International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-15

https://doi.org/10.70917/ijcisim-2026-0399
Article

Construction of English Language and Literature Translation Quality Assessment Model Based on Deep Neural Networks

Yu Zhao [1] and Yunxia Yan [2,*]

[1] Foreign Language Teaching Department, Changzhi Medical College, Changzhi, Shanxi, 046000, China
[2] Translation Department, Hebei University of Science & Technology, Shijiazhuang, Hebei, 050018, China
* Correspondence author: zhao2yu200@126.com

Abstract: English literature translation is a systematic and complex project, the study is based on the assessment of English language and literature translation quality, mBERT, XLM-R, mBART are selected for cross-language pre-training, and a translation quality estimation model based on the pre-training model of Transformer encoder-decoder structure (mBART) is constructed. The mainstream pre-trained language models are selected for performance comparison, and the mBART-based translation quality estimation model proposed in this paper achieves the best results, surpassing the baseline model under different translation tasks in multiple datasets. The average absolute error and root mean square error of the model training in this paper are less than 0.15, and the correlation coefficients in English and German translations and English and Chinese translations are higher than those of the baseline model by 0~4% and 0~13.3%, and are less affected by the number of translated sentence types and utterances, which indicates the accuracy and stability of the translation quality estimation model of English language and literature in this paper. In conclusion, the proposed model is suitable for use in the assessment of English language and literature translation quality.

Keywords: Transformer; mBART; cross-language pretraining; translation quality assessment; English language and literature

1. Introduction

With the continuous development of society, people's living standards are improving, and people's pursuit of culture is also increasing. This reflects that people begin to pay attention to the pursuit of spiritual culture, and it also means that literature may not be able to meet the needs of many readers, so the work for the introduction of English language and literature becomes indispensable [1-4]. But for the vast majority of readers directly reading English language literature is more difficult, the language is a major obstacle, if one by one to solve the obstacles in reading, it will take them a lot of time, which subverts the value of reading itself, which also reflects the importance of translating English language literature [5-8].

With the acceleration of globalization and the high cost of human translation, there is a growing demand for machine translation. Machine translation refers to the process of using computer software to automatically translate natural language, which includes the analysis and comprehension of the machine source language text and the generation of the target language [9-12]. However, the quality of machine translation has been a challenge due to the differences between languages and the complexity of semantics. In order to assess the quality of machine translation, it is important to construct a translation quality assessment model [13-15]. The traditional manual assessment is to score and evaluate the quality of machine translation by human assessors. The advantage of this method is higher accuracy, but it is costly and time-consuming [16-18]. In order to improve the accuracy of the evaluation of machine translation quality, some methods based on deep learning are widely used, and these methods mainly include deep neural network-based evaluation models and deep learning-based feature extraction methods. The advantage of deep neural network-based evaluation model is that it can learn more semantic information, thus improving the accuracy of evaluation [19-22].

With the development of globalization, the importance of machine translation has become increasingly prominent. Machine translation can help people translate texts in various languages, thus promoting various international exchanges and cooperation. However, ensuring its translation quality has become a hot issue in current research. Literature [23] aims to show the changes in the assessment of translation quality, pointing out a positive change in practice, where artificial intelligence is applied to translation assessment criteria, lifting barriers and achieving unprecedented translation standards. Literature [24] proposes the application of neural network algorithms in assessing the quality of English translation by developing a neural network model and evaluating its performance based on human assessment, revealing the effectiveness of the model, which contributes to improving the quality of translations in an academic environment. Literature [25] describes a framework for machine translation evaluation based on the use of neural networks in a two-by-two setting, aiming at selecting better translations, showing the flexibility of the framework, which allows for effective learning and scoring and provides machine translation evaluation metrics. Literature [26], as an example of refining the quality estimation of machine translation, proposes neural network features, noting that neural network features provide a significant improvement over the baseline, and that the system performance is better improved when neural network features are combined with baseline features. Literature [27] aims to solve the problems of unclear task objectives and incomplete feature learning in unsupervised quality assessment, saying that the proposed deep learning intelligent algorithm can effectively improve the relationship between machine scoring and human scoring in Chinese translation level evaluation. Literature [28] introduced a recursive neural network-based translation quality estimation model, and revealed the effectiveness of the method by applying the model to the sentence, word and phrase levels in the WMT16 quality assessment shared task. Literature [29] emphasizes the important role of translation quality assessment and the shortcomings of the current system, and exploits the synergy between two related tasks, word-level quality estimation and automatic post-editing, to achieve significant improvements on WMT16. Based on professionals' insights, literature [30] not only assessed the impact of AI on translation quality assessment, but also revealed that the quality of the source text has a significant impact on the accuracy of AI-assisted translation quality evaluation and emphasized that AI models are beneficial in improving the effectiveness of translation quality assessment. Literature [31] outlines a system for shared tasks for word-level quality estimation that incorporates a continuous-space deep neural network that learns bilingual feature representations from scratch and has a highly competitive performance in predicting word-level translation quality. Literature [32] describes an English translation evaluation system based on the BP neural network algorithm, which can provide a smarter machine translation service experience, can deeply optimize the problems such as underutilization of information in machine translation, and improve the performance of neural network machine translation.

