# Research on Chinese Text Error Correction Based on Sequence Model

Jianyong Duan
College of Information Science
North China University of Technology
Beijing, China
e-mail: duanjy@hotmail.com

Yang Yuan, Hao Wang, Xiaopeng Wei, Zheng Tan
College of Information Science
North China University of Technology
Beijing, China
e-mail: yuanyangeli@163.com

***Abstract*—When users input text, it will inevitably produce errors, and with the rapid development and popularization of smart devices, the situation becomes more and more serious. Therefore, text correction has become one of the important research directions in the field of natural language processing. As the grammatical error correction task, in this paper, the error correction process of Chinese text is regarded as the conversion process from wrong sentence to correct sentence. In order to adapt to this task, the (sequence-to-sequence) Seq2Seq model is introduced. The wrong sentence is used as the source sentence, and the correct sentence is used as the target sentence. Supervised training is carried out in units of characters and words. It can be used for correcting errors such as word of homophone, homotype, and near-sound, greatly reducing the artificial participation and expert support of feature extraction, improve model accuracy on specific errors. In order to solve the information loss caused by the conversion of long sequence to fixed length vector, the attention mechanism is introduced into the basic model. After adding the attention mechanism, the model's accuracy, recall rate and F1 value have been effectively improved.**

***Keywords- correction; Seq2Seq; attention***

## I. INTRODUCTION

Text proofreading technology is one of the important tasks of natural language processing. As early as the 1960s, some scholars conducted automatic proofreading research on English texts. IBM implemented a TYPO English spell checker by UNIX on IBM/360 and IBM/370 in 1960[1]; in 1971 Stanford University's Ralph Gorin implemented an English spell checker spell[2] on the EDC-10 machine. In recent years, with the continuous development of technology, Text proofreading research is also constantly making progress, and there are some commercialized results, such as Grammarly, Deal proof, Proofread and other special English word spell checking system. In the 1990s, Chinese scholars began to conduct research on Chinese text proofreading, but they developed rapidly. At present, many technology companies and universities or research institutions have invested a certain amount of human and material resources to carry out research in this area.

Among them, automatic error correction is an important part of automatic text proofreading. It provides suggestions for modification of error strings detected during automatic error detection, and assists users in correcting errors. The effectiveness of the proposed modification is the main indicator for measuring the automatic error correction performance. It has two requirements: the first proposed modification proposal should contain correct or reasonable suggestions, and the correct or reasonable recommendations should be arranged in front of all suggestions as much as possible. Therefore, the proposed algorithm and sorting algorithm are the two core topics of automatic error correction research. However, the Chinese text automatic proofreading theory is still not mature, and the discussion of automatic error correction is still rare.

At present, there are a number of commercial text proofing software products on the market. The representative products include the black horse proofreading system, the Founder Jinshan proofreading system, the woodpecker proofreading system, etc. They have obtained a certain degree of application in the publishing and printing industry and achieved good result. However, since most of the wrong language models are built based on the statistical features of words, the use of linguistic knowledge is still insufficient, and the error recovery and accuracy rate are still not up to the requirements of use. The efficiency of the error correction recommendation or the preferred correct rate is not high. It is difficult to find out the higher-level (such as sentence-level) errors, and there is still a big gap with the user's requirements. Therefore, the text proofreading technology needs to be further improved. And in the traditional Chinese text error correction field, the commonly used method is based on rules or statistical-based models. The disadvantage of the rule model is that it cannot fit all possible situations with simple rules, and new rules need to be added continuously. Both statistical models and rule models cannot take advantage of the semantic relationship of contexts.

With the successful application of deep learning in the NLP field, seq2seq has achieved good results in some end-to-end applications such as (automatic abstract, machine translation, text categorization, etc.). Different from the traditional method, the seq2seq model can learn the semantic relationship of the context, and can give more reasonable correction suggestions for the specific context.

## II. RELATED WORK

The traditional text proofreading is mainly divided into two parts. First, it is necessary to check the error, that is, in a sentence or an article, find the wrong position, and then replace the word or character in the wrong position with the correct word or character. Generally, several optimal options are recommended for the user to select during the error correction process.

At present, the error detection algorithm includes a context-based local language feature method; the transition probability is used to analyze the continuation relationship between adjacent words[3][4]; using rules or linguistic knowledge, such as grammar rules, word collocation, etc.; Statistical method. There are no obvious boundaries between the different methods and they can be mixed.

