution to the above tasks.

References

Yinhan Liu and Myle Ott and Naman Goyal and Jingfei Du and Mandar Joshi and Danqi Chen and Omer Levy and Mike Lewis and Luke Zettlemoyer and Veselin Stoyanov. 2019. *Roberta: A robustly optimized BERT pretraining approach*. CoRR,abs/1907.11692.

Giovanni Da San Martino and Alberto Barron-Cedeno,and Henning Wachsmuth and Rostislav Petrov, and Preslav Nakov. 2020. *SemEval-2020 task 11: Detection of propaganda techniques in news articles*. Proceedings of the 14th International Workshop on Semantic Evaluation, SemEval 20202. September Barcelona,Spain

Jacob Devlin and Ming-Wei Chang,and Kenton Lee, and Kristina Toutanova. 2018. *BERT: pre-training of deep bidirectional transformers for language understanding.*. CoRR, abs/1810.04805.

Giovanni Da San Martino, Alberto Barrn-Cedeo, and Preslav Nakov. 2019. *Findings of the nlp4if-2019 shared task on fine-grained propaganda detection.*.

Thomas Wolf, Julien Chaumond, Lysandre Debut, Victor Sanh, Clement Delangue, Anthony Moi, Pierric Cistac, Morgan Funtowicz, Joe Davison, Sam Shleifer, et al. 2020. 2020. *Transformers: State-of-theart natural language processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pages 38–45.*

Mohammad Sadegh Rasooli and Joel R. Tetreault. 2015. Yara parser: A fast and accurate dependency parser. Dawid Jurkiewicz, Łukasz Borchmann, Izabela Kosmala, and Filip Gralinski. 2020. ´semeval-2020 task 11: On roberta-crf, span cls and whether self-training helps them arXiv preprint arXiv:2005.07934.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov.2014

Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research 15(1):1929–1958.