A study of Grammar Analysis in English Teaching with Deep Learning Algorithm

Gang Zhang
Henan Medical College, Zhengzhou, Henan, China
`zg4ict@126.com`

Abstract—In English teaching, grammar is a very important part. Based on the seq2seq model, a grammar analysis method combining the attention mechanism, word embedding and CNN seq2seq was designed using the deep learning algorithm, then the algorithm training was completed on NUCLE, and it was tested on CoNIL-2014. The experimental results showed that $F_{0.5}$ of seq2seq+attention improved 33.43% compared to the basic seq2seq; in the comparison between the method proposed in this study and CAMB, the P value of the former was 59.33% larger than that of CAMB, the R value was 8.9% larger, and the value of $F_{0.5}$ was 42.91% larger. Finally, in the analysis of the actual students' grammar homework, the proposed method also showed a good performance. The experimental results show that the method designed in this study is effective in grammar analysis and can be applied and popularized in actual English teaching.

Keywords—Deep learning, English teaching, seq2seq attention mechanism, recall rate.

1 Introduction

In English teaching, grammar analysis is a very important part. Students can master the structure, rules and other knowledge of English through the study of grammar, so as to better master the use of English language and lay a solid foundation for the improvement of reading, speaking, writing and other abilities, which is also conducive to the cultivation of students' English application ability and the future work and learning of students. Therefore, the study of grammar is of great significance in English teaching. The correction of grammar works of students usually takes a lot of time and effort of teachers. With the development of computer technology, the application of artificial intelligence in education has become more and more popular [1], and methods such as machine learning [2] and deep learning have been widely used in language processing [3], which also provides a new possibility for the automatic analysis of English grammar. Tang et al. [4] designed an online patient question answering system based on deep learning. With that system, patients can find semantically similar questions in Q & A to get answers without waiting for doctors' answers. This

method was based on deep neural network and could predict the similarity between questions. They verified the effectiveness of the system through experiments. Chen et al. [5] analyzed the performance of deep learning in radiology text report classification, compared the performance of convolutional neural network (CNN) model and PeFinder, and found that the accuracy of CNN was 99%, F1 value was 0.938, and PeFinder was 0.867 in the internal validation report data set, which verified the reliability of CNN model. Zhang et al. [6] analyzed event recognition. After sentence segmentation, the words were divided into different categories and vectorized. Then, the deep belief network (DBN) was used for training. The test found that the maximum F value of the method was 85.17%, and the F value of the dynamic supervision DBN was 88.11%, showing better recognition performance. Yousfi et al. [7] recognized the text in Arabic video with Recurrent Neural Network (RNN), combined RNN model in long short-term memory, and introduced super parameters to improve the recognition effect. They found that the recognition rate increased by 16% and the response time was relatively reasonable. This study improved the seq2seq model, trained and tested it, and analyzed its performance in the grammar analysis, which makes some contributions to improving the efficiency of English teaching, reducing the burden of teachers and improving the learning effect of students.

2   Deep Learning Related Concepts

2.1   Deep learning

Deep learning can be said to be a kind of machine learning and a part of artificial intelligence. It is developed from artificial neural network (ANN). It can make machines have certain functions through some learning methods, so as to complete some tasks instead of human beings. It has a wide range of applications in many fields, such as image processing [8], automatic classification [9] and disease prediction [10].

2.2   RNN

When processing temporal order information, RNN [11] can take current input and previous input information into account simultaneously. It has a good applicability in language processing. Its structure mainly includes: (1) input layer $X = (x_0, x_1, \cdots, x_n)$; (2) output layer $Y = \{y_0, y_1, \cdots, y_n\}$; (3) hidden layer $H = \{h_0, h_1, \cdots, h_n\}$, and the weight of layers is w_{xh}, w_{hh} and w_{hy} respectively. The calculation formulas are:

$$h_n = f\left(h_{n-1}, x_n\right) \tag{1}$$

$$h_n = \tanh\left(w_{hh}h_{n-1} + w_{xh}x_n\right) \tag{2}$$

$$y_n = w_{hy}h_n \tag{3}$$

where f refers to the activation function. It is seen from the formulas that the output of RNN at time n is related not only to current input x_n, but also to all the previous inputs.