Context-based local language features such as the multi-feature-based Chinese automatic proofreading method designed and implemented by Microsoft China Research

Institute, which takes into account the local language features of words, words and part of speech in Chinese texts and the long-distance language features, and uses the Winnow method. Feature learning, using these context features to select the words in the target word confusion set, the main difficulty is how to convert the target sentence into multiple effective features and the acquisition of the confusion set[5]. Harbin Institute of Technology will search for possible candidates for each word in the proofreaded sentence, and form the word candidate matrix of the sentence. On this basis, using the structural features and statistical features of the language itself, from the candidate matrix. Select the best word candidate sequence for the sentence, compare it to the original sentence, find the wrong word, and correct it with the first candidate.

In the rule method, Yi Rongxiang, He Kekang, etc. use the revised grammar rules to proofread the manuscript[6]. If the sentence satisfies the rule of correcting grammar, the corresponding words are marked incorrectly according to the rules, but the limited rules are difficult to cover a large number of unpredictable The error phenomenon, limited ability to check. Liu Ting et al. used the clauses as a unit to scan Chinese sentences three times. Through automatic word segmentation, automatic recognition of new words, and the use of phrase rules to separate single words into phrases, and gradually bundle the correct strings[7]. The remaining single string that cannot be bundled is determined to be an error. The downside is that limited phrase binding rules are difficult to cover a large number of linguistic phenomena. Wu Yan et al.[8] also proposed a proofreading method combining word matching and grammar analysis. Using the combination of rules and statistics, the large-scale corpus is not used, the hash is found by the inverse maximum matching and the local corpus statistical algorithm, and the word matching and syntax analysis are performed on the hash, and then the candidate error string is found. The interactive method automatically corrects the error string and achieves a high error detection rate.

In the statistical-based method, Zhang Zhaohuang[9]proposed a method of using comprehensive approximate word set replacement and using statistical language model to score. The shortcoming is that it can only correct the so-called word errors, multi-words, missing words, and Bit errors are hard to find. Yu Yu, Yao Tianshun proposed a hybrid text proofreading method HMCTC, using pattern matching method to find the longest matching participle and finding long word errors. Then according to the ternary grammar, the frequency of co-occurrence with the adjacent words is less than a certain threshold. The word is marked as an error; finally, the word is marked with a grammatical attribute, and the error is marked in the sequence of the impossible grammar labeling sequence[10]. The disadvantage is that the error-checking criterion based on the co-occurrence frequency of words is limited by the size of the training corpus and the field of corpus selection, and the acquisition of the co-occurrence frequency data requires a large-scale segmented corpus, and such corpus is Hard to get. Sun Cai, Luo Zhensheng used the corpus statistical knowledge to guide the text proofreading[11], taking the sentence as the unit, treating the sentence as a field and a segment, calculating the average word frequency of the field and the average transition probability of the field; calculating the word transfer between words Probability, part-of-speech transition probability, the transition probability is used as the error-checking criterion, and the word or word whose transition probability is less than the threshold is taken as the detected error.

While error correction is another important component of text proofreading, the current mature theory is still rare. Yu Yu, Yao Tianshun, etc. used the pattern matching method to correct the long words[10], but did not make full use of the characteristics of the error string, the algorithm is computationally intensive. The IBM China Research Center[12] proposes an alternative word table combined with the main dictionary, which provides an error correction algorithm for modifying the detected error string by adding words and changing words, but the algorithm's error correction suggestions are limited to the replacement word. Table, without considering context-inspired information, mainly considers error correction for erroneous types, and has weaker error correction capabilities for missed words, multi-words, transpositions, multi-word substitutions, and English word spelling. Zhang Yangsen[13] proposed a candidate set generation algorithm for error correction based on likelihood matching, which greatly improved the error correction ability of missed words, multi-words, transpositions and multi-word substitutions.

