

Research Article

Shan Chen and Yingmei Xiao*

An intelligent error correction model for English grammar with hybrid attention mechanism and RNN algorithm

<https://doi.org/10.1515/jisys-2023-0170>

received September 14, 2023; accepted January 19, 2024

Abstract: This article proposes an English grammar intelligent error correction model based on the attention mechanism and Recurrent Neural Network (RNN) algorithm. It aims to improve the accuracy and effectiveness of error correction by combining the powerful context-capturing ability of the attention mechanism with the sequential modeling ability of RNN. First, based on the improvement of recurrent neural networks, a bidirectional gated recurrent network is added to form a dual encoder structure. The encoder is responsible for reading and understanding the input text, while the decoder is responsible for generating the corrected text. Second, the attention mechanism is introduced into the decoder to convert the output of the encoder into the attention probability distribution for integration. This allows the model to focus on the relevant input word as it generates each corrected word. The results of the study showed that the model was 2.35% points higher than statistical machine translation–neural machine translation in the CoNLL-2014 test set, and only 1.24 points lower than the human assessment score, almost close to the human assessment level. The model proposed in this study not only created a new way of English grammar error correction based on the attention mechanism and RNN algorithm in theory but also effectively improved the accuracy and efficiency of English grammar error correction in practice. It further provides English learners with higher-quality intelligent error correction tools, which can help them learn and improve their English level more effectively.

Keywords: attention mechanism, recurrent neural network, Transformer, English grammar error correction

1 Introduction

With the continuous development of social economy and information technology, English has become an important text carrier for information exchange [1]. Influenced by the language environment and the complexity of grammatical structure, learners from all occupations will inevitably make grammatical mistakes when using English [2]. English grammar intelligent error correction model plays an important role in the field of natural language processing and is widely used in various text processing and editing software to help improve grammar level as well as text quality and readability [3]. However, the traditional English grammar correction model often ignores contextual information and word order information, resulting in limited accuracy and efficiency of error correction [4]. To improve the accuracy of English expression and avoid the obstacles of foreign communication, this study proposes an English grammar intelligence error correction model that combines an attention mechanism and Recurrent Neural Network (RNN) algorithm. The aim is to

* **Corresponding author: Yingmei Xiao**, Humanities Quality Department, Hunan Technical College of Railway High-Speed, Hengyang, 427000, China, e-mail: Yingmei52Xiao023@outlook.com

Shan Chen: Department of Public Courses, Changde Vocational Technical College, Changde, 415000, China; Graduate School, University of Perpetual Help System DALTA, Manila, 0900, Philippines, e-mail: chen2shan1@163.com

improve the accuracy and efficiency of error correction by combining the powerful contextual capture capability of the attention mechanism with the excellent sequence modeling capability of RNN. This model not only broadens the research scope of English grammar correction methods but also enables English learners and text processing software to obtain higher-quality error correction results through this model, thus improving the effect of English learning and text processing. The error correction model based on the attention mechanism and RNN algorithm not only has a profound impact on the field of English grammar correction but also has a positive impact on other natural language processing tasks.

The research content includes five parts. The first part is the introduction, which describes the importance of the English grammar correction model to people's language communication and text learning under the background of rapid development of information technology. The second part is a literature review, the realization of the English grammar correction model and deep learning algorithm in various fields, and the research status of many scholars on the English grammar correction model. The third part is the study of the English grammar intelligent error correction model based on the mixed attention mechanism and RNN algorithm. The first section is the study of English grammar correction framework based on converters, the second section is the study of encoder model structure based on RNN, and the third section is the study of decoder model structure based on attention mechanism. In the fourth part, the proposed English grammar error correction model is tested and analyzed. The fifth part is the summary and prospect of the research methods and results.

2 Related works

With the continuous development of network information technology, deep learning technology with RNN or attention mechanism as the core has been applied in many research fields and has attracted wide attention in the field of natural language processing. Zhang et al. proposed a hybrid generative model based on RNNs and generative adversarial networks. The model learned the two-dimensional morphological characteristics and spatial relationships of pores through two-dimensional slices and restored the complete three-dimensional structure layer by layer. The research results showed that the model exhibited accuracy, diversity, and stability in reconstruction experiments of homogeneous and heterogeneous porous media and fractured cores, and accurately retained statistical features, morphological features, and long-distance connectivity based on 2D input images [5]. Wu and Li proposed a new neural symbolic reasoning method, RNNCTP, which consisted of a relational selector and a predictor, and refiltered the knowledge selection of the conditional theorem prover. After efficient interpretability training, the entire model dynamically generated the knowledge required for the inference of the predictor. The research results showed that this method was competitive with traditional methods in link prediction tasks and was more applicable than CTP on some data sets [6]. Shahkarami et al. proposed a new network architecture that consisted of convolutional neural network encoders and one-way many-to-one vanilla RNNs working together to capture one set of channel damages while compensating for the shortcomings of the other set of channels. The results showed that the hybrid model achieved the lowest error probability in both receiver configurations, and the complexity was reduced by more than 50% [7]. Zhang et al. proposed a parallel non-Cartesian convolutional RNN approach for the reconstruction of dynamic parallel MR Data from under-sampled non-Cartesian abdomens. This method used redundant information in the space and time domain to achieve high data fidelity reconstruction of non-Cartesian parallel MR Data. The research results showed that the performance of the method in different acceleration rates, motion modes, and imaging applications was significantly improved [8]. Gao et al. proposed a wood-node defect recognition model of SE-ResNet18, which combined convolutional neural networks, attention mechanisms, and transfer learning. The model also effectively reduced the parameters of the fully connected layer by replacing the fully connected layer with the global average pooling layer. The research results showed that the accuracy of the model in the test set was as high as 98.85%, which provided a new and effective method for non-destructive testing of wood [9].