2.3 Seq2seq model

Seq2seq model is a kind of deep learning model, which refers to the mapping from sequence to sequence [12]. English teaching grammar analysis can also be seen as the mapping from the text sequence with grammatical errors to the correct text sequence; therefore, seq2seq model can be used. Seq2seq has Encoder-Decoder structure, generally RNN structure, and a terminator <EOS> at the end. Taking the translation from sentence "ABC" to "WXYZ" as an example. Seq2seq structure is shown in Fig. 1.
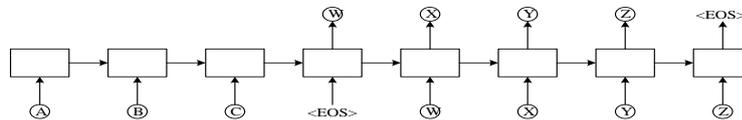
Fig. 1. Seq2seq structure

It is assumed that the input of seq2seq is $X = \{x_1, x_2, ..., x_n\}$, the output is $Y = \{y_1, y_2, ..., y_{n'}\}$, and n and n' may not be equal. RNN is used to evaluate conditional probability:

$$p(y_1, y_2, \cdots y_{n'} \mid x_1, x_2, \cdots, x_n) = \prod_{n=1}^{n'} p(y_n \mid h, y_1, y_2, \cdots, y_{n-1})$$

. In the training stage, the objective function can be expressed as:

$$L = \frac{1}{|Z|} \sum_{(I,O) \in Z} \log p(O \mid I)$$

, where Z stands for the training set, I is the input sequence, and O is the output sequence, i.e. the translated sentence. The translation result $\hat{O}$ can be obtained through the decoder RNN:

$$\hat{O} = \arg\max_{O} p(O \mid I)$$

. In grammar analysis, the input of seq2seq is the sentence with grammatical errors, and the output is the correct sentence.

3 Grammar Analysis in English Teaching by Using Deep Learning

3.1 Attention mechanism

In seq2seq, if the input sentence is long, the detail information may be lost. In order to solve this problem, the attention mechanism [13] can be introduced to enable the decoder to automatically focus on different parts of the sentence and improve the

decoding ability. It is assumed that Decoder has a state of s_n at the n-th time, the output is y_n, the semantic information is c_n, then: $s_n = f(s_{n-1}, y_{n-1}, c_n)$, where $c_n = \sum_{j=1}^{N} a_{nj} h_j$. a_{nj} is the semantic vector weight at time n and j, and $a_{nj} = \frac{\exp(e_{nj})}{\Sigma \exp(e_{nk})}$. e_{nj} is the relationship between the state of Decoder at time $n-1$ and the j-th state of Encoder, and $e_{nj} = a(s_{n-1}, h_n)$.

In grammatical analysis, the detection and correction of grammatical errors are generally determined by local words. For example, in the selection of articles, "a" and "an" depend on the following nouns. Therefore, after the introduction of the attention mechanism, the seq2seq model can perform better in grammatical analysis.

3.2 Word embedding

In order to make the computer understand the natural language, it is necessary to perform vectorization operation on the text to turn it into numbers. GloVe method [14] was used in this study: for a corpus, co-occurrence matrix is represented by M, and the times of word j appearing in the context of word i is represented by M_{ij}, then its co-occurrence probability is:

$$P_{ij} = P(j \mid i) = \frac{M_{ij}}{M_i}$$

, where M_{ij} stands for the times of any word appearing in the context of word i, $M_i = \sum_k M_{ik}$; the relationships between word vector v_{w_i}, v_{w_l} and v_{w_k} of word w_i, w_j and w_k can be expressed as: $\exp(v_{w_i} - v_{w_j})^T v_{w_k} = \frac{P_{ik}}{P_{jk}}$.

3.3 CNN-seq2seq

Compared with RNN, CNN has higher efficiency in capturing local information and shorter training time. In grammatical analysis, grammatical errors are generally local. Therefore, CNN can be used instead of RNN, and CNN-seq2seq is obtained.