The study combed the current machine translation quality assessment methods, selected mBERT, XLM-R and mBART as cross-language pre-training models, and carried out English language and literature translation quality evaluation by calculating the HTER values obtained from computerized machine translations and manual post-edited translations. Different ways are used to construct translation quality assessment models for different model structures, and a translation quality estimation model based on the pre-training model mBART is realized. Multiple datasets are selected for comparison experiments in the direction of English-Chinese and English-German translation, comparing the correlation coefficients and error indexes of the translation quality assessment models under different pre-training models, and exploring the evaluation performance of the proposed English language and literature translation quality assessment model. Then, we analyze the effects of translation sentence types and number of utterances on the performance of the model to verify the stability of the model. Finally, the proposed model suggests translation processing strategies for English language and literature translation.

2. Machine translation quality assessment methodology

Machine translation quality assessment plays an important role in improving the quality of translations produced by machine translation, improving the relevant algorithms of machine translation, and predicting the workload of post-translation editing after machine translation. With the technological development of machine translation, nowadays machine translation itself has experienced a long development, which has gone through rule-based translation methods, translation methods based on corpus statistics and so on.

2.1. Manual assessment

Manual evaluation is the main evaluation method for most of the translations before the emergence of machine translation, and there are various evaluation methods, including those based on functional linguistics and those using fuzzy mathematics with quantitative features, but there is no unified evaluation model so far. Therefore, the manual evaluation of the translation quality obtained by machine translation also has different evaluation characteristics depending on the evaluator. The manual evaluation of machine translation is mainly divided into two categories: the qualitative method using manual scoring and the quantitative method using fuzzy mathematics.

2.2. Automated assessment method

(1) Automatic evaluation method with reference translations

Automatic evaluation methods with reference translations are performed by counting the similarity between the reference translation as a standard translation result and the machine translation. The most used methods are the BLEU algorithm in N-element matching and the method based on edit distance.

BLEU is a kind of automatic evaluation method with reference translation proposed by IBM for the shortcomings of manual evaluation which is costly and cannot be reused.BLEU index has a very important significance in the field of evaluation of machine translation instructions because its evaluation value is close to that of manual evaluation, so it is the most used evaluation index in the automatic evaluation method with reference translation.BLEU's evaluation method adopts the N-gram The evaluation method of BLEU uses the N-gram matching rule to calculate a similarity ratio between the machine translation and the human translation in the range of n groups of words.

The editing distance refers to the minimum number of replacement, deletion, and insertion operations that the editor needs to go through in order to change the obtained machine translation into a human reference translation. The smaller the editing distance, i.e., the fewer the number of operations, the higher the similarity of the translations and the higher the quality of the machine translations.

(2) Automatic evaluation method for reference-free translations

Thanks to the development of artificial intelligence in recent years, more and more researchers focus on the automatic evaluation method of reference-free translation. The main technology involved is machine learning through the construction of neural networks, and the resulting model can be used to judge the translation. In this process, the training model is very important. Usually, the model is a neural network, whose structure contains an input layer, a hidden layer and an output layer. Among them, the hidden layer is the most important part for evaluating the effectiveness. In this paper, the deep neural network is used for the assessment of English language and literature translation quality, and the cross-language training model is combined to propose a model for the assessment of English literature translation quality.

