

## Sentence BERT Embedding on Hyperpartisan News Classification Task

Jungwoo Lim<sup>1</sup>, Yoon-na Jang<sup>1</sup>, Umadevi G R<sup>1</sup>, Heuseok Lim<sup>1</sup>

<sup>1</sup> Korea University, Republic of Korea  
{zxzxzxzx803, dbssk0604, umagr, limhseok} @korea.ac.kr

**Abstract.** Hyperpartisan news detection task is a classification task which requires a model that discriminates hyperpartisan articles from balanced articles. Though feature-based model ELMo + CNN had been proposed previously, the limitation that the model used average vector of word embedding instead of sentence embedding still remains. We proposed feature-based model using Sentence BERT that utilizes sentence embedding for hyperpartisan news classification task. We showed our model performance with our experiments using ELMo, BERT, Sentence BERT embedding utilized with CNN classifiers. In our experiments, the model showed 81% f1-score which is higher than state-of-the-art model.

**Keywords:** Hyperpartisan News Classification, Sentence BERT, Sentence embeddiing

### 1 Introduction

An article that exhibits blind, prejudiced, or unreasoning allegiance to one party, faction, cause, or person is called hyperpartisan news. As fake news and hyperpartisan news are recently increasing, those people hearing these news have difficulty in maintaining a balanced perspective. Therefore, a number of studies on this hyperpartisan news detection are being conducted along with the increased needs of an automatic news classification system. [1] Although the methodologies in the hyperpartisan news detection vary from knowledge-based [2, 3, 4] to style-based [5], an approach using deep learning [6] is significantly insufficient until diverse deep learning methods went out on SemEval2019 shared task 4<sup>1</sup> [8]. Jiang et al., who achieved the best accuracy score in the competition introduced the model that uses embedding of each article from pre-trained ELMo as an input of convolutional neural networks (CNNs). The accuracy score is followed by Hanawa et al. [9], who showed linear classifier using various features including BERT embedding. As the language models that understand contextual information such as OpenAI GPT [10] or BERT [11] outperformed in several tasks, most of the experiments were conducted using these language models. [12, 13]

---

<sup>1</sup> <https://pan.webbis.de/semEval19/>

The methods of applying these language models to other tasks are divided into two approaches [11], which are feature-based and fine-tuning. Surprisingly, the models with high performance in the hyperpartisan news detection were mostly feature-based. The reason why fine-tuning models had poor performance is that the long sequence of data used in the task keep fine-tuning models from capturing contextual information. Based on this finding, we assume that the task needs to be solved with a feature-based approach. Therefore, we decide to employ BERT with a feature-based approach as BERT obtained state-of-the-art results in eleven natural language processing tasks.

Previous studies using BERT sentence embedding as a feature, mostly created the fixed-size sentence embedding by averaging the value of the output layer in BERT or by using [CLS] token. [9, 14, 15] Still, feature-based BERT embedding of sentences showed poor performance in hyperpartisan news detection. To overcome this problem, we propose Sentence BERT embedding [16] which can be used a feature in the model. We conducted our experiment using Sentence BERT embedding and CNN classifiers.

## **2 Model**

In our research, we used Sentence BERT, which is modified to extract meaningful sentence embeddings from pre-trained BERT. Sentence embeddings from Sentence BERT can be compared to each other with cosine similarities [16]. To get fixed size sentence embeddings, the model is trained as follows. In the first step, mean value of the output vectors from BERT is evaluated. With the mean value, cosine similarities of sentence embeddings are calculated, and weights of embeddings are updated by siamese networks or triplet networks. The distance between embeddings which have similar meanings is closer to each other in the vector space.

We extracted these embeddings using pre-trained Sentence BERT, and used them as inputs of CNN classifier. As an article has different number of sentences, we padded each sequence. The size of filter is [3, 4, 5] and the dimension of each filter is 768. Used optimizer is Adam optimizer and the learning rate is  $3e-5$ . Weighted cross-entropy loss is used for loss function and dropout is 0.2.

## **3 Experiments**

### **3.1 Data**

In SemEval Shared Task 4 in 2019, “by-article” data was released as training data with XML format. As test data is not provided, we used 10% of the training data as test data. The number of the training data and test data is 432 and 213 correspondingly.

### **3.2 Data Preprocessing**

As the given data was composed of several sentences in one article, we preprocessed the data to use it as an input of the CNN model. In the first step, each sentence of the

articles entered into Sentence BERT and come out as sentence embeddings in 768 dimensions. Afterwards, sentence embeddings which corresponds to each sentence are added. Then, each article has embedding sequences with identical number of sentences.