In CNN-seq2seq model, both Decoder and Encoder use CNN and combine the gating linear unit (GLU) in the nonlinear part, and the attention mechanism is added at the same time. It is assumed that the word vector of the input sentence is $r = (r_1, r_2, ..., r_m)$, position vector is $t = (t_1, t_2, ..., t_m)$, then the word vector of the output sentence is $u = (r_1 + t_1, r_2 + t_2, ..., r_m + t_m)$. The output of the l-th layer of Encoder is $p^l = (p_1^l, p_2^l, ..., p_m^l)$. The output of the l-th layer of Decoder is $q^l = (q_1^l, q_2^l, ..., q_n^l)$. The size of convolution kernel is W^{kd*2d}, where k stands for the size of the convolution-

al window and d stands for the dimension. The vector generated by every time of convolution can be expressed as: $Y = [A\ B]$. The formula of GLU can be expressed as: $v([A\ B]) = A \otimes \sigma(B)$, where $\otimes$ stands for dot product and $\sigma(B)$ stands for the extent to which parts of A need to be partially retained.

The attention weight of the model is determined by the current output q_i^l of Decoder and all the outputs of Encoder. This weight can weight the output of Encoder to get input sentence vector c_i^l, as follows:

$$d_i^l = w_d^l q_i^l + b_d^l + g_i \tag{4}$$

$$a_{ij}^l = \frac{\exp\left(d_i^l p_j^u\right)}{\sum_{n=1}^{m} \exp\left(d_i^l p_n^u\right)} \tag{5}$$

$$c_i^l = \sum_{j=1}^{m} a_{ij}^l \left(z_j^u + e_j\right) \tag{6}$$

where g_i stands for the word vector information before q_i^l, e_j stands for the word vector information of the input word, and a_{ij}^l stands for the weight. CNN-seq2seq can fully find the hidden information in sentences and has a high computing speed. In addition, it combines GLU and attention mechanism, showing a better performance in grammatical analysis.

4 Experimental Results

4.1 Experimental data set

Training set: The NUCLE data set [15] of National University of Singapore was used, which is written by non-English native students according to certain topics, such as environmental pollution, medical health, etc. There are 1397 articles in total in the data set, which are marked and corrected by professional English teachers. Some errors are shown in Table 1.

Table 1. Some grammatical mistakes in NUCLE

Type	Content	Example sentence
Vt	Verb tense	A university [had conducted→conducted] the survey last year.
SVA	Subject predi-cate agreement	The boy [play→plays] soccer.
ArtOrDet	Article	From the ethical aspect,sex selection technology should not be used in [non-medical→a non-medical] situation.
Nn	Singular and plural nouns	Sex selection should therefore be used for medical [rea-son→reasons] and nothing else.

Test set: CoNIL-2014 was used as the test set [16], in which there are 1312 sentences. CoNIL-2014 covered 28 kinds of grammatical errors, and the errors have been marked and corrected by two native English speakers. Some errors are shown in Table 2.

Table 2. Some grammatical errors in CoNIL-2014

Type	Content	Example sentence
Vt	Verb tense	Medical technology during that time [is→was] not adcanced entough to cure him.
V0	Verb absence	However, there are also a great number of people [who→who are] against this technology.
SVA	Subject predicate agreement	The benefits of disclosing genetic risk information [out-weighs→outweigh] the costs.
Nn	Singular and plural nouns	A carrier may consider not having any [child→children] after getting married.
Prep	Preposition	This essay will [discuss about→discuss] whether a carrier should tell his relatives or not.

4.2 Evaluation criteria

The evaluation of CoNIL-2014 takes $F_{0.5}$ as the standard, and its calculation formula is as follows:

$$F_{0.5} = \frac{(0.5^2 + 1)PR}{0.5^2 P + R} \tag{6}$$

where P stands for accuracy rate, $P = \frac{A}{A+C}$, R stands for recall rate, $R = \frac{A}{A+B}$, A stands for the correctly marked sentence, B stands for the wrong sentence that is not marked out, and C stands for the correct sentence which is marked as wrong.

