Quick Task

Ward Bradt, bwb82@cornell.edu

February 6, 2021

1 Solution

After reading various papers on similar tasks [1] [2] [3], [4] [5], I settled on one of the ideas discussed in [5]. I used a multilingual implementation of BERT to encode Chinese-English source-candidate sentence pairs that were fed into a neural network. I got this idea from [5], which found this architecture to be one of the best for detecting neural machine translation. Due to having less training data than [5], I achieved less accuracy. I tried to counteract this by finding another suitable Chinese-English parallel corpus online for training, but was not able to due to time constraints. In my implementation, I used the simpletransformers Python library. I used the following hyperparameters: sequence length of 128, batch size of 8, Adam optimizer, learning rate of $1 \times e^{-5}$, `adam_epsilon` of $1 \times e^{-8}$. The implementation for my solution can be found at https://github.com/wardbradt/MT-Detection-Task.

2 Results

On the test set, I achieved the following results: precision of 0.632, recall of 0.524, an average F1 score of 0.625, and an unbalanced accuracy of 63.2%. I believe that these metrics would be improved if I were given more time to fine-tune the hyperparameters, access more data, and/ or try different sentence encoders.

References

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

[2] Roee Aharoni, Moshe Koppel, and Yoav Goldberg. Automatic detection of machine translated text and translation quality estimation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 289–295, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[4] David Kurokawa, Cyril Goutte, and P. Isabelle. Automatic detection of translated text and its impact on machine translation. 2009.

[5] Shivendra Bhardwaj, David Alfonso Hermelo, Phillippe Langlais, Gabriel Bernier-Colborne, Cyril Goutte, and M. Simard. Human or neural translation? In *COLING*, 2020.