So, H. J. et al. (Eds.) (2020). Proceedings of the 28th International Conference on Computers in Education. Australia: Asia-Pacific Society for Computers in Education

Multi-Channel CNN-BiLSTM for Chinese Grammatical Error Detection

Lung-Hao LEE[a], Yuh-Shyang WANG[a], Po-Chen LIN[a], Chih-Te HUNG[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: In this paper, we proposed a Multi-Channel Convolutional Neural Network with Bidirectional Long Short-Term Memory (MC-CNN-BiLSTM) model for Chinese grammatical error detection. The TOCFL learner corpus is adopted to measure the system capability of indicating whether a sentence contains errors or not. Our model performs better than a previous CNN-LSTM model that reflects the effectiveness of multi-channel embedding representation.

Keywords: Chinese as a foreign language, grammatical error diagnosis, deep neural networks

1. Introduction

Chinese as foreign language (CFL) learners usually make various kinds of grammatical errors, such as missing words, redundant words, incorrect word selection, or word ordering error, during their language acquisition and production process. An automated system able to detect such errors would facilitate the learning and teaching of CFL. Previous Chinese grammatical error detection approaches were based on linguistic rules (Lee et al., 2013), machine learning classifiers, or their hybrid methods (Lee et al., 2014). Deep neural network (i.e., deep learning) approaches have been widely applied recently and achieve dominating results in many natural language processing tasks. This trend motivates us to explore deep neural networks to detect errors written by CFL learners.

This study describes our proposed Multi-Channel Convolutional Neural Network with Bidirectional Long Short-Term Memory (MC-CNN-BiLSTM) model for Chinese grammatical error detection. The TOCFL learner corpus (Lee et al., 2018) is used to evaluate the performance. Compared with an previous approach on the same dataset, our proposed method achieved better F1-score which takes into account both detection precision rate and recall rate at the same time .

2. Multichannel CNN with Bidirectional LSTM (MC-CNN-BiLSTM)

Figure 1 illustrates our Multi-Channel Convolutional Neural Network with Bidirectional Long Short-Term Memory (MC-CNN-BiLSTM) model for Chinese grammatical error detection. The model has two main parts: 1) the multi-channel embedding representation and 2) a CNN along with a Bidirectional LSTM network (CNN-BiLSTM). In this model, an input sentence is represented as a sequence of words. The embedding vector of each word is pre-trained from a large corpus using different embedding techniques, such as Word2vec (Mikolov et al., 2013), GloVec (Pennington et al., 2014), and ELMo (Peters et al. 2018). Through different delicately designed training, the embedding vectors learn to become better semantic representations of their words in different ways. Therefore, instead of adopting only one single embedding technique for semantic representation, the multi-channel representations adopting multiple embedding sources as multi-channels are used in our model.

In each channel, a convolutional layer with k different convolution filters is used to extract k feature maps, each contains a kind of local n-gram (a consecutive of n word embeddings) features. The max-pooling layer is then used to aggregate the local n-gram features from each feature map to keep the most salient information corresponding to each input word. The obtained feature vectors are fed to the sequential BiLSTM layer to capture long-distance (global) features among the input words. In the BiLSTM, the backward LSTM is a reversed copy of the forward LSTM, so that we can take full advantage of the forward and backward word sequence information from the input sentence. The output of the BiLSTM denoting the input sentence representation from each channel is concatenated to form a long vector, which is fed to a neural network layer with the final softmax activation function for binary classification signifying whether the input sentence contains grammatical errors or not.

During the training phase, if a sentence contains at least one grammatical error judged by a human, its class is labeled as 1, and 0 otherwise. All the sentences with their labeled classes are used to train our MC-CNN-BiLSTM model to automatically learn all the corresponding parameters in this model. To classify a sentence during the testing phase, the sentence unseen in the training phase goes through the MC-CNN-BiLSTM architecture to yield a value corresponding to the error probability. If the probability of a sentence with class 1 (*i.e.*, with errors) exceeds a predefined threshold, it is considered as true as an erroneous sentence, and false otherwise.