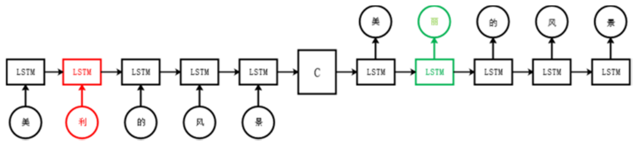

Figure 1 Figure of seq2seq model structure

However, it is worth noting that most of the above mentioned methods require complicated and cumbersome manual processing, as well as expert technical support, which consumes a lot of manpower and material resources. In the field of deep learning, some scholars try to use deep learning to correct grammatical errors[14][15]. To solve this problem, as the grammatical error correction task, this paper proposes a Chinese text error correction model based on sequence-to-sequence (hereinafter referred to as seq2seq), as shown in Fig 1, using the encoder-decoder structure to solve the conversion process of error text to the correct text, the left side is the encoding end, the right side is the decoding end, the encoding end and decoding. The LSTM[17] structure is adopted at the end, and the encoding end generates the semantic vector C of the entire sentence after the loop iteration, and the decoding end decodes the generated vector C into the corresponding text, and completes the conversion of the error text to the correct text. This model is different from the traditional rule-based and statistics-based methods, which effectively reduces manual participation and expert support, and can learn context-related semantic relations. When providing recommendations, it can be targeted for a certain sentence. Error correction scheme. At the same time, in order to alleviate the information loss caused by the seq2seq structure due to the long input sequence, the attention mechanism is introduced. The experiment proves that the seq2seq structure can effectively adapt to the text error correction task, and introduces the attention mechanism to effectively improve the evaluation indicators of the text error correction task.

# III. MODEL

The text error correction task can be regarded as the transformation process between different sequences. The error sentence is regarded as the source language and the correct sentence is regarded as the target language. Therefore, the seq2seq model is introduced into the error correction task as a sequence conversion model, and seq2seq is used as The deep learning model can better fit the conversion of the wrong statement to the correct statement compared to the traditional rule model and statistical model. In addition, in order to strengthen the correspondence between the encoding end and the decoding end, attention mechanism is introduced.

The specific process of the Seq2seq+attention model is expressed as follows: Let an input statement be s = $a_1, a_2, a_3, .., a_t$, where $a_i$ represents the $i$ th character and $m$ represents the length of the sentence. Each input statement will correspond to a correct sentence of the same correct length, and the content is one-to-one correspondence. In this paper, the error correction task is regarded as a machine translation task, the error sentence is treated as the source sentence, and the correct sentence is treated as the target sentence. Figure 2 shows the basic seq2seq+attention model structure, which rewrites the erroneous sentence “美利的风景（Murray's landscape）” into a “美丽的风景（beautiful landscape）” and completes the “translation process” from the wrong sentence to the correct sentence, in which the red The part indicates the wrong word in the original sentence, and the green part indicates the correct word in the target sentence.

The Embedding layer encodes the characters $a_1, a_2, a_3, .., a_t$ into word vectors, inputs them into the LSTM layer of the encoder, and after loop iteration in order, obtains the semantic vector C of the entire sentence, and the hidden state H = $h_0, h_1, h_2, ..., h_t$. Vector C is used to initialize the initial state of the decoder. The initial state of the input is fixed by "sos" and "0". After combining the two contents, as the initial input, the LSTM layer obtains an output $s_1$, $s_1$' at the attention layer. The calculation is performed with H to obtain the attention context. The context is merged with the character vector predicted in the previous step and then used as the input of the next LSTM cell at the decoding end to predict the information of the current node. The specific details of the Embedding layer, the LSTM layer, and the attention are described below.

## A. Embedding layer

We set the size of the word vector to a fixed dimension. The size of the word vector in the source sentence is 500 dimensions, and the target sentence is also 500 dimensions. As shown in Figure 2, for each input character $a_t$, the corresponding The word vector is represented as shown in equation(1), where e represents a lookup table of word vectors. That is, for the characters in the sentence, such as "利 (profit)", we all randomly generate a 500-dimensional vector by uniform distribution. In the process of training, the weight of each reverse derivative is constantly changing. At the end of training, the weight of the model Learning the context of the statement and the general law, the vector of the word "利（profit）" is the optimal expression of the wrong sentence.

$$X_t = \mathrm{e}(a_t) \qquad (1)$$

## B. LSTM layer

The model's encoder and decoder sides use the LSTM[16] structure. The LSTM is the optimized structure

