Sentence-Level Propaganda Detection Using BERT with Context-Dependent Input Pairs

Wenjun Hou, Ying Chen
College of Information and Electrical Engineering, China Agricultural University, China
{houwenjun, chenying}@cau.edu.cn

Abstract

The goal of fine-grained propaganda detection is to determine whether a given sentence uses propaganda techniques (sentence-level) or to recognize which techniques are used (fragment-level). This paper presents the system of our participation in the sentence-level subtask of the propaganda detection shared task. In order to better utilize the document information, we construct context-dependent input pairs (sentence-title pair and sentence-context pair) to fine-tune the pretrained BERT, and we also use the undersampling method to tackle the problem of imbalanced data[1].

1 Introduction

Propaganda detection is a process of determining whether a news article or a sentence is misleading. Several research works have been proposed to detect propaganda on document-level (Rashkin et al., 2017; Barrón-Cedeño et al., 2019b), sentence-level and fragment-level (Da San Martino et al., 2019). Sentence-level detection or classification (SLC) is to determine whether a given sentence is propagandistic and it is a special binary classification problem, while the goal of fragment-level classification (FLC) is to extract fragments and assign with given labels such as *loaded language*, *flag-waving* and *causal oversimplification*, and it could be treated as a sequence labeling problem.

Compared with document-level, sentence-level and fragment-level detection are much more helpful, since detection on sentences and fragments are more practical for real-life applications. However, these fine-grained tasks are more challenging. Although Da San Martino et al. (2019) indicates that multi-task learning of both the SLC and the FLC could be beneficial for the SLC, in this paper, we only focus on the SLC task so as to better investigate whether context information could improve the performance of our system. Since several pretrained language models (Devlin et al., 2019; Liu et al., 2019) have been proved to be effective for text classification and other natural language understanding tasks, we use the pretrained BERT (Devlin et al., 2019) for the SLC task. This paper elaborates our BERT-based system for which we construct sentence-title pairs and sentence-context pairs as input. In addition, in order to tackle the problem of imbalanced data, we apply the undersampling method (Zhou and Liu, 2006) to the training data, and we find that this method greatly boosts the performance of our system.

2 Related Work

Various methods have been proposed for propaganda detection. Rashkin et al. (2017) proposed to use LSTM and other machine learning methods for deception detection in different types of news, including *trusted*, *satire*, *hoax* and *propaganda*. Barrón-Cedeño et al. (2019b) proposed to use Maximum Entropy classifier (Berger et al., 1996) with different features replicating the same experimental setup of Rashkin et al. (2017) for two-way and four-way classifications. A fine-grained propaganda corpus was proposed in Da San Martino et al. (2019) which includes both sentence-level and fragment-level information. Based on this corpus and the pretrained BERT which is one of the most powerful pretrained language model, a multi-granularity BERT was proposed and it outperformed several strong BERT-based baselines.

3 Methodology

In our system, we utilize BERT as our base model and construct different kinds of input pairs to fine-tune it. When constructing the input representa-

[1] Code is available at https://github.com/Wenjun-Hou/Propaganda-Detection-SLC

Proceedings of the 2nd Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda, pages 83–86
Hong Kong, China, November 4, 2019 ©2019 Association for Computational Linguistics

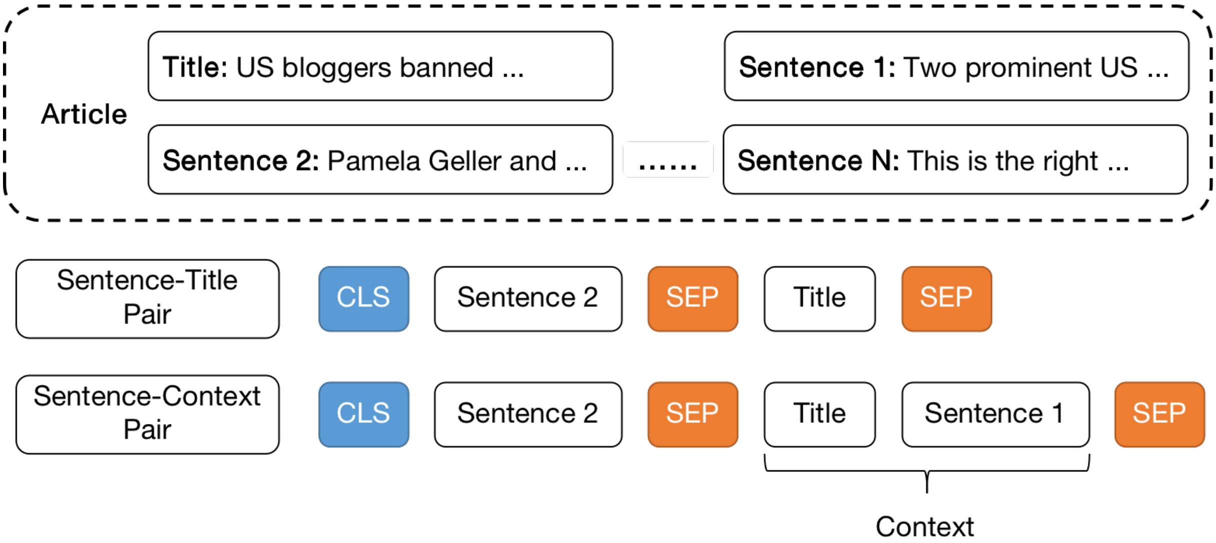
Article
Title: US bloggers banned ...
Sentence 1: Two prominent US ...
Sentence 2: Pamela Geller and ...
......
Sentence N: This is the right ...
Sentence-Title Pair
CLS
Sentence 2
SEP
Title
SEP
Sentence-Context Pair
CLS
Sentence 2
SEP
Title
Sentence 1
SEP
Context