Combining the advantages of attention mechanism and RNN to intelligently correct English grammar errors has opened a new perspective for the development of natural language processing technology, and many scholars have studied it. Ranalli and Yamashita evaluated Grammarly's automated written error correction feedback tool, which was based on complex error correction techniques and had the potential to discover and correct common errors in second languages. The research results indicated that the common types of errors in this model in a second language were ten times higher than those in the same student text corpus, which had practical significance for using written feedback tools in a second language environment [10]. Trappey et al. developed an intelligent summarization method that integrated natural language processing, text mining, and machine learning. The method used text analysis to automatically extract the basic specifications from it, and the K-means algorithm grouped the sentences for each specification. Research results showed that this system helped manufacturers improve complex product designs and improved cost estimation and quotation competitiveness [11]. Makwana et al. measured emotional intelligence using self-reported and performance tests and biases against various out-groups. The research highlighted the importance of emotional intelligence in interpersonal relationships, the key role of emotions in out-of-group bias, and the relationship between emotional intelligence and out-of-group bias. Results showed that those with stronger performance-based emotion management skills showed lower general racial prejudice and more positive attitudes toward immigrants and refugees [12]. Zhou and Liu proposed an English grammar error correction algorithm based on a classification model, analyzed the model architecture and model optimizer of the grammar error correction algorithm, and finally conducted a simulation experiment and analyzed the results. The research results showed that the proposed English grammar error correction algorithm based on the classification model continuously improved its classification accuracy, reduced processing time, saved storage space, and simplified processing flow [13]. Solymann et al. proposed rule-based statistical machine translation and neural machine translation. To overcome the exposure bias problem, a bidirectional regularization term utilizing Kullback–Leibler divergence was introduced into the training target to improve the consistency between right-to-left and left-to-right models. The results showed that compared with the existing Arab GEC system, the proposed model obtained the best *F1* score [14]. The reference comparison table is shown in Table 1.

In summary, with the development of deep learning technology and the continuous enrichment of training data, the English grammar intelligent error correction model integrating attention mechanism and RNN has significant application effects in various fields. Aiming at the problem of intelligent error correction in English grammar, this research method shows remarkable innovation compared with the traditional model. In this model, an attention mechanism is introduced to improve the ability of context capture, and RNN is used to optimize the processing of word order so that the accuracy and efficiency of the model are significantly improved compared with the traditional methods. However, the model still has room for improvement in dealing with advanced and complex grammatical structures. However, the model proposed in this study is a great progress in the field of English grammar intelligent error correction and provides a new direction for further research. By further optimizing the combination of attention mechanism and RNN algorithm, as well as the processing of advanced and complex grammatical structures, the model will be pushed forward to higher accuracy and wider application. It is expected to promote the further development of the field of English grammar correction and provide more efficient learning tools for English learners.

3 English grammar error correction model construction

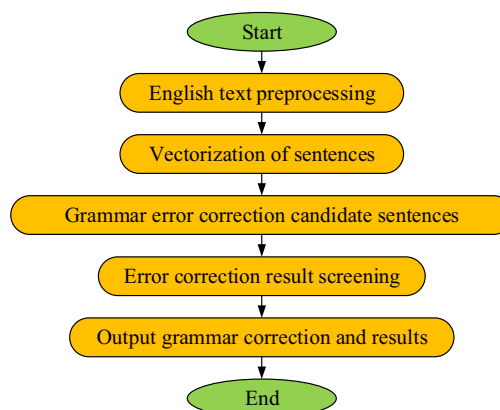
3.1 Transformer-based English grammar error correction framework

Intelligent English grammar error correction refers to the automatically corrected grammatical errors that occur in English texts using computer programs. The mainstream approach is to transform the process of grammar error correction into the translation process according to the principle of machine translation, i.e., to

Table 1: Comparison of references

Author	Method proposed	Method description	Research results
Wu and Li	Neural symbolic reasoning method, RNNCTP	Consists of a relational selector and a predictor, and re-filters the knowledge selection of the conditional theorem prover	Competitive with traditional methods in link prediction tasks, and more applicable than CTP on some data sets [6]
Shahkarami et al.	New network architecture	Consists of convolutional neural network encoders and one-way many-to-one vanilla RNNs	Achieves the lowest error probability in both receiver configurations, and the complexity is reduced by more than 50% [7]
Zhang et al.	Parallel non-Cartesian convolutional recurrent neural network	Uses redundant information in space and time domain to achieve high data fidelity reconstruction of non-Cartesian parallel MR data	Performance has been significantly improved in different acceleration rates, motion modes, and imaging applications [8]
Gao et al.	Wood-node defect recognition model of SE-ResNet18	Combines convolutional neural networks, attention mechanisms, and transfer learning	The accuracy of the model in the test set is as high as 98.85%, which provides a new and effective method for non-destructive testing of wood [9]
Trappey et al.	Intelligent summarization	Natural language processing, text mining, machine learning, K-means algorithm	System improves complex product designs and cost estimation and quotation competitiveness [11]
Makwana et al.	Emotional intelligence	Self-reported and performance tests	Those with stronger performance-based emotion management skills show lower general racial prejudice and more positive attitudes toward immigrants and refugees [12]
Zhou and Liu	English grammar error correction	Classification model algorithm	The proposed model improves classification accuracy, reduces processing time, saves storage space, and simplifies processing flow [13]
Solyman et al.	Rule-based statistical machine translation and neural machine translation	Bidirectional regularization term utilizing Kullback–Leibler divergence	The proposed model obtains the best <i>F1</i> score compared to the existing Arab GEC system [14]

re-translate the grammatically problematic sentences into correct ones through an algorithmic model [15,16]. Based on this, the study takes the Transformer model, which is more effective in machine translation, as the base model of the correction system, and achieves the correction of incorrect grammar in the text by improving the encoder and decoder. Figure 1 shows its overall flow chart.