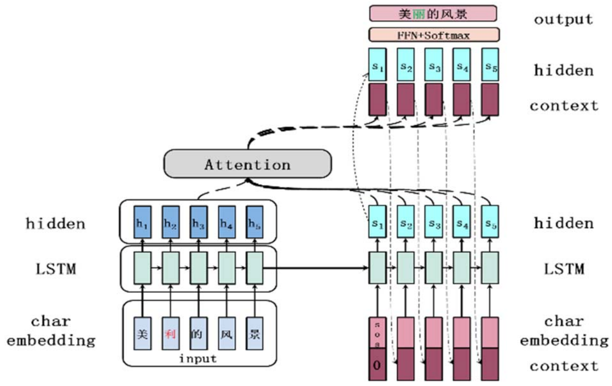


Figure 2 Seq2seq+attention structure

of the RNN. It uses an additional gate mechanism to effectively maintain long distances. The information and the problem of gradient disappearance and gradient explosion of the RNN structure are alleviated. As shown in equation (2)(3)(4):

$$\begin{bmatrix} o_t \\ f_t \\ \tilde{c}_t \\ i_t \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ tanh \\ \sigma \end{bmatrix} \left( W^T \begin{bmatrix} x_t \\ h_{t-1} \end{bmatrix} + b \right) \qquad (2)$$

$$C_t = C_{t-1} * f_t + i_t * \tilde{C}_t \qquad (3)$$

$$h_t = o_t * \tanh(C_t) \qquad (4)$$

Where $\mathrm{f}_t$, $o_t$ and $i_t$ are forgotten gates, output gates and input gates respectively, $W^T$ and b are parameters of the model, σ represents the sigmoid function, tanh represents the hyperbolic tangent function, $\tilde{c}_t$ is the input data of the current node, $C_{t-1}$ is the output information of the previous LSTM node. In order to make full use of the information in both directions, this paper uses a bidirectional LSTM structure. Each input sentence $a_1, a_2, a_3, .., a_t$, respectively Processing in both positive and negative directions to obtain past and future textual information.

$$\overrightarrow{h_1}, \overrightarrow{h_2}, \dots, \overrightarrow{h_t} = \overrightarrow{LSTM}(x_1, x_2, \dots, x_t) \qquad (5)$$

$$\overleftarrow{h_1}, \overleftarrow{h_2}, \dots, \overleftarrow{h_t} = \overleftarrow{LSTM}(x_1, x_2, \dots, x_t) \qquad (6)$$

Then merge the two hidden layers into the final output.

$$\mathrm{h_i} = [\overrightarrow{h_\iota}; \overleftarrow{h_\iota}] \qquad (7)$$

## C. Attention layer

The During the conversion process from the wrong text to the correct text, we found that in the process of transforming “美利的风景（Murray's Landscape）” to the “美丽的风景（beautiful landscape）”, the source sentence “美利的风景（Murray's Landscape）” finally generates a fixed vector in the loop iteration, and the information is lost seriously. when decoding "利(profit)", in the fixed vector generated by the encoding segment, the corresponding "利(profit)" information has been depleted, which will lead to errors in the conversion process of the wrong text and the correct text.

In order to solve the problem of information loss caused by long sequence to fixed length vector transformation, this paper introduces the Attention mechanism, as shown in Figure 3. At the encoding end is a forward-propagating RNN that generates translation results by the equation(8):

$$\mathrm{p}(y_t|y_{t<1},\mathrm{X}) = \mathrm{g}(y_{t-1}, s_t, m_t) \quad (8)$$

g() is a linear function, $\mathrm{s}_t$ and $\mathrm{m}_t$ represent the jth time step decoding state and the source text content, respectively, where st is calculated as equation (9).

$$s_t = f(s_{t-1}, y_{t-1}, m_t) \quad (9)$$

f() is an activation function, such as the LSTM function. According to the principle of the attention mechanism, this article defines $\mathrm{m}_t$ as the weighted sum of the source output $\mathrm{h}_i$:

$$m_t = \sum_{i=1}^{I} \alpha_{t,i}.h_i \quad (10)$$

$\alpha_{t,i}$ indicates that the degree of matching between $s_{t-1}$and $\mathrm{h}_i$ is as equation (11)(12):

$$\alpha_{t,i} = \frac{\exp(e_{t,i})}{\sum_{i'=1}^{I}\exp(e_{t,i'})} \quad (11)$$

$$e_{t,i} = v_a^T\, tan\mathrm{h}(\, W_a s_{t-1} + U_a \mathrm{h}_i + b) \quad (12)$$

$\mathrm{W}_a$ $\mathrm{U}_a$ and $\mathrm{V}_a$ are the weight matrix of the attention matrix. With this model, the decoder automatically selects the vocabulary in the source statement associated with the target word being generated.

