A Study on Error Feature Analysis and Error Correction in English Translation Through Machine Translation

Guifang Tao

Wuhan Vocational College of Communications and Publishing, Wuhan, Hubei 430223, China

E-mail: gui6095@163.com

English translation is the most frequently encountered problem in English learning, and fast, efficient and correct English translation has become the demand of many people. This paper studied the most frequently encountered English grammatical error problem in English translation by the Transformer grammatical error correction model in machine translation and explored whether machine translation can analyze the features of the errors that may occur in English translation and correct them. The results of the study showed that the precision of the Transformer model reached 93.64%, the recall rate reached 94.01%, the $F_{0.5}$ value was 2.35, and the value of the Bilingual Evaluation Understudy was 0.94, which were better than those of the other three models. The Transformer model also showed stronger error correction performance than Seq2seq, convolutional neural network, and recurrent neural network models in analyzing error correction instances of English translation. This paper proves that it is feasible and practical to identify and correct English translation errors by machine translation based on the Transformer model.

Povzetek: Jezikovni pretvornik je bil uporabljen za odkrivanje napak v prevodu v angleščino. Rezultati so pokazali visoko točnost v primerjavi z drugimi pristopi.

1   Introduction

English is a universal language, and countless people are learning English; however, many people are not native English speakers and are prone to English translation errors. To reduce the occurrence of such translation errors, people use computer programs for English translation, i.e., machine translation. With the advent of big data and deep learning technology [1], machine translation technology has been optimized and the quality of translation has been improved, and machine translation has become an important aid to human translation. This paper studied the most frequently encountered English grammatical error problem in English translation by the Transformer-based grammatical error correction model and explored whether machine translation can analyze the features of errors in English translation and correct the errors. The experimental results of various models, such as the Seq2seq model, the convolutional neural network (CNN) model, and the RNN model, were compared by evaluation indicators, thereby ensuring that the Transformer-based grammatical error correction model was of application value.

2   Related works

Table 1: Summary of relevant work

	Model	Results	Limitations
Satir et al. [2]	A hybrid system by guiding NMT	The proposed method can improve	Applications in different languages
	decoding using the output sentences of the phrase-based SMT systems	translation quality.	
Hnamte et al. [3]	Neural machine translation	Achieve 42.65 BLEU score on the 4 grams.	Dependence on corpus during model training
Singh et al. [4]	A semi-supervised neural machine translation system	The proposed semi-supervised system outperforms the supervised, the pretrained mBART and existing semi-supervised baselines in terms of automatic score and subjective evaluation parameters by a significant	The proposed semi-supervised approach is robust to handle rare words and long-term dependencies as evident from the error analysis based on word translation

		margin up to +4.5 and +1.2 BLEU improvements against the supervised and mBART baselines respectively.	accuracy and BLEU scores grouped by the sentence length; an imbalanced combination of the synthetic data is found to deteriorate the overall performance.
Laskar et al. [5]	An Assamese pre-trained language model	With the use of both prior alignment and a pre-trained language model, the transformer-based neural machine translation model shows improvement, and we have achieved state-of-the-art results for the English-to-Assamese and Assamese-to-English translation, respectively.	The multilingual transfer learning-based approach for further research
Loubser et al. [6]	The neural network model for core language technologies	The neural model performs comparably with the baseline on Afrikaans and disjunctive languages (accuracy within 1%), and slightly worse on conjunctive languages, falling short of the baseline by 2.3% on average. Neural networks can	The experiments in this paper evaluated only one neural architecture for each task.
		be viable implementations of core language technologies for resource-scarce South African languages.	

Through the analysis of the current related research, we can find that in the research of machine translation, how to further improve the quality of translation and expand the application of machine translation in different languages is still a focus of current research. Due to the low quality of machine translation, many translation errors often occur, and this problem can be effectively solved by improving the quality of translation. Therefore, this paper investigates the problem of error elimination and correction in English translation.