4.3 Experimental results

This study analyzed the performance of the grammar analysis method which combined the Attention mechanism, word embedding and CNN-seq2seq and compared the grammar analysis effect of the basic seq2seq, seq2seq + attention, CAMB [17] and the method designed in this study on CoNIL-2014. The results are shown in Fig. 2.
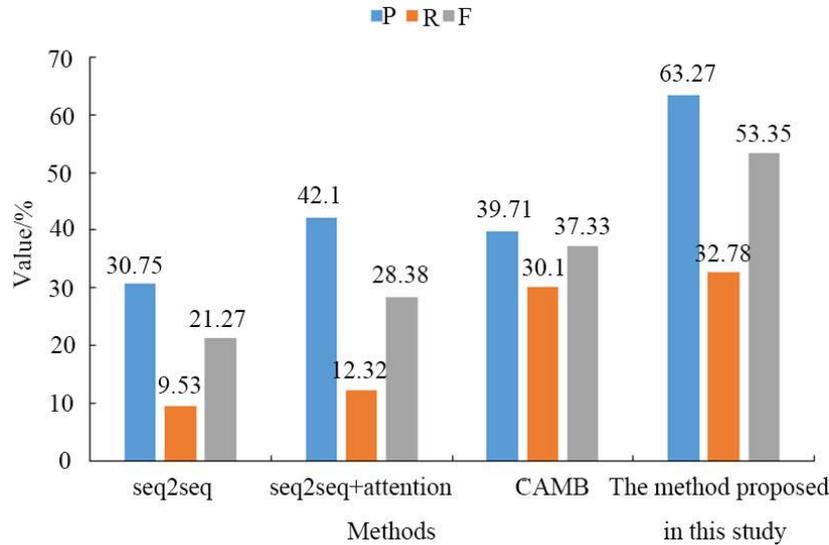
Fig. 2. Comparison of analysis results between different methods

It was seen from Fig. 2 that the effect of the basic seq2seq on grammar analysis was relatively poor, with low P and R values, and the value of $F_{0.5}$ was only 21.27%, which showed that the efficiency of this method in the process of grammar analysis was relatively low if the correct sentences are marked as errors or the wrong sentences are not marked out. After the attention mechanism was added, the effect of grammar analysis significantly improved, the P value increased by 36.91%, the R value increased by 29.28%, and $F_{0.5}$ was 28.38%, 33.43% higher than the basic seq2seq, which verified that the attention mechanism improved the model. Compared with seq2seq + attention, the R value of CAMB significantly improved, which was about twice of the former, indicating that the CAMB method had better performance in identifying sentence errors, but slightly lower accuracy. Values of P, R and $F_{0.5}$ of the method proposed in this study were large; compared to CAMB, the P value improved 59.33%, the R value improved 8.9%, and $F_{0.5}$ improved 42.91%, which verified the performance of the proposed method in the grammar analysis.

The actual grammar homework of students were corrected by the method proposed in this study, and 100 sentences were collected. Some results of the grammar analysis are shown in Table 3.

Table 3. The analysis results of actual grammar homework of students

	Input	Output
1	Mary should't have received my letter, otherwise she would have replied before now.	Mary couldn't have received my letter, otherwise she would have replied before now.
2	Nobody but Jim and Mike were on the playground.	Nobody but Jim and Mike was on the playground.
3	She knows you had been in Beijing for five years.	She knows you have been in Beijing for five years.
4	It is great pleasure to talk with you.	It is a great pleasure to talk with you.
5	She said that she and my schoolmate all wished me success.	She said that she and my schoolmates all wished me success.
6	Suddenly, I caught a sight of my English teacher in the crowd.	Suddenly, I caught a sight of my English teacher in the crowd.

In (1), the method corrected the error of modal verbs in the sentence; in (2), it corrected the error of subject predicate agreement; in (3), it corrected the error of verb tense; in (4), it added the true article "a"; in (5), it corrected the error of singular and plural nouns; but in (6), the article "a" in the sentence was redundant and should be deleted, but the method determined it as the correct sentence output. Among these 100 sentences, 71 sentences were found and corrected correctly. Although there were also cases of misjudgment and missed judgment, the overall performance was high.