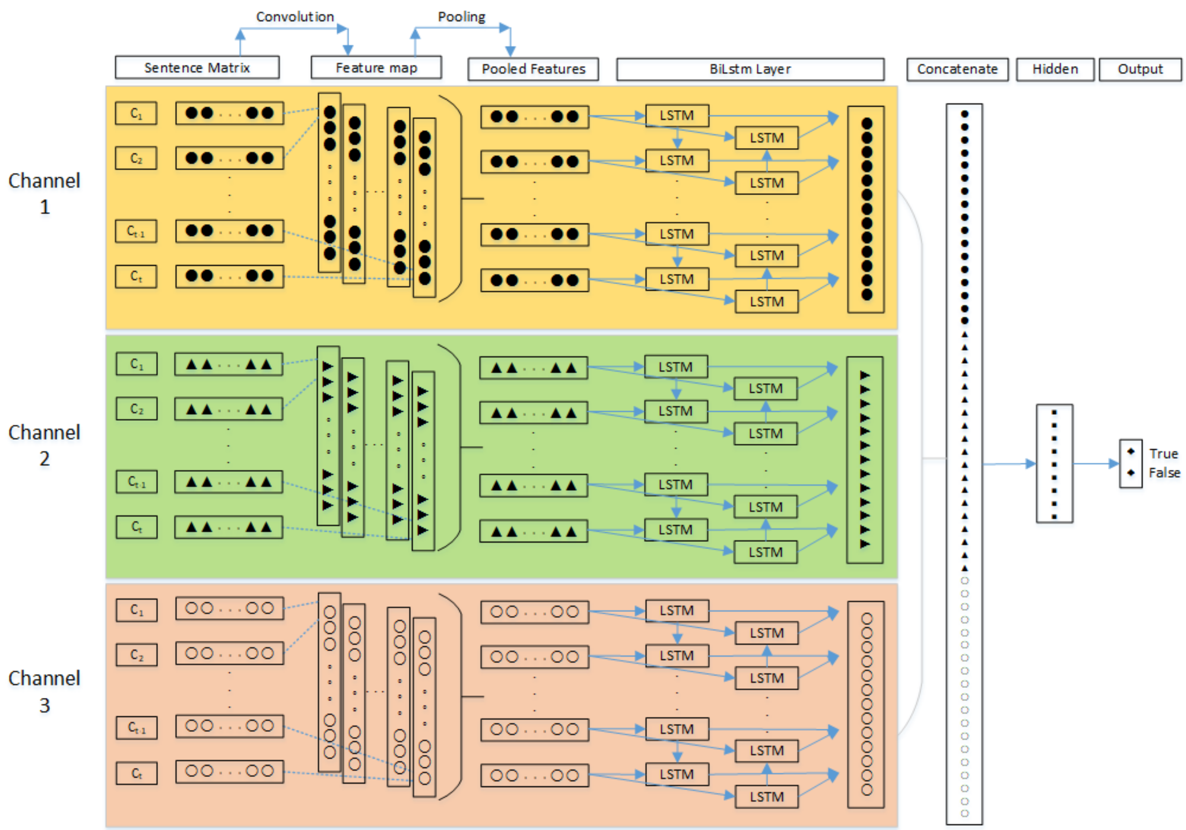

Figure 1. The illustration of our MC-CNN-BiLSTM model.

3. Experiments and Evaluation 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. The result is that a total of 25,057 sentences contain at least one grammatical error, while the remaining 63,446 sentences are grammatically correct (an unbalanced distribution with 28.31% sentences having grammatical errors). Five-fold cross validation evaluation was used to measure the performance.

For Word2vec, GloVe, and ELMo embedding representations, 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. For convolution operation, the number of filters was k=300 and their length was 3. The loss function was categorical cross-entropy and the optimizer was Adam. The number of training iteration (*i.e.*, epochs) was set to 3 to learn the CNN-BiLSTM network parameters. If the error probability of an input sentence exceeds 0.2, it was considered as an erroneous sentence.

The following methods were compared to show their performance. (1) CNN-LSTM (Lee et al., 2017): this method integrated CNN with LSTM. According to their suggestions, Word2Vec embedding was used and the hyperparameters were set up as follows: The number of filters was 300 and their length was 3, epochs were 5, and the threshold is 0.3. (2) MC-CNN-BiLSTM: this is our proposed model for Chinese grammatical error detection. We also do some ablation experiments, i.e., comparing the performance using different number of channels in terms of individual embedding combinations.

Table 1 shows the results. The single-channel MC-CNN-BiLSTM was degraded as CNN-BiLSTM, which was still slightly better than the CNN-LSTM (Lee et al., 2017). It reveals that bidirectional LSTM in the sequence layer can enhance the performance. Comparing our MC-CNN-BiLSTM model with the CNN-LSTM, all our models has the improvement in terms of F1-score no matter which number of channels was used. In our observation, an increasing number of channels will achieve better performance if different embedding representations were used in a proper combination. It is noted that due to the large number of testing examples (17700=(25057+63446)/5), the small number of improvement (over 0.01 absolute improvement or over 3.72% relative improvement) is still statistically significant.

Table 1: Evaluation on Chinese grammatical error detection.

<table>
<tr><th colspan="3">Method</th><th>Precision</th><th>Recall</th><th>F1</th></tr>
<tr><td colspan="3">CNN-LSTM (Lee et al., 2017)</td><td>0.3812</td><td>0.6544</td><td>0.4808</td></tr>
<tr><td rowspan="3">MC-CNN-BiLSTM (our model)</td><td>Single-channel</td><td>Word2vec</td><td>0.3386</td><td>0.8656</td><td>0.4862</td></tr>
<tr><td>Dual-channel</td><td>Word2vec + GloVe</td><td>0.3667</td><td>0.7801</td><td>0.4982</td></tr>
<tr><td>Triple-channel</td><td>Word2vec + GloVe + ELMo</td><td>0.3669</td><td>0.7845</td><td>0.4987</td></tr>
</table>

4. Conclusions

Chinese grammatical error detection is a very hard problem such that even MS Word has no such function for Chinese (but has this function for English for many years). This study describes the MC-CNN-BiLSTM model for Chinese grammatical error detection. We use the TOCFL learner corpus to demonstrate the system performance. Our system achieved an improvement of F1-score 0.4987 for predicting whether a given Chinese sentence contains any grammatical errors or not, which roughly corresponds to a half of input sentences were judged correctly under the unbalanced error distribution.

Acknowledgements

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

References

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., 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

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.

Pennington, J., Socher, R., & Manning C. D. (2014). GloVe: global vectors for word representation. *Proceedings of EMNLP'14* (pp. 1532-1543), Doha, Qatar: ACL Anthology.

Petters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *Proceedings of NAACL-HLT'18* (pp. 2227-2237), New Orleans, Louisiana, USA: ACL Anthology.