3 English translation error correction method

3.1 Analysis of English translation error characteristics

English is currently the most widely spoken language in the world [7], and the most common difficulty in the process of learning English is English translation. Since languages are not the same from country to country, the influence of factors such as Chinglish and insufficient vocabulary leads to frequent translation errors. The analysis of the English translation error characteristics of the data set studied in this paper reveals that most of the errors in the data set are grammatical translation errors, and grammatical errors can be divided into two categories: lexical errors and syntactic errors. Lexical errors mainly include lexical errors and singular and plural mistranslation, while syntactic errors mainly include subject mistranslation, confusion of logical relations, and verb tense errors [8]. For example, in terms of the vocabulary error, "红茶" is translated as "red tea", but the actual correct translation is "black tea"; in terms of the singular and plural mistranslation, "她有 5 天的带薪年假" is translated as "She has five day of paid annual leave", but the actual correct translation is "five days"; in terms of the grammatical error, "我们的土地已经得到很好的开发" is translated as "Our land has well developed", but the actual correct translation is "has been"; in terms of the verb tense error, "他说他第二天要去爬山" is translated as "He said he will go climbing the next day", but the actual correct translation is "would".

3.2 Transformer model

In this paper, the Transformer model is used as an error corrector for machine translation. A transformer error correction model is a neural machine translation model [9] with efficient model training efficiency, which can

automatically correct a large number of recognition errors, especially the substitution errors in recognition results. The model is mainly composed of two parts: the encoder and the decoder. The encoder consists of a self-attention layer and a feedforward neural network layer. The self-attention layer contains multi-headed attention, summation, and normalization, and the feedforward neural network layer contains feedforward neural network, summation, and normalization. The decoder consists of the self-attention layer, the encoder-decoder attention layer, and the feedforward neural network layer, which also contains summation and normalization. The equation for the self-attention layer is:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_K}}\right)V , \quad (1)$$

where Q is the query vector, K is the key vector, V is the value vector, $\sqrt{d}$ is a fixed factor, and d is the dimension of the hidden layer. The Transformer model splits the self-attention mechanism into (Q, K, V) and applies multi-headed attention in the self-attention layer, where Q query vector, K key vector, and V value vector are all from the output of the previous sub-layer.

The Transformer model for error correction works by inputting the source error sentence, outputting the corresponding feature vectors after encoder processing, inputting the feature vectors to the decoder for re-processing and outputting the target corrected sentence. In the training process of the Transformer-based error correction model, since the model does not contain convolutional network and recurrent network, the sequential order of words in the data cannot be obtained, and the model needs to be embedded with position encoding features; the position information of words is added to the embedded vector to allow the model to discriminate words in different positions. The model position coding used in this paper is sine/cosine position coding [10], and the specific formulas are given in Equations (2) and (3):

$$PE_{(pos,2i)} = sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right), \quad (2)$$
$$PE_{(pos,2i+1)} = cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right), \quad (3)$$

where PE is the abbreviation for position encoding, pos is the specific position of a word in a sentence, i is the component of every value in the position encoding, and d_{model} is the uniform dimension of the input and output of different layers in the model. When the word is in an even position, the sine coding is used, and when it is in an odd position, the cosine coding is used.

4   Experimental analysis

4.1   Data collection and processing

The dataset for this study comes from the British National Corpus (BNC), Lang-8, the Tsinghua University Chinese-English Parallel Corpus (THUMT), and the CoNLL-2014 test set. Five million pieces of data were used as the initial data for this study. Three million pieces of data were used as the training set for continuous training and adjustment of the model, and the remaining two million pieces were used as the test set to test the final machine translation results of the model. The data specifications are listed in Table 1.

Table 2: Experimental data sources and quantities

Data distribution	Corpus name	Number of data (piece)
Training set data	British National Corpus (BNC)	1.5 million
	Lang-8	1.5 million
Test set data	Tsinghua University Chinese-English Parallel Corpus (THUMT)	1 million
	CoNLL-2014 test set [11]	1 million

After collecting the data from the experiments, they were processed to meet the input requirements of the model. The data processing methods used in this paper are as follows. Firstly, the data with a large number of duplicates in the dataset were deleted to avoid the deviation of machine translation results caused by duplicate data in the process of model training. Secondly, the length of the data in the dataset was unified to avoid the step of splitting the long data. The maximum length of the data set was set as 100, and the part exceeding this length was directly truncated. Thirdly, special symbols, such as ¥ and m^2, were processed. The data in the dataset will inevitably have some special symbols. Since the model could not translate the special symbols, the data containing special symbols were deleted. Fourthly, the initial data was transformed into grammatically incorrect data by means of noise perturbation. The experiment aimed to prove that the machine translation based on the Transformer model could correct wrong English translation, so data containing a large number of grammatical errors were needed for the experimental operation. The sentences in the dataset were treated by Byte Pair Encoding (BPE) [12], and then the words were deleted, replaced, or inserted by other words with random probability, thus generating data containing a large amount of grammatical errors for the model analysis.

