Chinese Grammatical Error Detection Using Adversarial ELECTRA Transformers

Lung-Hao LEE[a], Man-Chen HUNG[a], Chao-Yi CHEN[a], Rou-An CHEN[a] & Yuen-Hsien TSENG[b*]
[a]*Department of Electrical Engineering, National Central University, Taiwan*
[b]*Graduate Institute of Library and Information Studies, National Taiwan Normal University, Taiwan*
*samtseng@ntnu.edu.tw

Abstract: We explore transformer-based neural networks for Chinese grammatical error detection. The TOCFL learner corpus is used to measure the model capability of indicating whether a sentence contains errors or not. Experimental results show that ELECTRA transformers which take into account both transformer architecture and adversarial learning technique can achieve promising effectiveness with an improvement of F1-score.

Keywords: Grammatical error diagnosis, adversarial learning, transformers, neural networks

1. Introduction

Chinese learners may make various kinds of grammatical errors, such as missing words, redundant words, incorrect word selection, or word ordering error, during their language acquisition process. An automated system able to detect such errors would facilitate Chinese learning. Previous Chinese grammatical error detection approaches were based on linguistic rules (Lee et al., 2013), machine learning classifiers (Liu et al., 2016), or their hybrid methods (Lee et al., 2014). Deep learning approaches had also been applied to detect Chinese grammatical errors (Lee et al, 2017; 2020). Recently, novel transformer-based network architectures (e.g., BERT, RoBERTa, and XLNet) achieve dominating results in many natural language processing tasks. This trend motivates us to explore transformer-based neural networks to detect Chinese grammatical errors.

This study describes our application of ELECTRA transformer architecture (Clark et al., 2020) for Chinese grammatical error detection. The TOCFL learner corpus (Lee et al., 2018) is used to evaluate the performance. Compared with previous approaches on the same dataset, ELECTRA transformers achieved an impressive improvement of F1-score which considers both detection precision and recall at the same time.

2. ELECTRA Transformers

Figure 1 illustrates our adapted ELECTRA model (Clark et al., 2020) for Chinese grammatical error detection. ELECTRA (Efficiently Learning as Encoder that Classifiers Token Replacements Accurately) is a new pre-training approach that aims to match or exceed the downstream performance of a Masked Language Modeling (MLM) pre-trained model while using less computational loading (Clark et al., 2020). During the training phase, ELECTRA trains two transformer models: the generator, which replaces the tokens in a sequence for training a masked language model; and the discriminator, which tries to identify which tokens in the sequence were replaced by the generator. If a sentence contains at least one grammatical error judged by a human, its class is labeled as 1, and 0 otherwise. The Word2Vec embedding (Mikolov et al., 2013) is used to represent sentences. All the sentences with their labeled classes are used to train our adapted ELECTRA model to automatically learn all the corresponding parameters. To classify a sentence during the testing phase, the sentence unseen in the training phase goes through the ELECTRA architecture to yield a probability value corresponding to each class. The class with the larger probability will be returned as the prediction result.

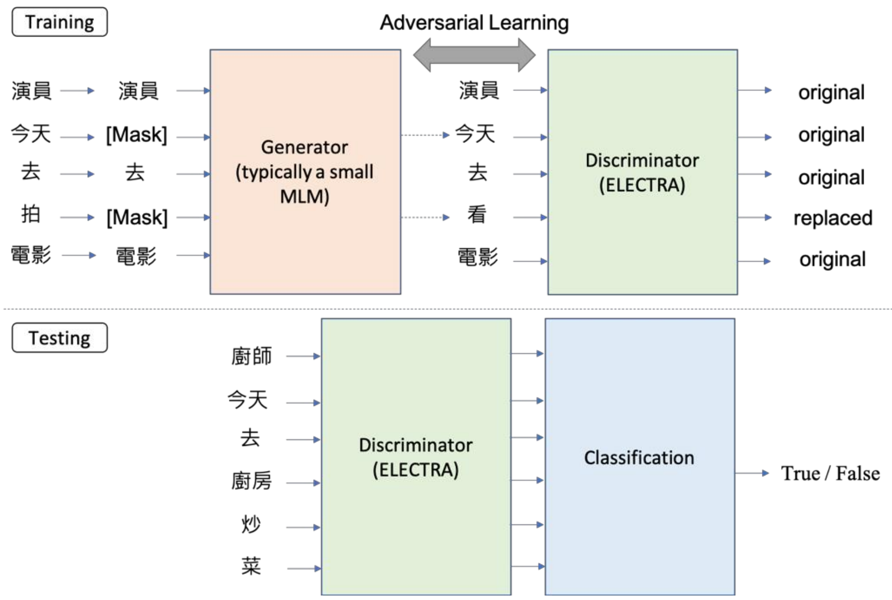

Figure 1. The illustration of ELECTRA Transformers for Chinese Grammatical Error Detection.

3. Experiments and Results

The experimental data came from the TOCFL learner corpus (Lee et al., 2018), including grammatical error annotation of 2,837 essays written by Chinese language learners originating from 46 different mother-tongue languages. Each sentence in each essay is manually labeled. This yields an annotated corpus having a total of 25,057 sentences containing at least one grammatical error, while the remaining 63,446 sentences being grammatically correct (an unbalanced distribution with 28.31% sentences having grammatical errors).