### 3.3 Experiments

As can be seen from Table 1, using BERT embedding as a feature-based method reduces the performance regardless of the type of classifiers. In this study, we implemented and evaluated CNN with ELMO embedding, which showed the highest performance among existing models, it showed similar performance with suggested performance. While CNN classifier with ELMO embedding showed 80.6% of accuracy, Sentence BERT showed the highest accuracy with 81% to CNN. These results suggest that replacing BERT, which is hard to be used as feature-based method, with Sentence BERT embedding increases the performance compared to using existing models with ELMO embedding.

**Table 1.** Experiment

Model	Accuracy	Precision	Recall	F1-score
BERT + CNN	0.629	0.627	0.633	0.630
ELMo + CNN [8]	0.806	0.801	0.775	0.788
Sentence BERT + CNN	0.810	0.797	0.805	0.801

## 4 Conclusion

To classify hyperpartisan news with feature-based approach, Sentence BERT embedding was used in this research. We experimented on CNN with ELMO, BERT and SBERT to prove the performance of the model suggested. Based on the results, our model performed 1.3%p higher than existing state-of-the-art models. In this study, we analyzed several methods that show good performance on hyperpartisan news classifier, and adopted Sentence BERT which is a feature-based method. This approach would have better performance, when applied with other classifier models like CNN.

## 5 Acknowledgement

“This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2018-0-01405) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)”

**The 3<sup>rd</sup> International Conference on Interdisciplinary research on  
Computer science, Psychology, and Education (ICICPE' 2019)  
December 17-19, 2019. Phu Quoc, Vietnam**

## References

1. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36.
2. Wu, Y., Agarwal, P. K., Li, C., Yang, J., & Yu, C. (2014). Toward computational fact-checking. *Proceedings of the VLDB Endowment*, 7(7), 589-600.
3. Shi, B., & Weninger, T. (2016, April). Fact checking in heterogeneous information networks. In *Proceedings of the 25th International Conference Companion on World Wide Web* (pp. 101-102). International World Wide Web Conferences Steering Committee.
4. Lee, N., Wu, C. S., & Fung, P. (2018). Improving large-scale fact-checking using decomposable attention models and lexical tagging. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 1133-1138).
5. Wang, W. Y. (2017). "liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*.
6. Horne, B. D., & Adali, S. (2017, May). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Eleventh International AAAI Conference on Web and Social Media*.
7. Kiesel, J., Mestre, M., Shukla, R., Vincent, E., Adineh, P., Corney, D., ... & Potthast, M. (2019, June). Semeval-2019 task 4: Hyperpartisan news detection. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 829-839).
8. Jiang, Y., Petrak, J., Song, X., Bontcheva, K., & Maynard, D. (2019, June). Team Bertha von Suttner at SemEval-2019 Task 4: Hyperpartisan News Detection using ELMo Sentence Representation Convolutional Network. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 840-844).
9. Hanawa, K., Sasaki, S., Ouchi, H., Suzuki, J., & Inui, K. (2019, June). The Sally Smedley Hyperpartisan News Detector at SemEval-2019 Task 4. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 1057-1061).
10. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. URL [https://s3-us-west-2.amazonaws.com/openai-assets/researchcovers/languageunsupervised/language understanding paper. pdf](https://s3-us-west-2.amazonaws.com/openai-assets/researchcovers/languageunsupervised/language%20understanding%20paper.pdf).
11. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
12. Lee, N., Liu, Z., & Fung, P. (2019, June). Team yeon-zi at SemEval-2019 Task 4: Hyperpartisan News Detection by De-noising Weakly-labeled Data. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 1052-1056).
13. Ning, Z., Lin, Y., & Zhong, R. (2019, June). Team Peter-Parker at SemEval-2019 Task 4: BERT-Based Method in Hyperpartisan News Detection. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 1037-1040).
14. Hassan, H. A. M., Sansonetti, G., Gaspiretti, F., Micarelli, A., & Beel, J. (2019). BERT, ELMo, USE and InferSent Sentence Encoders: The Panacea for Research-Paper Recommendation?.
15. Miller, D. (2019). Leveraging BERT for Extractive Text Summarization on Lectures. *arXiv preprint arXiv:1906.04165*.
16. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv preprint arXiv:1908.10084*.