Chen, W. et al. (Eds.) (2017). Proceedings of the 25th International Conference on Computers in Education. New Zealand: Asia-Pacific Society for Computers in Education

Chinese Grammatical Error Detection Using a CNN-LSTM Model

Lung-Hao LEE[a], Bo-Lin LIN[b,c], Liang-Chih YU[b,c] & Yuen-Hsien TSENG[a*]
[a]*Graduate Institute of Library and Information Studies, National Taiwan Normal University, Taiwan*
[b]*Department of Information Management, Yuan Ze University, Taiwan*
[c]*Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taiwan*
*samtseng@ntnu.edu.tw

Abstract: In this paper, we proposed a Convolution Neural Network with Long Short-Term Memory (CNN-LSTM) model for Chinese grammatical error detection. The TOCFL learner corpus is adopted to measure the system performance of indicating whether a sentence contains errors or not. Our model performs better than other neural network based methods in terms of accuracy for identifying an erroneous sentence written by Chinese language learners.

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

1. Introduction

Chinese as foreign language learners usually make different kinds of grammatical errors during second language acquisition process (Lee et al., 2016a). Automated grammatical error detection and correction are emerging important research directions and a number of competitions have been organized to encourage innovation (Leacock et al., 2014). Recently, the Natural Language Processing Techniques for Educational Applications (NLPTEA) workshops have hosted a series of shared tasks for Chinese grammatical error diagnosis (Yu et al., 2014; Lee et al., 2015; Lee et al., 2016b). All of these activities attracted global participations and enhanced research developments.

Language models have been adopted to detect various types of Chinese errors written by US learners (Wu et al., 2010). A probabilistic inductive learning algorithm has been proposed to diagnose Chinese grammatical errors (Chang et al., 2012). Linguistic rules have been manually constructed to detect Chinese erroneous sentences (Lee et al., 2013). Support Vector Machine based classifiers have been used to explore useful features for detecting word-ordering errors in Chinese sentences (Yu and Chen, 2012). A sentence judgment system has been developed to detect grammatical errors in Chinese sentences using both n-gram statistical analysis and rule-based linguistic analysis (Lee et al., 2014). Gated recurrent neural network models have been explored to select the best prepositions for Chinese grammatical error diagnosis (Huang et al., 2016). In recent NLPTEA workshops (Lee et al., 2015; Lee et al., 2016b), neural approaches have been explored for identifying Chinese grammatical errors. This observation motivates us to explore neural networks to detect errors written by Chinese learners.

This study describes our proposed Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM) model, a kind of deep neural network, for Chinese grammatical error detection. The TOCFL learner corpus is used to evaluate and compare performance. Error detection systems that indicate grammatical errors in a given sentence are useful to leaners for computer-assisted language learning.

2. Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM)

Figure 1 shows our Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM) architecture for Chinese grammatical error detection. An input sentence is represented as a sequence of words. Each word refers to a row looked up in a word embedding matrix generating from Word2Vec (Mikolov et al., 2013). A single convolution layer is

adopted. We use convolutions overs the sentence matrix to extract features. The full convolutions are obtained by sliding the filters over the whole matrix. Each filter performs the convolution operations on the sentence matrix and generates a feature map. A pooling layer is then used to subsample features over each map. We apply the max operation to reduce the dimensionality for keeping the most salient features. To capture long-distance dependency across features, LSTM is used in the sequential layer for vector composition. After the LSTM memory cells sequentially traverse through all feature vectors, the last state of the sequential layer is regarded as input for neural computing. The final softmax layer then receives computing results and uses it to classify the sentence.

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 CNN-LSTM model to automatically learn all the corresponding parameters in this model.

To classify a sentence during the testing phase, the sentence goes through the CNN-LSTM 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.

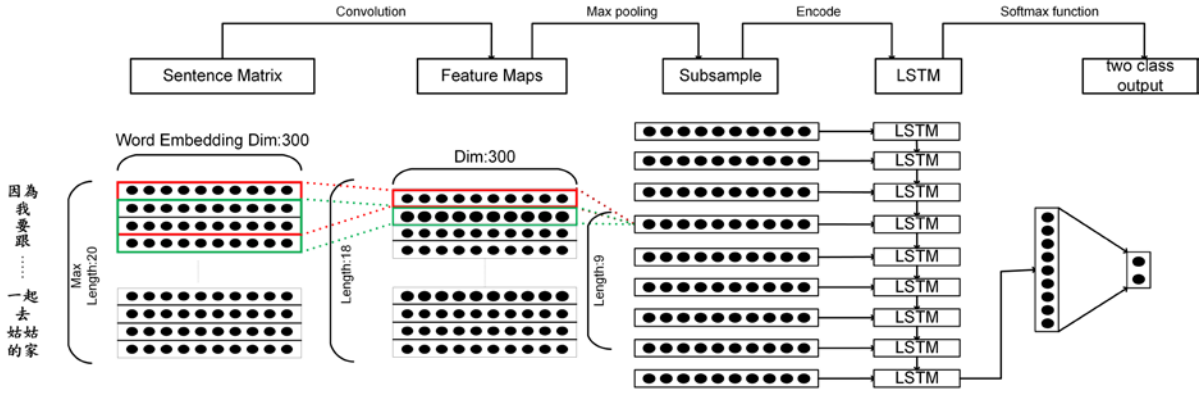

Figure 1. The illustration of our CNN-LSTM model for Chinese grammatical error detection.