**Figure 1:** Flow chart of grammar error correction model.

In Figure 1, the sentences need to be pre-processed and vectorized before text syntax error correction. The preprocessing process uses data augmentation methods to add tags to the sentences. Vectorization, on the other hand, uses the ALBERT algorithm to obtain information about the lexical and semantic aspects of the sentence [17,18]. Then, the Transformer model corrects errors and generates a set of candidate sentences. Finally, the correct results are filtered and output by Beam search [19]. In the English text preprocessing stage, this study uses a fluency-based data augmentation method [20,21]. The expression for calculating fluency is shown in equation (1).

$$\begin{cases} f(x) = \frac{1}{1 + H(x)} \\ H(x) = -\frac{\sum_{i=1}^{|x|} \log P(x_i|x_{<i})}{|x|}. \end{cases} \quad (1)$$

In equation (1), x and $|x|$ are the string and sentence length, respectively. $|x|$ denotes the probability of the occurrence of word x_i when $x < i$. $f(x)$ and $H(x)$ are the fluency score and cross-entropy of the sentence. On this basis, the fluency of the corrected sentences is boosted by combining equation (2). $D(x)$ is the set of eligible candidate sentences. x^o and x^r are the sentences with grammatical errors and the given standard sentences. y_n is the prediction result obtained based on x^r . σ can specify the degree of similarity between the output result and the standard sentences in fluency, which is set to 80% here. The candidate sentences with higher quality can be filtered by equations (1) and (2).

$$D(x^r) = D(x^r) \cup \left\{ x^o \mid x^o \in y_n(x^r; \theta_{\text{crt}}) \wedge \frac{f(x^o)}{f(x^r)} \geq \sigma \right\}. \quad (2)$$

The candidate sentences obtained from the above operations were augmented with three corpora, ICNALE, NUCLE, and CLEC, respectively. The sentences are then fed into the Transformer model using the dynamic word vector representation of the ALBERT model trained by the NUCLE corpus. In Figure 2, the model includes an encoder and a decoder, and this study focuses on improving the encoder and decoder parts.

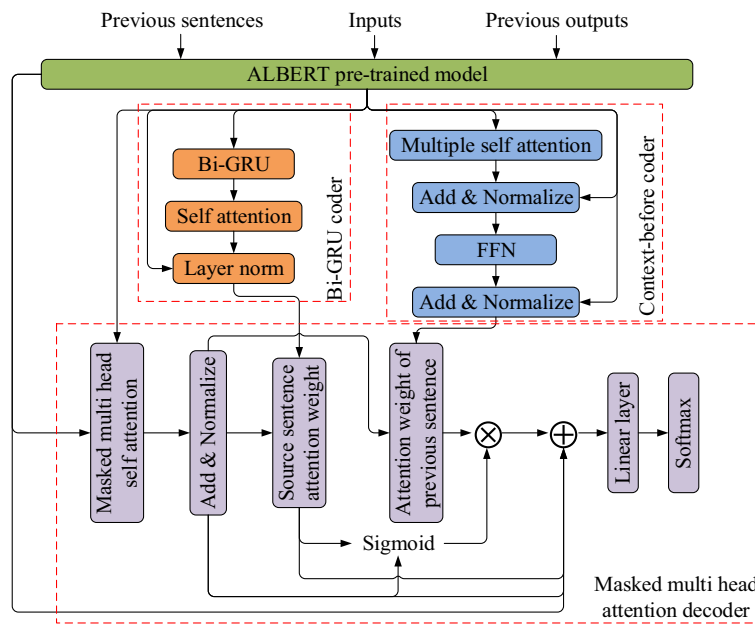


Figure 2: Improved Transformer model structure.

3.2 RNN-based encoder model architecture

A correct utterance expression not only requires accurate wording but also considers the logical relationships between utterances and the phrasing environment. Traditional encoders can only focus on the internal connections of individual sentences and cannot extract the information between sentences, which can affect the effectiveness of the application of grammar error correction models. Inspired by the auxiliary encoder proposed by Mohammad et al. [22], the study considers adding another RNN encoder to the encoding module of the Transformer model for the learning ability of the encoding module on the semantic relations between sentences. RNNs are more fundamental algorithmic models in deep learning algorithms, and their structure contains feedback connections that allow them to be cyclically passed through the network and retain unit states for processing and memorization of time sequences [23,24]. Sentences in natural language can be viewed as sequences of words formed in temporal order, so RNNs are commonly used in machine translation, speech recognition, etc. RNN includes one input, one output, and one hidden layer [25].

The RNN can transform the sequence $X = \{x_1, x_2, \dots, x_n\}$ in the input layer into the hidden state s_t at a fixed length by the weight matrices U and W and obtain the output result o_t by the weight matrix V . The calculations associated with o_t and s_t are shown in equation (3). f and g both represent activation functions, which can be sigmoid, tanh, softmax, etc.

$$\begin{cases} s_t = f(U \cdot x_t + W \cdot s_{t-1} + b_s) \\ o_t = g(V \cdot s_t + b_o). \end{cases} \quad (3)$$

The earlier the sequence elements are input during the update process, the smaller their occupied weights are, and they exhibit temporal correlation. However, for long sequence input, the RNN model is prone to gradient disappearance and gradient explosion. When applied to the English grammar error correction model, it shows that the longer the sentence length is, the less information is retained and the worse the actual effect of the model is. The RNN-based Gated Recurrent Units (GRU) can solve this problem [26,27], and its structure is shown in Figure 3.

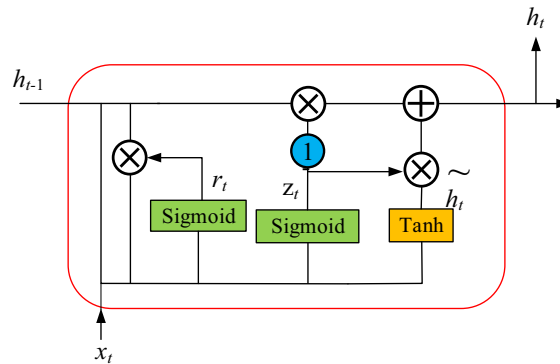


Figure 3: Structure of the gating cycle unit.