5 Discussion

With the progress of computer and the development of artificial intelligence technology, computer has been more and more widely used in the processing of natural language. Writing a good computer program can realize the analysis and processing of human language, such as the classification of different texts, the translation of different languages [18], the analysis of emotions contained in languages [19], and opinion mining [20], and reading text automatically. In English teaching, grammar is a difficult but very important part, which will directly affect listening, speaking, writing, etc. [21]. In the process of homework, there are often errors such as tense errors and misuse of words. In the process of correction, teachers inevitably make mistakes or omissions. If teachers can use computer technology to detect and correct the grammatical errors in students' work, it will be of great positive significance for teachers and students.

Based on the deep learning method, this study designed a grammar analysis method which combined the attention mechanism, word embedding and CNN-seq2seq model. It was trained on NUCLE and tested on CoNIL-2014. Compared with other models, it was found that the proposed method had a good performance in the grammar analysis. Firstly, after combining with the attention mechanism, the $F_{0.5}$ of the basic seq2seq improved 33.43%, which verified the effectiveness of the attention mechanism. Then it was found from the comparison with CAMB that the P value and R value of the proposed method increased by 59.33% and 8.9% respectively, and $F_{0.5}$ improved 42.91%, which showed that the method had a good performance in marking and correcting grammatical errors in the process of grammatical analysis and

could achieve accurate grammatical analysis. Then, in the actual students' grammatical analysis, it was also found that the proposed method could accurately find and modify the grammatical errors in sentences and showed a good performance although there were some wrong and missed corrections.

Although some achievements have been made in the research of deep learning-based grammar analysis method in English teaching in this study, there are still some shortcomings that need to be solved in the future work, such as:

1. Further improving the recall rate of grammatical analysis methods.
2. Expanding the scale of data set to train the model better.
3. Considering the performance of other deep learning methods in grammar analysis.

6   Conclusion

Based on the seq2seq model in deep learning, this study designed a grammar analysis method which combined the attention mechanism, word embedding and CNN-seq2seq and trained and tested it. The results showed that:

1. The addition of the attention mechanism significantly improved the accuracy of the model.
2. Compared with CAMB, the P value and R value of the proposed method were 59.33% and 8.9% larger respectively, and $F_{0.5}$ was 42.91% larger.
3. The proposed method also had a good performance in the analysis of the actual grammar homework of students.

7   References

[1] Muniasamy, A., Alasiry, A. (2020). Deep Learning: The Impact on Future eLearning. International Journal of Emerging Technologies in Learning (iJET), 15(1): 188. https://doi.org/10.3991/ijet.v15i01.11435

[2] Hadioui, A., Faddouli, N.E.E., Touimi, Y.B., Bennani, S. (2017). Machine Learning Based on Big Data Extraction of Massive Educational Knowledge. International Journal of Emerging Technologies in Learning, 12(11): 151. https://doi.org/10.3991/ijet.v12i11.7460

[3] Wang, C., Shen, Y., Jia, J., Lu, Y., Chen, Z., Wang, B. (2018). SingleCaffe: An Efficient Framework for Deep Learning on a Single Node. IEEE Access, 6: 69660-69671. https://doi.org/10.1109/access.2018.2879877

[4] Tang, G.Y., Ni, Y., Xie, G.T., Fan, X.L., Shi, Y.L. (2017). A Deep Learning-Based Method for Similar Patient Question Retrieval in Chinese. Studies in Health Technology & Informatics, 245: 604-608.

[5] Chen, M.C., Ball, R.L., Yang, L., Moradzadeh, N., Chapman, B.E., Larson, D.B., Langlotz, C.P., Amrhein, T.J., Lungren, M.P. (2017). Deep Learning to Classify Radiology Free-Text Reports. Radiology, 286(3): 171115. https://doi.org/10.1148/radiol.2017171115

[6] Zhang, Y., Liu, Z., Zhou, W. (2016). Event Recognition Based on Deep Learning in Chinese Texts. plos one, 11(8): e0160147. https://doi.org/10.1371/journal.pone.0160147

[7] Yousfi, S., Berrani, S.A., Garcia, C. (2016). Contribution of recurrent connectionist language models in improving LSTM-based Arabic text recognition in videos. Pattern Recognition, S0031320316303697. https://doi.org/10.1016/j.patcog.2016.11.011