The following detection methods were compared to show their performance. (1) CNN-LSTM (Lee et al., 2017): this method integrated Convolution Neural Networks (CNN) with Long Short-Term Memory (LSTM). (2) MC-CNN-BiLSTM (Lee et al., 2020): this model had two main parts, including the multi-channel embedding representation and a CNN along with a Bidirectional LSTM (BiLSTM). (3) BERT (Devlin et al., 2018): this model was a bidirectional transformer encoder pretrained using combination of masked language modeling objective and next sentence prediction. (4). RoBERTa (Liu et al., 2019): it's a replication study of BERT pre-training that modifies key hyperparameters, removes the next-sentence pre-training objective, and applies training with much larger mini-batches and learning rates. (5) XLNet (Yang et al., 2019): this is an extension of the Transformer-XL model pretrained using an autoregressive method to learn bidirectional contexts by maximizing the expected likelihood over all permutation of the input sequence factorization order. (6). ELECTRA (Clark et al., 2020): this is our used model based on adversarial learning for Chinese grammatical error detection.

Five-fold cross validation evaluation was adapted. For Word2vec embedding representation, the whole Chinese Wikipedia (zh_tw version on Dec. 24th, 2019) was firstly segmented into words and then the segmented sentences were used to train 300 dimensional vectors for 849,217 distinct words. The hyperparameters of CNN-LSTM and MC-CNN-BiLSTM were set up according to their suggestions (Lee et al., 2017; 2020). Pretrained Chinese transformer-based models (i.e. BERT-wwm, RoBERTa-wwm, XLNet-mid and ELECTRA-base) were downloaded from HuggingFace. The configured hyper-parameters were as follows: training batch size 128, learning rate 4e-5, and max sequence length 128. F1-score, which is a harmonic mean of the precision and recall, was used as the main evaluation metric to measure the performance.

Table 1 shows the results. The four transformer-based models achieved better F1-scores that outperformed two stack-based methods. Comparing ELECTRA with the other three transformers, the former achieved F1-score of 0.6353 which is an improvement over the latter (F1-score around 0.56). It reveals that adversarial learning can enhance the performance. Besides, it's noted that ELECTRA has similar precision and recall performance without a clear bias.

Table 1. *Evaluation on Chinese Grammatical Error Detection*

Method		Precision	Recall	F1
Stack-based	CNN-LSTM	0.3812	0.6544	0.4808
	MC-CNN-BiLSTM	0.3669	0.7845	0.4987
Transformer-based	BERT	0.6593	0.4725	0.5501
	RoBERTa	0.6691	0.4830	0.5606
	XLNet	0.6436	0.4928	0.5580
	ELECTRA	0.6406	0.6303	0.6353

4. Conclusions

This study explores the transformed-based neural networks for Chinese grammatical error detection. We use the TOCFL learner corpus to demonstrate the model performance. The ELECTRA model, which is a transformer network architecture along with adversarial learning technique, achieved an improvement of F1-score 0.6353 for detecting whether a given Chinese sentence contains any grammatical errors or not, which is the best performance in this task as far as we know.

Acknowledgements

This study was partially supported by the Ministry of Science and Technology, under the grant 106-2221-E-003-030-MY2, 108-2218-E-008-017-MY3 and 109-2410-H-003-123-MY3.

References

Clark, K., Luong, M.-T., Le, Q. V., & Manning, C. D. (2020). ELECTRA: pre-training text encoders as discriminators rather than generators. *Proceedings of ICLR'20*, https://arxiv.org/abs/2003.10555

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint, https://arxiv.org/abs/1810.04805

Lee, L.-H., Chang, L.-P., Lee, K.-C., Tseng, Y.-H. & Chen, H.-H. (2013). Linguistic rules based Chinese error detection for second language learning. *Proceedings of ICCE'13* (pp. 27-29), Bail, Indonesia: Asia-Pacific Society for Computers in Education.

Lee, L.-H., Lin, B.-L., Yu, L.-C., & Tseng, Y.-H. (2017). Chinese grammatical error detection using a CNN-LSTM model. *Proceedings of ICCE'17* (pp. 919-921), Christchurch, New Zealand: Asia-Pacific Society for Computers in Education.

Lee, L.-H., Tseng, Y.-H., & Chang, L.-P. (2018). Building a TOCFL learner corpus for Chinese grammatical error diagnosis. *Proceedings of LREC'18* (pp. 2298-2304), Miyazaki, Japan: ACL Anthology.

Lee, L.-H., Wang, Y.-S., Lin, P.-C., Hung, C.-T., & Tseng, Y.-H. (2020). Multi-channel CNN-BiLSTM for Chinese grammatical error detection. *Proceedings of ICCE'20* (pp. 558-560), online: Asia-Pacific Society for Computers in Education.

Lee, L.-H., Yu, L.-C., Lee, K.-C., Tseng, Y.-H., Chang, L.-P., & Chen, H.-H. (2014). A sentence judgment system for grammatical error detection. *Proceedings of COLING'14* (pp. 67-70), Dublin, Ireland: ACL Anthology.

Liu, Y., Han, Y., Zhou, L., & Zan, H. (2016). Automatic grammatical error detection for Chinese based on conditional random field. *Proceedings of NLPTEA'16* (pp. 57-62), Osaka, Japan: ACL Anthology.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V. (2019). RoBERTa: a robustly optimized BERT pretraining approach. arXiv preprint, https://arxiv.org/abs/1907.11692

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Proceedings of NIPS'13* (pp. 1-10), Stateline, Nevada.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. & Le, Q. V. (2019). XLNet: generalized autoregressive pretraining for language understanding. arXiv preprint, https://arxiv.org/abs/1906.08237