In Figure 3, according to the hidden state h_{t-1} at the moment $t - 1$ and the input t at the moment t , the reset gate r_t and the update gate z_t can be obtained, and the related calculation is shown in equation (2). w_r and w_z are the weights of the reset gate and update gate, respectively; the function sigmoid can map the output of the hidden layer to the interval $[0, 1]$.

$$\begin{cases} r_t = \sigma(w_r[h_{t-1}, x_t]) \\ z_t = \sigma(w_z[h_{t-1}, x_t]). \end{cases} \quad (4)$$

The input information at t can be remembered by h'_t , and the activation function \tanh maps the value of h'_t to the interval $[-1, 1]$. Combined with z_t , the degree of retention and forgetting of the historical information in the output of h_t can be decided, so that the information can be updated. The process can be represented by equation (3), where $z_t \cdot h'_t$ means that the update gate remembers the current information and $(1 - z_t) \cdot h_{t-1}$ means that the update gate forgets the historical information.

$$h_t = z_t \cdot h'_t + (1 - z_t) \cdot h_{t-1}. \quad (5)$$

Since the hidden layer of the GRU model can only be passed in one direction, grammar error correction is the process of linking the preceding and following text information to determine whether the vocabulary is correct at that position. Therefore, a hidden layer in the opposite direction is added to the GRU structure to form a Bidirectional Gated Recurrent Unit (Bi-GRU); i.e., the hidden layers learn the preceding and following text information respectively [28]. The structure is shown in Figure 4.

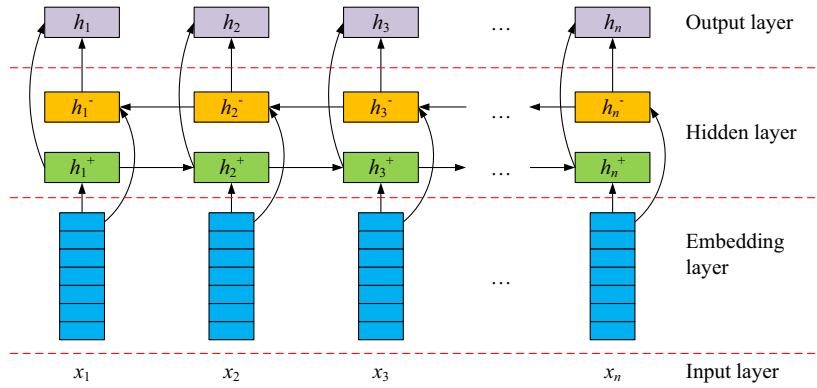


Figure 4: Diagram of Bi-GRU bi-directional operation.

In Figure 4, h^+ is the forward output, and h^- is the reverse output; then, the output of the hidden layer at the moment of t is a cascade of forward and reverse, which can be represented by equation (4).

$$h_t = h_t^+ \oplus h_t^-. \quad (6)$$

If Bi-GRU has m layers, the final states of the forward and reverse hidden layers are h_m^+ and h_1^- , and the final state of the Bi-GRU hidden layer h_{final} is $h_m^+ \oplus h_1^-$. However, the output of the Bi-GRU is given the same meaning at this point, which will affect the operation of the decoding phase. Therefore, attention weight calculation is inserted after the Bi-GRU model to clarify the focus of each semantic based on the weights. Taking a_m to denote the attention probability distribution of the hidden layer state to the final state of the Bi-GRU hidden layer at a certain moment, then a_m can be calculated by equation (5).

$$\begin{cases} a_m = \frac{\exp(h'_m)}{\sum_{i=0}^N \exp(h'_i)}, \\ h'_m = h_m^T U h_{\text{final}}. \end{cases} \quad (7)$$

where U and N in equation (5) represent the weight matrix and the number of tokens of the input sentences. Finally, the Bi-GRU hidden layer state is normalized by equation (6) to obtain the final encoding result.

$$\begin{cases} h_{t-\text{enc}} = f \left[\frac{g}{\sigma_t} \odot \left(a_t - \frac{\sum_{i=1}^H a_t^i}{H} \right) + b \right] \\ a_t = W_{\text{hh}} h_{t-1} + W_{\text{xh}} x_t \\ \sigma_t = \sqrt{\frac{1}{H} \sum_{i=1}^H \left(a_t^i - \frac{\sum_{i=1}^H a_t^i}{H} \right)^2}. \end{cases} \quad (8)$$

In equation (6), g , b and H , respectively, denote the gain matrix, offset matrix, and the number of cells in the hidden layer. W_{xh} represents the weight matrix between the input and the hidden layer, and W_{hh} represents the weight matrix between the hidden layers.

3.3 Decoder model architecture based on attention mechanism

The decoder can decode the result obtained by the encoder to obtain the corresponding sentence. Traditional decoders can only decode by semantic vectors and each word in the utterance has the same impact on the output at the same moment. This would result in the output at each moment not corresponding to the input content that is more associated with it, thus losing some information in the decoding stage. To change this situation, the study introduces an attention mechanism in the decoder that can focus on the words it decodes at each step of the decoding process, depending on the different components of the sentence. The algorithmic structure of the attention mechanism is shown below [29,30]. y_t denotes the decoded word of the encoding result, which is a composite state obtained from the weighted average of h_t . Then, α is the weight value obtained by the model autonomously learning [31].

Specifically, to focus the attention on the decoder's hidden state h_t at the moment t , the attention score is obtained by calculating the impact of each hidden state output by the encoder on the overall result through the attention model. Based on the weighted value of the attention score and the hidden states of the encoder output s_t , the intermediate semantic vector c_t for predicting words is obtained. The relevant calculations are shown in equation (9). W and v are the parameters obtained from the model training.

$$\left\{ \begin{array}{l} \text{Score}(s, h) = \begin{cases} s^T h & \text{dot} \\ s^T W h & \text{generral} \\ v^T \tan h(W \cdot [s; h]) & \text{concat} \end{cases} \\ c_t = \sum_{i=1}^T \alpha_{t,i} s_i \\ \alpha_{t,i} = \frac{\exp(\text{score}(s_t, h_i))}{\sum_{i=1}^T \exp(\text{score}(s_t, h_i))} \end{array} \right. \quad (9)$$