4.2   Experimental design

The Transformer-based error correction model was used as the training model. In this model, both the encoder and decoder had six layers, there were eight multi-headed attention layers, the dimension of the hidden layer of the feedforward network was 2,048, and the dimension of the word embedding was 512. Before experiment, enough data were collected for the model analysis. Then, the data set was divided and preprocessed to meet the training requirements of the Transformer-based-training and fine-

tuned to reach the optimum by inputting the training set data. After the optimal model was obtained, the test set data were input to the model for machine translation. Two evaluation indicators, M^2 and Bilingual Evaluation Understudy (BLEU), were used to evaluate the translation results to verify the effectiveness of the model. Moreover, the translation results were compared with the machine translation results of Seq2seq [13], CNN [14], and RNN models [15] to ensure that the Transformer-based error correction model is of practical value in machine translation self-correction. In addition, the initial parameters of different models were set uniformly to ensure the reliability and validity of the experimental results. The Adam optimization function was used during model training, the learning rate was set as 0.001, the dropout was set as 0.5, the training batch size was 100, and the number of iterations was set as 40.

4.3 Evaluation indicators

4.3.1 The maximum matching score (M^2)

The maximum matching score (M^2) [16] is one of the most commonly used methods to assess English grammar error correction models. The first evaluation index used in this paper is the maximum matching score (M^2) in which precision (P), recall rate (R), and $F_{0.5}$ value as the main evaluation indexes. Their calculation formulae are:

$$P = \frac{\sum_{i=1}^{n} |A_i \cap B_i|}{\sum_{i=1}^{n} |A_i|}, \quad (4)$$
$$R = \frac{\sum_{i=1}^{n} |A_i \cap B_i|}{\sum_{i=1}^{n} |B_i|}, \quad (5)$$

where A_i is the set of corrective edits output by the model and B_i is the set of corrective edits for manual annotation.

The reason for choosing $F_{0.5}$ as the evaluation indicator is because the accuracy of machine translation is valued more than the number of translations in the error correction model, so the weight of accuracy is set at twice the recall rate. The value of $F_{0.5}$ is a combination of precision and recall rate. When the $F_{0.5}$ value is higher, the better the translation output of the model is, and vice versa. The calculation formula is:

$$F_{0.5} = \frac{(1+0.5^2)*precision*recall\ rate}{0.5^2*precision+recall\ rate}. \quad (6)$$

4.3.2 BLEU value

The second evaluation indicator used was BLEU [17], which is very common in the evaluation metrics of machine translation, and it was used to evaluate the difference values in the model-generated machine translation text and the actual correct text. Its value was between 0 and 1. If two texts matched perfectly, then the value of BLEU was 1; otherwise, the value of BLEU was 0.

4.4 Analysis of results

As seen in Figure 1 above, the word error rate decreased as the number of hidden layers increased, but too many hidden layers resulted in a large number of neurons, which increased the computational load of the model. In Figure 1, the word error rate of the four different models became the lowest when the number of hidden layers of the model was 6; however, the word error rate tended to increase after that, even though the number of hidden layers increased. Therefore, the number of hidden layers was set at 6 to avoid the increase in computation and decrease in accuracy of the model due to the excessive number of hidden layers.
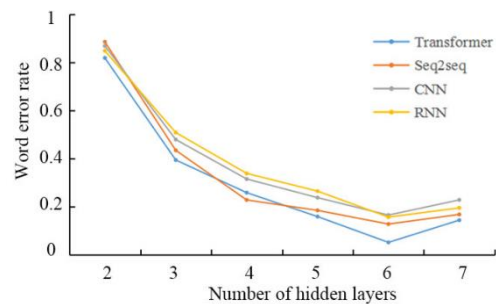

Figure 1: The number of batches and BLEU variation of different models.

Table 3: Comparison of error-corrected translation results of different models with correct translations in some cases.