Figure 1: Two kinds of input pairs for BERT. [CLS] and [SEP] are two special tokens.

tion, a special token [CLS] is padded in front of every sentence and another token [SEP] is added at the end of it. In addition, for each input pair, a [SEP] is added between a sentence and its context or title. Finally, a linear layer and a *sigmoid* function are applied to the final representation of [CLS] to obtain the probability for classification. For comparison, we also use the official method (Random) as baseline which randomly labels sentences.

3.1 Data

The dataset is provided by NLP4IF 2019 Shared Task (Barrón-Cedeño et al., 2019a), and the training set, the development set, and the test set contain approximately 16,000, 2,000 and 3,400 sentences respectively. According to the statistics, only 29% of the training sentences are labeled as propaganda, and thus in this paper, we treat propaganda sentences as positive samples and non-propaganda sentences as negative samples. More details of the dataset could be found in Da San Martino et al. (2019).

3.2 Input pairs

Sentence Only: We only use the current sentence to fine-tune the model and models trained with this kind of input are used as baselines for those models trained with the following two kinds of input pairs.

Sentence-Title Pair: As described in Da San Martino et al. (2019), the source of the dataset that we use is news articles, and since the title is usually the summarization of a news article, we use the title as supplementary information.

Sentence-Context Pair: In addition to setting the title as the supplementary information, we construct the sentence-context pair which also includes preceding sentences as additional context, since preceding sentences usually convey the same or related events and this historical content is closely related to the current sentence. Figure 1. shows the details of this kind of input pair in which the preceding sentence and the title are directly concatenated.

3.3 Undersampling

As mentioned above, there are only 29% of training sentences labeled as propaganda (positive). In order to tackle the problem of imbalanced data, we first collect positive samples which size is S_{pos} and negative samples, then we resample S_{neg} (X percent of S_{pos}) from negative samples at the beginning of each training epoch. Finally, we combine and shuffle both positive samples and sampled negative samples as a new training set $S_{sampled}$.

$$S_{neg} = X * S_{pos} \quad (1)$$
$$S_{sampled} = S_{neg} + S_{pos} \quad (2)$$

3.4 Experiment Details

In this paper, we use the pretrained uncased version of BERT$_{\text{BASE}}$ and BERT$_{\text{LARGE}}$ [2] for the SLC, and more details of these two models could

[2] https://github.com/google-research/bert

Model	Input	Sample Rate	Precision	Recall	F_1
Random	-	-	44.38	50.74	47.35
BERT$_{\text{BASE}}$	Sentence Only	-	72.76	52.77	61.18
	Sentence-Title	-	70.54	56.70	62.87
		0.8	57.83	77.94	66.40
		0.9	60.77	70.64	65.33
		1.0	63.70	68.88	66.19
	Sentence-Context	-	71.10	54.94	61.98
		0.8	57.53	77.54	66.05
		0.9	60.95	73.07	66.46
		1.0	63.44	66.44	64.90
BERT$_{\text{LARGE}}$	Sentence Only	-	**73.19**	50.61	59.84
	Sentence-Title	-	71.23	54.26	61.60
		0.8	58.69	75.37	66.00
		0.9	61.89	64.82	63.31
		1.0	60.85	71.31	65.67
	Sentence-Context	-	71.88	49.12	58.36
		0.8	59.43	**79.30**	**67.94**
		0.9	63.73	66.58	65.12
		1.0	62.28	73.07	67.25

Table 1: Experiment results of different models on the SLC task, and the model with the highest F_1 score which has been underlined is chosen to be evaluated on the test set. '-' in sample rate means undersampling is not used.