To obtain more accurate results, the study uses a masked multi-head attention mechanism to focus and capture different information comprehensively. The multi-head attention mechanism combines the query weight matrix W_q , key weight matrix W_k and value weight matrix W_v corresponding to each input as a set of attention heads and merges the information in different subspaces after obtaining them, respectively. Finally, the results of multi-head attention are output. The calculation process is shown in equation (10).

$$\left\{ \begin{array}{l} \text{Multihead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_n)W \\ \text{head}_i = \text{Attention}(QW_{i,q}, KW_{i,k}, VW_{i,v}). \end{array} \right. \quad (10)$$

In equation (10), Q , K , and V are the sets of the products of the inputs x_i and the corresponding matrices $W_{i,q}$, $KW_{i,k}$, and $VW_{i,v}$. The masked multi-headed attention mechanism, on the other hand, hides the information after the i th word and reduces the information interference after the position i by the casual mask based on the multi-headed attention mechanism. The diagram of the casual masking process is shown in Figure 5, where the green part is the information that can be read by the decoder and the red part is the information that cannot be read.

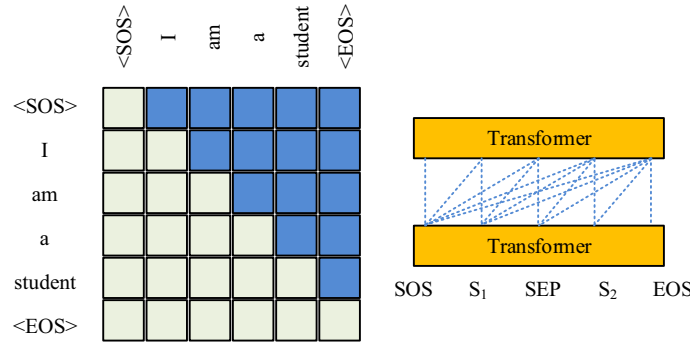


Figure 5: Casual masking process.

According to the previous Transformer model structure, it is known that a gating structure is set up after the residual concatenation and layer normalization of the results of the output of the masked multi-head attention mechanism. This structure is used to integrate the attention weights obtained by the two encoders. Assuming that the attention weights obtained by the Bi-GRU encoder are G_t and the attention weights obtained by the contextual encoder are \hat{G}_t , the output of the gating module Y_t can be calculated from equation (11).

$$\begin{cases} Y_t = G_t + \hat{G}_t + \Phi_t \odot \hat{G}_t + G_{t-1} \\ \Phi_t = \sigma(\text{LIN}(G_t) + \text{LIN}(\hat{G}_t)). \end{cases} \quad (11)$$

In equation (11), $\Phi_t \odot \hat{G}_t$ represents the influence of the preceding and following information on the predicted value at that moment, and the symbol \odot represents the Hadamard product. The probability of outputting candidate sentences is mapped by the softmax function after a linear combination of attention weights, and the output of candidate sentences is controlled by the Beam Search algorithm and dynamic cluster search [32].

4 Performance evaluation of English grammar error correction model

To verify the effectiveness of the model, this study used three corpora, ICNALE, NUCLE, and partially CLEC, and trained the model using the CoNLL-2013 test set and JFLEG Dev set as validation sets. Then, the model was tested using the CoNLL-2014 test set, JFLEG test set, and partially untrained CLEC data set. The encoder includes a three-layer Bi GRU and a six-layer Transformer encoder, with a six-layer shielded multi-head attention layer set on the decoder side. The other related parameters are set in Table 2.

Table 2: Model parameter settings

Parameter type	Parameter settings	Parameter type	Parameter settings
Word vector dimension	256	Dropout of activation function	0.1
Model for generating word vectors	ALBERT	Ir	0.01
Dictionary size	30K	Parameter optimization algorithm	Adam
Number of Transformer encoder hidden units	512	batch_size	20
Number of Bi-GRU hidden units	256	Epoch	40
Dropout of Bi-GRU	0.2	Maximum sentence length	50
Heads of multiple attention	6	Dynamic bundle search probability threshold	0.90–0.99
Number of neurons in feedforward neural network	2,048	Length penalty parameter	0.6
Activation function	Relu	—	—

To ensure that the performance of the model is in an optimal state when tested, the experiments first screened the probability thresholds of the cluster search that affect the model greatly. The three metrics of accuracy, recall, and $F1$ score were tested on the training set, and the appropriate probability threshold was selected according to the $F1$ score shown in Figure 6.

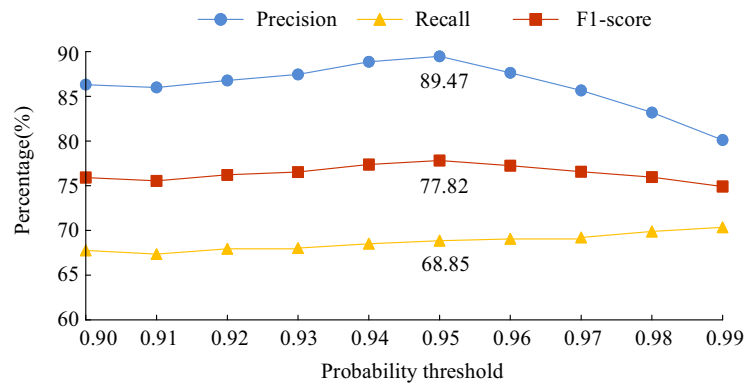


Figure 6: Evaluation index data of the model under different probability thresholds.

Figure 6 demonstrates that the model achieves its highest accuracy rate and $F1$ score of 89.47 and 77.82%, respectively, when the probability threshold is set at 0.95. This indicates optimal performance, with a high accuracy rate and more retained candidate results. Therefore, the probability threshold is established at 0.95. The results were further evaluated based on precision rate, recall rate, and $F0.5$ score using the CoNLL-2014 Test Set test, and compared to previously designed models such as Nested-GRU, MLConv (4 ens.)-EO, and SMT-NMT. Figure 7 illustrates the comparison results.