3. Experiments and Evaluation Results

The experimental data came from the TOCFL learner corpus (Lee et al., 2016a), 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 25,277 sentences contain at least one grammatical error, while the remaining 68,982 sentences are grammatically correct (an unbalanced distribution with 26.82% sentences having grammatical errors). Five-fold cross validation evaluation was used to measure the performance.

To implement the system, a python library Theano was used. For Word2Vec representation, Chinese Wikipedia 2016 was trained to generate 300 dimensional vectors for 655,247 words and phrases. The number of filters was 300 and their length is 3. The number of iteration (*i.e.*, epochs) was set up as 5 to learn the CNN-LSTM network parameters. If the error probability of an input sentence exceeds 0.3, it was considered as an erroneous sentence.

The following three methods were compared to demonstrate their performance. (1) CNN only: this method only considers the CNN part of our proposed model. (2) LSTM only: this approach only focuses on the LSTM part of our proposed model (3) CNN-LSTM: this is our proposed model for Chinese grammatical error detection.

Table 1 shows the results. The CNN only and CNN-LSTM model respectively had the best recall and precision. Considering the tradeoff, the LSTM only model reflected the best F1-score of 0.4859 (the improvement compared to the lowest F1-score is 5.4%). In addition to best precision, our

proposed CNN-LSTM model also achieved the best accuracy of 0.6905 (the improvement compared to the lowest accuracy is 12.77%).

Table 1: Evaluation on Chinese grammatical error detection.

Method	Accuracy	Precision	Recall	F1-score
CNN only	0.6123	0.3745	0.6488	0.4717
LSTM only	0.6599	0.4179	0.6049	0.4859
CNN-LSTM	0.6905	0.4439	0.5057	0.4610

4. Conclusions

This study describes the CNN-LSTM model for Chinese grammatical error detection. We use the TOCFL learner corpus to demonstrate system performance. Our system achieved the best accuracy of 0.6905 for predicting whether a given sentence contains grammatical errors or not, which roughly corresponds to 7 out 10 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 MOST 103-2221-E-003-013-MY3, MOST 105-2221-E-155-059-MY2, MOST 106-2221-E-003-030-MY2 and the "Aim for the Top University Project" and "Center of Language Technology for Chinese" of National Taiwan Normal University, sponsored by the Ministry of Education, Taiwan, ROC.

References

Chang, R.-Y., Wu, C.-H., & Prasetyo, P. K. (2012). Error diagnosis of Chinese sentences using inductive learning algorithm and decomposition-based testing mechanism. *ACM Transactions on Asian Language Information Processing*, 11(1), Article 3.

Huang, H.-H., Shao, Y.-C., Chen, H.-H. (2016). Chinese preposition selection for grammatical error diagnosis. *Proceedings of COLING'16* (pp. 888-899). Osaka, Japan: ACL Anthology.

Leacock, C., Chodorow, M., Gamon, M., & Tetreault, J. (2014). Automated Grammatical Error Detection for Language Learners (2nd Edition). Morgan & Claypool Publishers.

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., Chang, L.-P., & Tseng, Y.-H. (2016a). Developing learner corpus annotation for Chinese grammatical errors. *Proceedings of IALP'16* (pp. 254-257), Tainan, Taiwan: IEEE Digital Library.

Lee, L.-H., Rao, G., Yu, L.-C., Xun, E., Zhang, B., & Chang, L.-P. (2016b). Overview of the NLP-TEA 2016 shared task for Chinese grammatical error diagnosis. *Proceedings of NLPTEA'16* (pp. 40-48), Osaka, Japan: ACL Anthology.

Lee, L.-H., Yu, L.-C., & Chang, L.-P. (2015). Overview of the NLP-TEA 2015 shared task for Chinese grammatical error diagnosis. *Proceedings of NLPTEA'15* (pp. 1-6), Beijing, China: 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.

Yu, C.-H., & Chen, H.-H. (2012). Detecting word ordering errors in Chinese sentences for learning Chinese as a foreign language. *Proceedings of COLING'12* (pp. 3003-3017), Bombay, India: ACL Anthology

Yu, L.-C., Lee, L.-H., & Chang, L.-P. (2014). Overview of grammatical error diagnosis for learning Chinese as a foreign Language. *Proceedings of NLPTEA'14* (pp. 42-47), Nara, Japan: Asia-Pacific Society for Computers in Education.

Wu, C.-H., Liu, C.-H., Harris, M., & Yu, L.-C. (2010). Sentence correction incorporating relative position and parse template language model. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6), 1170-1181.