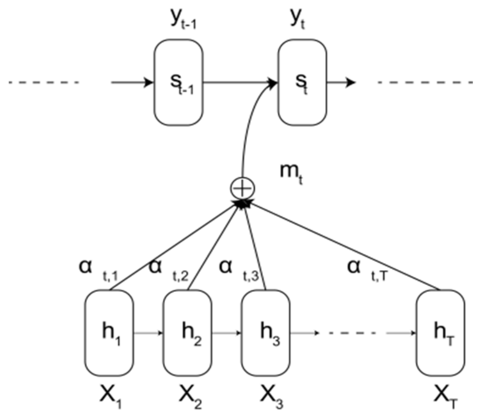

Figure 3 Attention structure

The key operation here is to calculate the weight of the correlation between the encoder and the decoder state, and obtain the Attention distribution, so that the weight of the input position that is more important for the current output position will account for a larger proportion when predicting the output. For example, in the process of transforming "美利的风景（Murray's Landscape）" to "美丽的风景（beautiful landscape）", when the word "丽（beautiful）" is decoded, the corresponding weight α of the word "利（profit）" in the source sentence is gradually increased after training, that is, The decoding end adds the information input of the word "利（profit）", and corrects the error of the "利（profit）" word to improve the accuracy of the model.

## IV. EXPERIMENT

### A. Data

The text errors generated by the input method mainly focus on homophone errors, homonym errors, and near-word errors. This paper mainly uses the above three types of errors as error correction objects. The dataset of this paper is derived from the tencent dataset[17], with 230,000 training sets and 10,000 validation sets and test sets. The training set verification set, the data format is the form of the wrong sentence corresponding to the correct sentence, in the test set, replace the four thousand error data into four thousand correct data, used to calculate the accuracy, recall rate and other evaluation values. In all data sets, each sentence has only one error, the same type, the same tone or the near sound. The data is shown in the following table:

TABLE 1:TRANING DATA、VALITATION DATA、TEST DATA

| dataset | | consititution | total_num | right/wrong | num |
|---|---|---|---|---|---|
| tencent | Train | src-train | 230000 | right | 0 |
| | | | | wrong | 230000 |
| | | tgt-train | 230000 | right | 230000 |
| | | | | wrong | 0 |
| | Valid | src-val | 10000 | right | 0 |
| | | | | wrong | 10000 |
| | | tgt-val | 10000 | right | 10000 |
| | | | | wrong | 0 |
| | Test | src-test | 10000 | right | 4000 |
| | | | | wrong | 6000 |

### B. Parameter settings

As shown in Table 2, for the model, we set the word vector dimension to 500, the dropout to 0.3, and the parameter uniform distribution to (-0.1, 0.1). This article uses the ADAM optimizer[18] and sets the learning rate. For 0.001, a total of 100,000 rounds of training are set, and the model effect is checked every 10,000 rounds. After 100,000 rounds, the training is ended. The maximum batch of training is 64. In the coding segment, different coding modes RNN, BRNN, and different encode layer (2, 4) is set for different models, and the RNN node is uniformly set to the LSTM structure for whether to use the add attention mechanism to set the parameter to none or generate.

TABLE 2: MODEL SUPER PARAMETER

| parameter | value | parameter | value |
|---|---|---|---|
| encoder_type | rnn/brnn | attention | none/generate |
| optim | adam | rnn_type | lstm |
| learning_rate | 0.001 | dropout | 0.3 |
| enc_layers | 2/4 | enc hidden size | 500 |
| dec hidden size | 500 | embedding size | 500 |

### C. Evaluation standard

In order to reasonably evaluate the model, this paper designs the following evaluation criteria, TP: number of correct corrections for incorrect sentences; FP: number of wrong corrections or not corrected for incorrect sentences; TN: no corrections for correct sentences; FN:

changes of correct sentences.The accuracy, recall and F1 are calculated using the equation(13)(14):

$$P = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN} \tag{13}$$

The F1 value is calculated as shown in the following formula:

$$F_1 = \frac{2PR}{P + R} \tag{14}$$

*D. Experimental design*

The experiment is divided into two parts. The first experiment is the model tuning part. In this part, five sets of comparative experiments are performed, namely seq2seq, seq2seq+attention(2 layers)/(4 layers), and seq2seq+attention+brnn (2 layers)/(4 layers), in which each set of experiments was set up for multiple experiments in order to obtain stable results, and finally averaged. Here 2 layers and 4 layers represent the number of layers of the LSTM layer at the encoding end. In the second part, the data is sent to the model in units of words and words for training, and the effects of different granularity data on the model's effect are compared and analyzed.