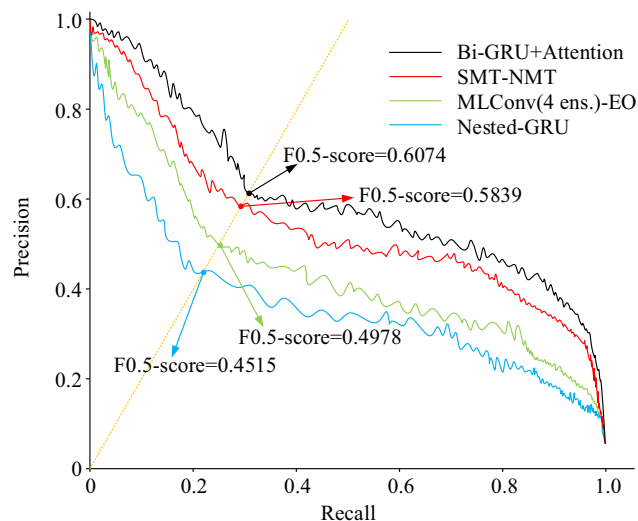


Figure 7: Plot of $F0.5$ scores of the model in CoNLL-2014 Test Set.

The test results showed that the Bi-GRU + Attention model proposed in this study had an accuracy of 71.28%, a recall of 38.16%, and an $F0.5$ of 60.74%. The $F0.5$ of the Nested-GRU, MLConv(4 ens.)-EO, and SMT-NMT models scores were 45.15, 49.78, and 58.39%, respectively. Comparing the curves in Figure 7 with the experimental data, the Bi-GRU + Attention model has made greater progress in grammar error correction. Among them, Nested-GRU also used an additional non-public corpus for training, and its $F0.5$ score was still six

percentage points lower. The model also outperformed the more advanced SMT-NMT hybrid model by 2.35 percentage points in the F0.5 score compared to the more advanced SMT-NMT hybrid model. This indicates that the Bi-GRU + Attention model is effective in automatically correcting English grammatical errors. In addition, to evaluate adequacy and fluency, the experiment tested the GLUE value of the model on the JFLEG Test Set, which is a manual measurement used to check the degree of matching between machine error correction results and translation results of professional English users. Figure 8 shows the GLUE scores, and the GLUE value of Human performance is introduced here as a reference.

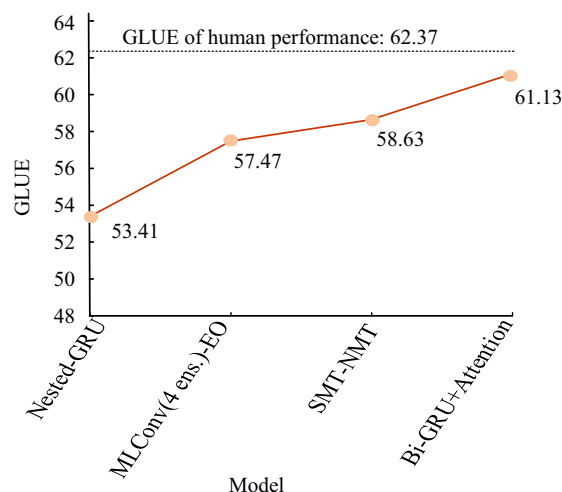


Figure 8: Comparison of GLUE values of models in JFLEG test set.

Comparing the GLUE values of Bi-GRU + Attention, Nested-GRU, MLConv (4 ens.)-EO, and SMT-NMT models, the Bi-GRU + Attention model had the highest GLUE score of 61.13. It was closer to the human assessment score of 62.37, which indicated that the model had better adequacy and fluency and was better for the application of English grammatical error correction. The analysis may be due to the introduction of a Bi-GRU encoder and Attention decoder on top of the Transformer model with powerful feature extraction ability, which enables the machine to extract a wider range of semantic information, thus enhancing the fluency of the utterance. The utility of the model was also explored in this study. The model was randomly selected from the untrained CLEC corpus with 1,000 essay tests of non-English majors under five headings for the fourth- and sixth-grade exams and segmented by 100, 200, 500, 800, and 1,000, and the error correction accuracy rate, recall rate, and F1 score at each stage are shown in Figure 9.

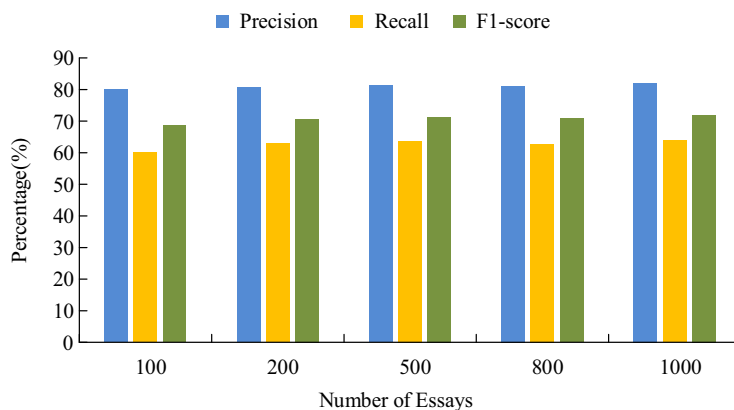


Figure 9: Model measurement at different number of articles.

In Figure 9, although there are small fluctuations in the accuracy rate, recall rate, and *F1* score for grammatical error correction for different numbers of essays, the results do not differ much, indicating that the model error correction is less influenced by the number of texts and has better stability. Even in the case of one hundred texts, the model can maintain an 80% accuracy rate, probably because the model adopted a fluency-based data augmentation method and ALBERT-based dynamic word vector representation in the preprocessing stage, which made the model obtain more comprehensive information with less computational resources. In terms of the test results of 1,000 essays, the accuracy, recall, and *F1* scores were 82.13, 63.85, and 71.85%, respectively, which were at a high level overall, indicating that the model had good practicality and effectively criticized students' essays. The statistics of the above types of grammatical errors in English composition according to the CoNLL-2014 classification criteria are shown in Figure 10. The number of errors marked, detected, and corrected was represented by bar graphs, and the accuracy rate, recall rate, and *F1* scores detected by the model are represented by line graphs.