Original sentence	这个村庄在山腰 The village on side of the mountain
Transformer model	The village is **on the side of** the mountain
Seq2seqmodel	The village is **by the side of** the mountain
CNN model	The village is **on the hillside**
RNN model	The village is **halfway up** the hill
Correct translation	The village is **on the side of** the mountain
Original sentence	报纸上都是最新的新闻 The newspapers are full of the newest new
Transformer model	The newspapers are full of the **latest** news
Seq2seqmodel	The newspapers are full of the **newest** news
CNN model	The newspapers are full of the **latest** news
RNN model	The newspapers are full of the **newest information**
Correct translation	The newspapers are full of the **latest** news
Original sentence	只有我们互相包容的时候，才能更好的合作

Transformer model	Only when we toleranting of each other we can have better cooperate
Seq2seqmodel	**Only** when we are tolerant of each other can we cooperate better
CNN model	When we are tolerant of each other
RNN model	**then** we can cooperate better
Correct translation	When we are tolerant of each other**,** we can cooperate better
	When we are tolerant of each other**,** we can cooperate better
	Only when we are tolerant of each other can we cooperate better

Through the above three case studies of error-corrected parallel sentences, it was seen that the correction results of the Transformer model were the same as the correct English translations for reference, and its error correction result was better than the other three correction models. In Case 1, the correct translation focuses on "on the side of the mountain", and the overall meaning of the translations of Transformer and RNN models was consistent with it, but the Seq2seq and CNN models were not. In Case 2, the correct translation highlights "the latest", but "new" does not mean the latest. In Case 3, the correct translation highlights "only". Although the results of the error-corrected translation by the Seq2seq, CNN, and RNN models did not differ much from the overall meaning of the correct sentence, they did not translate the word "only".

Table 4: Comparison of experimental results between different error correction models.

	Secondary evaluation indexes of the confusion matrix			BLEU
	Precision	Recall rate	$F_{0.5}$	
Transformer model	93.64%	94.01%	2.35	0.94
Seq2seq model	87.26%	85.39%	2.15	0.86
CNN model	83.55%	83.61%	2.09	0.81
RNN model	84.37%	83.94%	1.99	0.82

It was observed in Table 3 that the Transformer model had a precision of 93.64%, a recall rate of 94.01%, and an $F_{0.5}$ value of 2.35; they were much higher than the precision, recall rate, and $F_{0.5}$ value of the Seq2seq, CNN, and RNN models. In terms of BLEU, the BLEU value of the Transformer model was 0.94, which was very close to 1. This indicated that the Transformer model had a very good performance in error correction for machine translation of English. Although the BLEU values of the Seq2seq, CNN, and RNN models were 0.86, 0.81, and 0.82, respectively, suggesting good translation performance, there was still a distance with 1. Therefore, it was concluded that the Transformer model was better than the Seq2seq, CNN, and RNN models in terms of error correction for machine translation of English.

5   Discussion

The translation error in the process of machine translation directly affects the quality of translation and the reliability of the machine translation model, so how to further improve the quality of machine translation has become an urgent problem to solve. As English is a widely used language, the study of its translation errors is of great practical value, so this paper studied the characteristics of errors in English translation and the error correction.

The comparison between the Transformer model and other models showed that the Transformer model had a lower error rate in the English translation task, and the translation results obtained were closer to the correct translation, i.e. the translation quality was higher. From the analysis in Table 2, it was found that the Transformer model handled translation details better and had fewer translation errors compared to models such as CNN and RNN. Then, from the comparison in Table 3, it can be seen that the Transformer model outperformed the Seq2seq model, CNN model, and RNN model in terms of precision, recall rate, and $F_{0.5}$ value, with precision and recall rate above 90%. The BLEU of the Transformer model was 0.94, which was 9.3%, 16.04%, and 14.63% higher than Seq2seq, CNN, and RNN models, respectively, further proving the reliability of Transformer model in English translation error correction.

The results of the study prove the effectiveness of the Transformer model in error feature analysis and error correction of English translation, which further improves the quality of English translation based on the current study, but there are also some limitations. The research in this paper is based on English translation, and the applicability of the Transformer model for error correction in other languages is unclear. Although some results have been achieved in improving BLEU scores compared to other current methods, further validation on larger datasets is still needed, and these are issues that need to be addressed in future work.

6   Conclusion

This paper briefly introduced the characteristics of English translation errors and the Transformer model, and analyzed whether the Transformer model can analyze error features and correct errors by machine translation. A sufficient amount of data was collected and processed before the experiment, the training set data was input into the Transformer model for training, the test set data was input after the model was continuously adjusted to reach the optimal model, and the final machine translation results were obtained. Finally, the experimental results were evaluated using two major evaluation indexes, namely, M^2 and BLEU. The results showed that the machine translation precision of the Transformer model reached 93.64%, the recall rate reached 94.01%, the $F_{0.5}$ value reached 2.35, and the BLEU value reached 0.94. The analysis of the correction cases of some English

mistranslations showed that the Transformer model was more effective than the other three models in the error correction of English translation. It proves that machine translation based on the Transformer model can be used to identify and correct English translation errors in the future.