*E. Experiment results and analysis*

As shown in Table 3, the basic seq2seq model has the worst performance. After adding the attention mechanism, the scores are greatly improved. The results show that the attention mechanism is very helpful for the improvement of text correction.

Among all the experimental average results, Seq2seq+attention (4 layers) has the highest accuracy, recall and F1 values, seq2seq+attention+brnn (4 layers) accuracy, and the recall and F1 values are equally high, respectively, 58.3%, 68.2%, 31.4%. However, after comparing Seq2seq+attention(4layers), seq2seq+attention+brnn (4 layers), it is found that after the BRNN is used, there is no improvement in the error correction effect, indicating that the BRNN structure does not help the error correction task.

TABLE 3:VALUATE SCORE OF DIFFERENT MODEL

| | Model | P | R | F1 |
|---|---|---|---|---|
| Enc 2 layers | Seq2seq | 23.18 | 38.46 | 28.93 |
| | Seq2seq+attention | **57.47** | **73.37** | **64.30** |
| | Seq2seq+attention+brnn | 49.92 | 57.91 | 53.62 |
| Enc 4 layers | Seq2seq+attention | **58.51** | **81.46** | **68.11** |
| | Seq2seq+attention+brnn | 57.92 | 81.84 | 67.83 |

In order to make the data more accurate, this paper designed a number of experiments, and finally averaged, the results obtained confirm the original idea, the best model is Seq2seq+attention (4 layers). It can be found that in the error correction task, the BRNN may introduce noise and affect the effect of the model.

Comparing the two models of Seq2seq+attention(2 layers) and Seq2seq+attention(4 layers), we can find the four-layer LSTM structure, which is very helpful for the Chinese text correction task and can help to adapt to the error correction task. And effectively improve the accuracy, recall rate, F1 value and other evaluation indicators. A large part of the reason for the data increase is that the size of the parameters has increased, and the model can better fit the text error correction task, thereby improving various evaluation indicators.

In order to better evaluate the model effect, we analyzed the results of the test data prediction. The analysis results are as Figure 4.

Figure 4 take the model's generated data for analysis. The model parameters are (adam lr(0.001) rnn attention 2layers), and the test data are all sentences with errors. Among all the data generated by the test data, there are 10049 pieces of data. The length of the sentence of the prediction result is the same as the length of the sentence of the standard answer, but there are also some pieces of data whose predicted answer is different from the sentence length of the standard answer. This type of data accounts for 2%. There are 11 numbers in the same length of data that have errors (that is, the wrong sentence is the same as the correct sentence), and this data needs to be filtered out. There are 5,857 correct correctly sentences, accounting for 57.1% of all data. There are 4248 uncorrected and corrected data, accounting for 41.43%, of which 1768 are correct errors, accounting for 17.2%. What can be found is that error correction and uncorrected errors account for a large proportion of the total data, and the model has a lot of room for improvement.

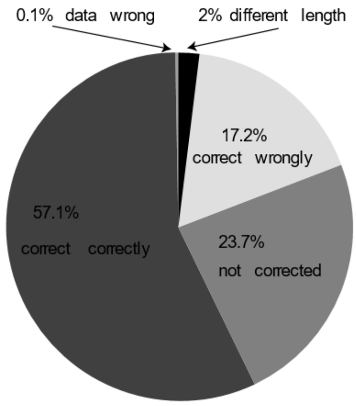


Figure 4 Analysis of prediction results

TABLE 4:RATING RESULT OF CHAR SIZE AND WORD SIZE MODEL

| Model | P | R | F1 |
|---|---|---|---|
| Char | **58.51** | **81.46** | **68.11** |
| Word | 28.11 | 25.42 | 13.31 |

The second experiment is based on the first experiment. The pre-processed Chinese text data is firstly segmented using the jieba word segmentation tool, and then sent to the model for training. The results are shown in Table 5.

The results show that the results obtained using the word segmentation data are generally inferior to the evaluation indicators of the character-based data. The reason is that when using the word segmentation tool for word segmentation, it is inevitable to introduce word segmentation errors, and in the process of "translation", training in word units, the lexicon shared by both parties will be reduced, in "translation" Extra-territorial words (OOV) often occur, resulting in a decrease in error correction.