[8] Kamilaris, A., Prenafeta-Boldú, F.X. (2018). Deep Learning in Agriculture: A Survey. Computers and Electronics in Agriculture, 147(1): 70-90. https://doi.org/10.1016/j.compag.2018.02.016

[9] Grassmann, F., Mengelkamp, J., Brandl, C., Harsch, S., Zimmermann, M.E., Linkohr, B., Peters, A., Heid, I.M., Palm, C., Weber, B.H.F. (2018). A Deep Learning Algorithm for Prediction of Age-Related Eye Disease Study Severity Scale for Age-Related Macular Degeneration from Color Fundus Photography. Ophthalmology, 125(9): 8280209. https://doi.org/10.1016/j.ophtha.2018.02.037

[10] Kiral-Kornek, I., Roy, S., Nurse, E., Mashford, B., Karoly, P.J., Carroll, T., Payne, D., Saha, S., Baldassano, S., O'Brien, T., Grayden, D.B., Cook, M.J., Freestone, D.R., Harrer, S. (2018). Epileptic Seizure Prediction Using Big Data and Deep Learning: Toward a Mobile System. EBioMedicine, 27: 103. https://doi.org/10.1016/j.ebiom.2017.11.032

[11] Gregor, K., Danihelka, I., Graves, A., Wierstra, D. (2015). DRAW: A Recurrent Neural Network for Image Generation. computer science, 1462-1471.

[12] Dušek, O., Jurcicek, F. (2016). Sequence-to-Sequence Generation for Spoken Dialogue via Deep Syntax Trees and Strings. Meeting of the Association for Computational Linguistics. https://doi.org/10.18653/v1/p16-2008

[13] Cho, K., Courville, A., Bengio, Y. (2015). Describing Multimedia Content Using Attention-Based Encoder-Decoder Networks. Multimedia, IEEE Transactions on, 17(11): 1875-1886. https://doi.org/10.1109/tmm.2015.2477044

[14] Naili, M., Chaibi, A.H., Ben Ghezala, H.H. (2017). Comparative study of word embedding methods in topic segmentation. Procedia Computer Science, 112: 340-349. https://doi.org/10.1016/j.procs.2017.08.009

[15] Dahlmeier, D., Ng, H.T., Wu, S.M. (2013). Building a large annotated corpus of learner English: The NUS corpus of learner English. Proceedings of the Eighth Workshop on Innovative use of NLP for Building Educational Applications. https://doi.org/10.3115/v1/w15-0614

[16] Ng, H.T., Wu, S.M., Briscoe, T., Hadiwinoto, C., Susanto, R.H., Bryant, C. (2014). The Co NLL-2014 shared task on grammatical error correction. Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task. 1-14. https://doi.org/10.3115/v1/w14-1701

[17] Felice, M., Yuan, Z. (2014). Generating artificial errors for grammatical error correction. Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics. https://doi.org/10.3115/v1/e14-3013

[18] Zhang, J., Zong, C. (2015). Deep Neural Networks in Machine Translation: An Overview. IEEE Intelligent Systems, 30(5): 16-25. https://doi.org/10.1109/mis.2015.69

[19] Tang, D., Qin, B., Liu, T. (2015). Deep learning for sentiment analysis: successful approaches and future challenges. Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery, 5(6): 292-303. https://doi.org/10.1002/widm.1171

[20] Sun, S., Luo, C., Chen, J. (2016). A Review of Natural Language Processing Techniques for Opinion Mining Systems. Information Fusion, 36: 10-25. https://doi.org/10.1016/j.inffus.2016.10.004

[21] Pan, D., Zhou, H. (2018). English Learning System Design for College Students Personalized English Grammar Check and Diagnosis. International Journal of Emerging Technologies in Learning, 13(04): 21-32. https://doi.org/10.3991/ijet.v13i04.8467

8   Author

Gang Zhang, born in 1978, has gained the master's degree and is an associate professor in Henan Medical College. He is engaged in College English teaching and research works and is interested in college English teaching and research, English language and literature, British and American literature and English translation.