Model	Data	Prec.	Rec.	F_1
Random	Dev.	44.38	50.74	47.35
	Test	38.80	49.42	43.47
BERT$_{\text{LARGE}}$	Dev.	59.43	79.30	67.94
	Test	51.81	74.44	61.10

Table 2: Experiment results of the chosen model and the random baseline for the SLC task.

be found in Devlin et al. (2019). Before fine-tuning, sentences are first converted to lower case and their maximum sequence length is set to 128. For a sentence-context pair, the maximum length of context is set to 100. If the sequence length of an input pair exceeds 128, then the context or title is truncated to meet the length.

When fine-tuning, we use the Adam (Kingma and Ba, 2014) with learning rate 2e-5 for 2 epochs, the batch size is 32 and the dropout probability is kept at 0.1. Since the title or context information could help improve the performance, we only apply the undersampling method to input pairs (sentence-title and sentence-context). For those models involved with undersampling, the sample rate X is set to 0.8, 0.9 or 1.0 empirically. During the training stage, all training samples are used. We directly evaluate all the models on the development set, and the best model is chosen to generate predictions of the test data.

4 Result

Our approach is evaluated on Propaganda Detection@NLP4IF SLC dataset. In the development stage, we use three kinds of input and three different sample rates for BERT. Table 1. shows the results of the development set. From Table 1., without considering undersampling, we can see that using the sentence-title pair could boost the performance of BERT$_{\text{BASE}}$, compared with the model using only the current sentence and the random baseline. While using the sentence-context pair could improve the F_1 score of BERT$_{\text{BASE}}$ by 0.8% with precision rising to 71.10 and recall decreasing to 54.94, the performance of BERT$_{\text{BASE}}$ drops by around 1% with recall dropping significantly to 49.12.

We also observe that both performances of BERT$_{\text{BASE}}$ and BERT$_{\text{LARGE}}$ trained with original training sentences are competitive compared with the random baseline. However, the precision of BERT$_{\text{BASE}}$ at 70.54 and the one of BERT$_{\text{LARGE}}$ at 71.23 are significantly higher than the recall of both models, at 56.70 and at

54.26 respectively, and this may result from the problem of imbalanced instances. Thus, we introduce the undersampling technique using 0.8, 0.9 or 1.0 sample rate to tackle this issue. We observe from Table 1. that the F_1 score of $\mathrm{BERT}_{\mathrm{BASE}}$ with the sentence-title pair and 0.8 sample rate rises around by 5% and the same model using the sentence-context pair and 0.9 sample rate performs similarly. As for $\mathrm{BERT}_{\mathrm{LARGE}}$, while using the sentence-title pair has the similar performance as it is employed in the base version model, using the sentence-context pair strongly boosts the F_1 score, at 67.94 with 0.8 sample rate and at 67.25 with 1.0 sample rate. In addition, it is worth noting that there is a better trade-off between precision and recall with 1.0 sample rate than the one with 0.8.

In the test stage, since we are only allowed to submit a single run on the test set, we choose the model with the highest F_1 score (67.94) to generate predictions and the evaluated results are listed in Table 2. From Table 2., we can see that the recall raises by nearly 5% and the precision of it drops significantly, by around 7%, compared with the results on the development set, while the recall of Random Baseline also drops by approximately 5.5% and the precision of it remains nearly the same.

5 Conclusion and Future Work

In this paper, we examine capability of the context-dependent BERT model. In the sentence-level propaganda detection task, we construct sentence-title pairs and sentence-context pairs in order to better utilize context information to improve the performance of our system. Furthermore, the undersampling method is utilized to tackle the data imbalanced problem. Experiments show that both sentence-title/context pairs and the undersampling method could boost the performance of BERT on the SLC task.

In the future, we plan to apply multi-task learning to this context-dependent BERT, similar to the method mentioned in Da San Martino et al. (2019) or introducing other kinds of tasks, such as sentiment analysis or domain classification.

References

Alberto Barrón-Cedeño, Giovanni Da San Martino, and Preslav Nakov. 2019a. Findings of the nlp4if-2019 shared task on fine-grained propaganda detection. In *Proceedings of the 2nd Workshop on NLP for Internet Freedom (NLP4IF): Censorship, Disinformation, and Propaganda*, NLP4IF EMNLP 2019.

Alberto Barrón-Cedeño, Israa Jaradat, Giovanni Da San Martino, and Preslav Nakov. 2019b. Proppy: Organizing the news based on their propagandistic content. *Information Processing & Management*, 56(5):1849–1864.

Adam Berger, Stephen Della Pietra, and Vincent Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*.

Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeño, Rostislav Petrov, and Preslav Nakov. 2019. Fine-grained analysis of propaganda in news articles. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing*, EMNLP-IJCNLP 2019.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Fine-grained analysis of propaganda in news articles. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL 2019.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference for Learning Representations*, ICLR 2014.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. In *arXiv preprint arXiv:1409.0473*.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing*, EMNLP 2017.

Zhi-Hua Zhou and Xu-Ying Liu. 2006. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Trans Knowledge and Data Engineering*, 18(1):63–77.