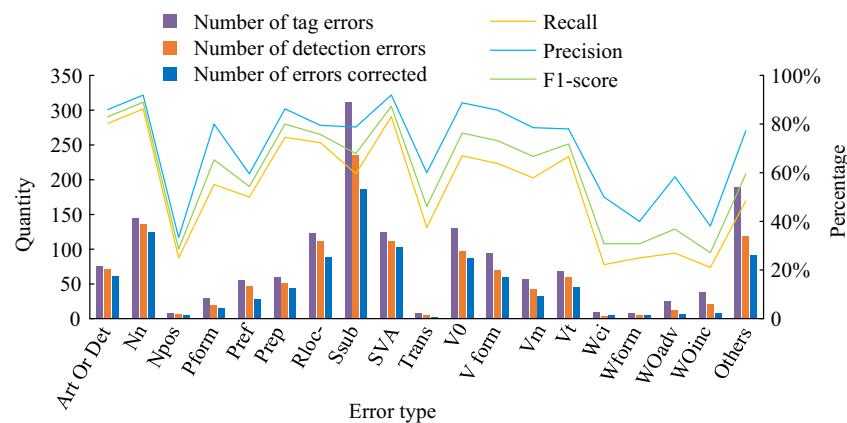


Figure 10: Results of model identification and correction of different grammatical errors.

In Figure 10, the model has a good correction effect for most grammatical error types, especially in correcting errors in the use of coronals and qualifiers, errors in the singular and plural forms of nouns, errors in the form of verbs, errors in the use of prepositions, subject–predicate inconsistency, and missing verbs, with accuracy rates above 80%. The correction effect for the six error types of *Npos*, *Trans*, *Wci*, *Wform*, *WOadv*, and *WOinc* was less effective, and their *F1* scores did not exceed 50%. Combined with the corresponding number of grammatical errors in Figure 9, it is speculated that the reason may be that the number of samples for these six error types is small, and the relevant information available for model training is reduced, which reduces the correction effect. Based on the above results, the model was used to score English compositions, and the results were compared with manual scoring to evaluate its practical application. The results are shown in Figure 11, and the model is set to score an upper limit of thirty.

The comparison between the model score and the manual score showed a high degree of overlap between the two scores, and the scores were very close. The data also showed that the average score was 26.14 points, and the average score for manual work was 25.66 points, with an error of only 1.9%. This indicated that the model could replace teachers in correcting English compositions within a certain range and has practical application value.

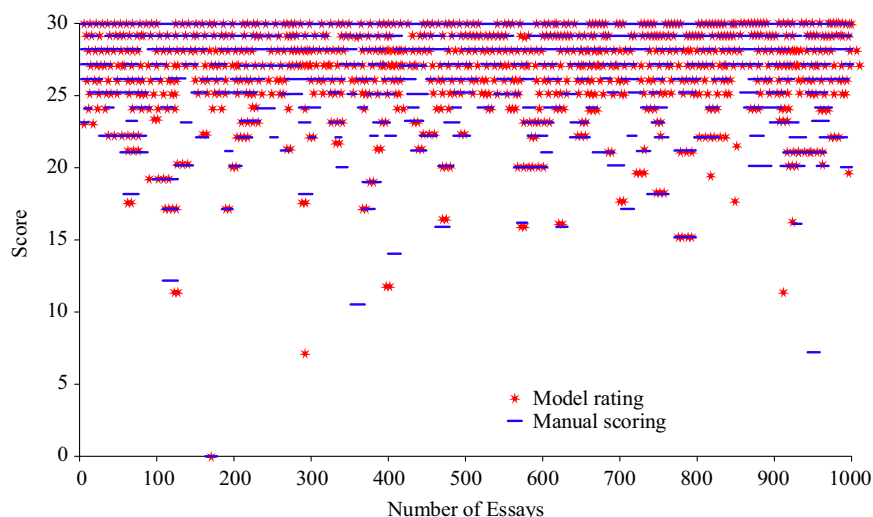


Figure 11: Comparison of model scoring and manual scoring.

5 Conclusion

In recent years, with the wide use of English in the world, the mastery of English grammar is a challenge for learners of other languages, which makes the demand for intelligent error correction models of English grammar growing day by day. However, the traditional English grammar correction model is often unable to deal with complex grammar errors. To improve the accuracy and efficiency of English grammar correction, an English grammar intelligent error correction model based on mixed attention mechanism and RNN algorithm is proposed. The results showed that the F0.5 score of this model on the CoNLL-2014 test set was 60.74%, which was 2.35 percentage points higher than that of the more advanced SMT-NMT model, showing the advantage of this model in dealing with English grammar errors. On the JFLEG test set, the GLUE score of the model reached 61.13, which was only 1.24 points different from the human evaluation, which proved the ability of the model to process real text. In the CET-4 and CET-6 compositions for non-English majors, the *F1* score was as high as 71.85%, and the average English composition score of the model was only 0.48 points different from that of the manual score, which verified the validity and accuracy of the model. The model has achieved remarkable results in improving the accuracy and efficiency of English grammar correction. However, there may still be some limitations in the processing of advanced and complex grammatical structures. In the future, it is still necessary to improve the processing of complex grammatical structures, further improve the accuracy and efficiency of English grammar correction, and provide English learners with higher-quality learning tools.

Funding information: The research is supported by Authorized by Hunan Province Office of Education, Research on Teaching Innovation in Railway ESP from the Perspective of the “the Belt and Road” Initiative (NO. ZJGY2021018).

Author contributions: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Shan Chen and Yingmei Xiao. The first draft of the manuscript was written by Shan Chen and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conflict of interest: The authors report there are no competing interests to declare.