## V. CONCLUSION

Based on the basic seq2seq model, the Attention mechanism effectively improves the error correction effect, and the multi-layer LSTM Layer helps improve the accuracy of the model. Training is required to use character-based data for training, which can reduce the impact of word segmentation. It can be seen that the Seq2seq model is well adapted to the Chinese text error correction task, and the accuracy, recall rate and F1 value can reach 58.4%, 68.7%, and 31.6%, respectively, and have achieved good results. It can be seen that the text-based error correction model based on the sequence model has a higher score on the error correction task. In t

he case of similar scores, this model can greatly reduce the workload of manually extracting features, and adopts the end-to-end approach directly. Solve the problem of error correction. However, there are still many improvements in the model, such as the problem of miscorrecting and correcting, how to make targeted changes to the attention mechanism to reduce the occurrence of miscorrecting and correcting errors. Modify the input mode of the model so that the model can use the information of the word and use the character information to help the model improve the effect. The next work is mainly to solve these two problems.

## ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (61672040, 61972003), Beijing Urban Governance Research Center and the North China University of Technology Startup Fund.

## REFERENCES

[1] Kukich K. Techniques for automatically correcting words in text[J]. Acm Computing Surveys (CSUR), 1992, 24(4): 377-439.

[2] James L. Peterson. 1980. Computer programs for detecting and correcting spelling errors. Commun. ACM 23, 12 (December 1980), 676-687. DOI: https://doi.org/10.1145/359038.359041

[3] Riseman E M, Ehrich R W. Contextual word recognition using binary digrams[J]. IEEE Transactions on Computers, 1971, 100(4): 397-403.

[4] Golding A R, Schabes Y. Combining trigram-based and feature-based methods for context-sensitive spelling correction[C]//Proceedings of the 34th annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 1996: 71-78.

[5] Lau R, Rosenfeld R, Roukos S. Trigger-based language models: A maximum entropy approach[C]//1993 IEEE International Conference on Acoustics, Speech, and Signal Processing. IEEE, 1993, 2: 45-48.

[6] Yi Rongxiang, He Kekang. Computer Chinese proofreading system [D]. , 1997.

[7] Liu Ting, et al. Principles of Chinese Computer Aided Proofreading System [J]. Chinese Information, 1997.

[8] Wu Yan, Li Xiukun, Liu Ting, et al. Research and implementation of Chinese automatic proofreading system [J]. Journal of Harbin Institute of Technology, 2001, (2): 60-64.

[9] Chang, C.H.. (1994). A pilot study on automatic Chinese spelling error correction. Communication of COLIPS. 4. 143-149.

[10] Yu Wei, Yao Tianshun. A mixed Chinese text proofreading method [J]. Chinese Informatics, 1998, 12(2).

[11] Sun Cai, Luo Zhensheng. Research on the Correction of Chinese Text Proofreading Words[A]. Proceedings of the Fourth Conference on Computational Linguistics (Language Engineering) [C]. Beijing: Tsinghua University Press, 1997. 319 -324

[12] Guo Zhili, et al. Algorithm for modifying suggestions in Chinese proofreading system [A]. Proceedings of the 4th Conference on Computational Linguistics (Language Engineering) [C] . Beijing: Tsinghua University Press, 1997. 325-330 .

[13] Zhang Yangsen. The construction of error correction knowledge base in Chinese proofreading system and the generation of error correction suggestions [J]. Chinese Journal of Information, 2001, 15( 5): 33-39.

[14] Grammatical error correction using neural machine translation. Zheng Yuan and Ted Briscoe. NAACL 2016.

[15] A Multilayer Convolutional Encoder-Decoder Neural Network for Grammatical Error Correction. Shamil Chollampatt, Hwee Tou Ng. AAAI 2018.

[16] Mnih V, Heess N, Graves A. Recurrent models of visual attention[C]//Advances in neural information processing systems. 2014: 2204-2212.

[17] Wang D, Song Y, Li J, et al. A Hybrid Approach to Automatic Corpus Generation for Chinese Spelling Check[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018: 2517-2527.

[18] Kingma D, Ba J. Adam: a method for stochastic optimization (2014)[J]. arXiv preprint arXiv:1412.6980, 2015, 15.