References

[1] Zhang B, Xiong D, Su J (2020) Neural Machine Translation with Deep Attention. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42, pp. 154-163.
https://doi.org/10.1109/TPAMI.2018.2876404

[2] Satir E, Bulut H (2021) A Novel Hybrid Approach to Improve Neural Machine Translation Decoding using Phrase-Based Statistical Machine Translation. *2021 International Conference on Innovations in Intelligent Systems and Applications (INISTA).* https://doi.org/10.1109/INISTA52262.2021.954840 1

[3] Hnamte V, Thangkhanhau H, Hussain J, Lalnunmawii C, Tlaisun L, Vanlalruata (2022) Mizo to English Machine Translation: An Evaluation Benchmark. *2022 International Conference on Futuristic Technologies (INCOFT)*, pp. 1-6, https://doi.org/10.1109/INCOFT55651.2022.10094 376

[4] Singh SM, Singh TD. (2022) Low resource machine translation of English-Manipuri: A semi-supervised approach. *Expert Systems with Applications*, 209, pp. 118187, https://doi.org/10.1016/j.eswa.2022.118187

[5] Laskar SR, Paul B, Dadure P, Manna R, Pakray P, Bandyopadhyay S (2023) English-Assamese neural machine translation using prior alignment and pre-trained language model. *Computer Speech & Language*, 82, pp. 101524. https://doi.org/10.1016/j.csl.2023.101524

[6] Loubser M, Puttkammer MJ (2020) Viability of Neural Networks for Core Technologies for Resource-Scarce Languages. *Information (Switzerland)*, 11, pp. 41. https://doi.org/10.3390/info11010041

[7] Xu C, Li Q (2021) Machine Translation and Computer Aided English Translation. *Journal of Physics: Conference Series*, 1881, pp. 1-8. https://doi.org/10.1088/1742-6596/1881/4/042023

[8] Huyen N T (2020) Common grammatical errors in English writing - A case study with second-year students of information technology at HAUI. *Can Tho University Journal of Science*, 11, pp. 37. https://doi.org/10.22144/ctu.jen.2020.005

[9] Nguyen T, Nguyen L, Tran P, Nguyen H (2021) Improving Transformer-Based Neural Machine Translation with Prior Alignments. *Complexity*, 2021, pp. 1-10. https://doi.org/10.1155/2021/5515407

[10] Ling G, Yang X (2022) Knowledge Base Question Answering Based on Multi-head Attention Mechanism and Relative Position Coding. *Journal of Physics: Conference Series*, 2203, pp. 1-7.

[11] Guo D M (2020) Jointly Part-of-Speech Tagging and Semantic Role Labeling Using Auxiliary Deep Neural Network Model. *Computers, Materials & Continua*, 2020, pp. 529-541. https://doi.org/10.32604/cmc.2020.011139

[12] Amalia A, Sitompul O S, Mantoro T, Nababan E B (2021) Morpheme Embedding for Bahasa Indonesia Using Modified Byte Pair Encoding. *IEEE Access*, 9, pp. 155699-155710. https://doi.org/10.1109/ACCESS.2021.3128439

[13] Li Y, Li J, Zhang M (2021) Deep Transformer modeling via grouping skip connection for neural machine translation. *Knowledge-based Systems*, 234, pp. 1-12. https://doi.org/10.1016/j.knosys.2021.107556

[14] Ren Q, Su Y, Wu N (2020) Research on Mongolian-Chinese machine translation based on the end-to-end neural network. *International Journal of Wavelets Multiresolution & Information Processing*, 18, pp. 46-59. https://doi.org/10.1142/S0219691319410030

[15] Datta D, Evangeline P, Mittal D, Jain A (2020) Neural Machine Translation using Recurrent Neural Network. *International Journal of Engineering and Advanced Technology*, 9, pp. 1395-1400. https://doi.org/10.35940/ijeat.D7637.049420

[16] Qi L, Guo FY, Zhang J, Wang Y W (2022) An internet review topic hierarchy mining method based on modified continuous renormalization procedure. *Fractals*, 30, pp. 1-25. https://doi.org/10.1142/S0218348X22501341

[17] Phan-Vu H H, Tran V T, Nguyen V N, Dang H V, Do P T (2019) Neural Machine Translation between Vietnamese and English: an Empirical Study. *Journal of Computer Science and Cybernetics*, 35, pp. 147-166. https://doi.org/10.15625/1813-9663/35/2/13233