Data availability statement: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

References

- [1] Zhong Y, Yue X. On the correction of errors in English grammar by deep learning. *J Intell Syst.* 2022;31(1):260–70.
- [2] Qin M. A study on automatic correction of English grammar errors based on deep learning. *J Intell Syst.* 2022;31(1):672–80.
- [3] Nava E, Hesham F. Grammatical and context-sensitive error correction using a statistical machine translation framework. *Softw: Pract Exper.* 2012;43(2):187–206.
- [4] Zhao Z, Wang H. MaskGEC: Improving neural grammatical error correction via dynamic masking. *Proceedings of the AAAI Conference on Artificial Intelligence.* Vol. 34, Issue 1; 2020. p. 1226–33.
- [5] Zhang F, He X, Teng Q, Wu X, Cui J, Dong X. PM-ARNN: 2D-TO-3D reconstruction paradigm for microstructure of porous media via adversarial recurrent neural network. *Knowl Syst.* 2023;264(15):2–16.
- [6] Wu YH, Li HB. RNNCTPs: A neural symbolic reasoning method using dynamic knowledge partitioning technology. *Knowl Syst.* 2023;268(23):2–9.
- [7] Shahkarami A, Yousefi M, Jaouen Y. Complexity reduction over Bi-RNN-based nonlinearity mitigation in dual-pol fiber-optic communications via a CRNN-based approach. *Optical Fiber Technol.* 2022;74(10):2–12.
- [8] Zhang Y, She H, Du YP. Dynamic MRI of the abdomen using parallel non-Cartesian convolutional recurrent neural networks. *Magn Reson Med.* 2021;86(2):964–73.
- [9] Gao M, Wang F, Liu J, Song P, Chen J, Yang H, et al. Estimation of the convolutional neural network with attention mechanism and transfer learning on wood knot defect classification. *J Appl Phys.* 2022;131(23):2–10.
- [10] Ranalli J, Yamashita T. Automated written corrective feedback: Error-correction performance and timing of delivery. *Lang Learn Technol.* 2022;26(1):1–25.
- [11] Trappey AJC, Chang AC, Trappey CV, Chien JCC. Intelligent RFQ summarization using natural language processing, text mining, and machine learning techniques. *J Glob Inf Manag.* 2022;30(7):3193–218.
- [12] Makwana AP, Dhont K, Sancho EG, Berrocal PF. Are emotionally intelligent people less prejudiced The importance of emotion management skills for outgroup attitudes. *J Appl Soc Psychol.* 2021;51(6):98–127.
- [13] Zhou S, Liu W. English grammar error correction algorithm based on classification model. *Complexity.* 2021;21(2):2–11.
- [14] Solyman A, Wang Z, Tao Q, Rui Z, Zeinab M, Mohammed EAA. Automatic Arabic Grammatical Error Correction based on Expectation-Maximization routing and target-bidirectional agreement. *Knowl Syst.* 2022;241(6):2–13.
- [15] Zhan W, Chen Y. Application of machine learning and image target recognition in English learning task. *J Intell Fuzzy Syst.* 2020;39(4):5499–510.
- [16] Wang Y, Xu C, Hu H, Tao C, Wan S, Dras M, et al. Neural rule-execution tracking machine for transformer-based text generation. *Adv Neural Inf Process Syst.* 2021;34:16938–50.
- [17] Xu W, Carpuat M. EDITOR: An edit-based transformer with repositioning for neural machine translation with soft lexical constraints. *Trans Assoc Comput Ling.* 2021;9:311–28.
- [18] Duran-Karaoz Z, Tavakoli P. Predicting L2 fluency from L1 fluency behavior: The case of L1 Turkish and L2 English speakers. *Stud Second Lang Acquis.* 2020;42(4):671–95.
- [19] Firat O, Cho K, Sankaran B, Vural YFT, Bengio Y. Multi-way, multilingual neural machine translation. *Comput Speech Lang.* 2017;45:236–52.
- [20] Ning K, Cai M, Xie D, Wu F. An attentive sequence to sequence translator for localizing video clips by natural language. *IEEE Trans Multimed.* 2020;22(9):2434–43.
- [21] Wang D, Su J, Yu H. Feature extraction and analysis of natural language processing for deep learning English language. *IEEE Access.* 2020;8:46335–45.
- [22] Mohammad EB, Shahla N, Moloud A, Erik C, Rajendra AU. ABCDM: An Attention-based bidirectional CNN-RNN deep model for sentiment analysis. *Future Gener Comput Syst.* 2021;115:279–94.
- [23] Zhang B, Xiong D, Xie J, Su J. Neural machine translation with GRU-gated attention model. *IEEE Trans Neural Netw Learn Syst.* 2020;31(11):4688–98.
- [24] Nicolson A, Paliwal KK. Masked multi-head self-attention for causal speech enhancement. *Speech Commun.* 2020;125:80–96.
- [25] Aguiar-Pérez JM, Pérez-Juárez MÁ. An insight of deep learning based demand forecasting in smart grids. *Sensors.* 2023;23(3):1467.
- [26] Mclellan G. Practitioners respond to Icy Lee's 'Teacher written corrective feedback: Less is more'. *Lang Teach.* 2021;54(1):144–8.
- [27] Jin L, Schwartz LO, Velez FD, Miller T. Depth-bounded statistical PCFG induction as a model of human grammar acquisition. *Comput Linguist.* 2021;47(1):181–216.
- [28] Yang S, Kong X, Wang Q, Li Z, Cheng H, Xu K. Deep multiple auto-encoder with attention mechanism network: A dynamic domain adaptation method for rotary machine fault diagnosis under different working conditions. *Knowl Syst.* 2022;249(5):2–17.

- [29] Niu D, Yu M, Sun L, Gao T, Wang K. Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism. *Appl Energy*. 2022;313(1):2–17.
- [30] Hui T, Xu YL, Jarhinbek R. Detail texture detection based on Yolov4-tiny combined with attention mechanism and bicubic interpolation. *IET Image Process*. 2021;15(12):2736–48.
- [31] Sagnika S, Mishra BSP, Meher SK. An attention-based CNN-LSTM model for subjectivity detection in opinion-mining. *Neural Computing and Applications*. 2021;33(24):17425–38.
- [32] He Z. English grammar error detection using recurrent neural networks. *Sci Program*. 2021;21(5):2–8.