3. Model for assessing the quality of English language and literature translations

The translation quality estimation task aims to predict the quality of a machine translation given the source text and the machine translated translation, thus the translation quality estimation task is a cross-lingual task. The performance of translation quality estimation is enhanced by feature extraction of information from source and target languages in a pre-training order model, and then migrating the cross-language representation capability learned from word mask prediction to the translation quality estimation task. This chapter proposes a translation quality estimation framework based on mBART, a cross-language pre-trained model with a complete Transfomer encoder-decoder structure, for English language and literature translation quality assessment.

3.1. Cross-language pre-training model

Cross-language pre-training model-based translation quality estimation model for English language and literature Three mainstream cross-language pre-training models are selected nowadays: mBERT, XLM-R, and mBART. While the infrastructure of the first two pre-trained models is based on the encoder structure of the Transformer, the backbone networks of mBART are all based on the encoder-decoder structure of the full Transformer.

3.1.1. Transformer

The Transformer model consists of two main parts: the Encoder and the Decoder, both of which use the same multilayer structure, in which the self-attention mechanism is the core of the Transformer model, which allows the model to focus on the entire input sequence while processing each input, thus being able to capture long-term dependencies in the sequence. In machine translation, the positional information of individual elements in sequence processing is crucial for understanding the overall structure and context of the sequence. However, the Transformer model itself does not have the ability to directly utilize sequence positional information. To solve this problem, it adopts a positional encoding approach, which aims to embed the positional information of each element in the sequence into the network so that the model can make full use of this positional information to better understand the sequence. Through position encoding, Transformer is able to take into account not only the content of the element itself but also its position in the sequence when processing the sequence, thus capturing the structural and semantic information of the sequence more accurately.

The specific operation is shown in the following equation:

$$PE(pos,2i) = \cos\frac{pos}{10000^{\frac{2i}{d_{model}}}} \tag{1}$$

$$PE(pos,2i+1) = \cos\frac{pos}{10000^{\frac{2i}{d_{model}}}} \tag{2}$$

where pos denotes the position of the word and i denotes the dimension of the word. Each dimension of the position encoding corresponds to a sinusoidal curve, and PE_{p+k} at any moment can be obtained from PE_p by a bit of trigonometric variation:

$$\sin(\alpha \pm \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta \tag{3}$$

$$\cos(\alpha \pm \beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta \tag{4}$$

The word vector embeddings are used as inputs to the Transformer model after being summed up with the position encoding and then fed into the Self-Attention of the encoder. Note that the Self-Attention calculation introduces three vectors of Query, Key, and Value denoted by Q, K, and V, respectively. The Query vector of each word is calculated by doing the inner product with the vectors of the entire sequence of words, respectively. The correlation scores between each element in the input sequence are calculated, and these correlation scores reflect the closeness of the relationship between different elements. Then, based on these scores, the model performs a weighted summation of the input sequence to obtain a new representation vector. This vector contains information about all the elements in the input sequence and emphasizes the part that is most relevant to the current element. The exact formula is given below:

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}}) \tag{5}$$

where $\sqrt{d_k}$ denotes the scaling factor. d_k is the dimension of Q,K,V.

The multi-head self-attention mechanism is an important component of the Transformer model, which splits the input sequence into multiple heads (or "layers" or "subspaces"), each of which independently performs self-attention computation. Each head then performs a weighted summation of the input sequence based on its own attention distribution to obtain the output of that head. The outputs of all heads are subsequently stitched together to form a richer representation. This representation contains information from different heads, so it is able to capture multiple contextual relationships in the input sequence. Finally, the spliced representation is converted into the final output by a linear transformation. The process is formulated as follows:

$$MultiHead(Q,K,V) = Concat(head_1,\cdots,head_n)W^\circ \tag{6}$$

$$head = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{7}$$

where W_i^Q, W_i^K, W_i^V is the weight of i nd and W° is the contextual linear mapping parameter generated by the splicing.

3.1.2. The mBERT model

The mBERT model is based on the Transformer encoder structure and is trained in a self-supervised manner on a Wikipedia corpus containing 104 languages. The model uses the WordPiece tool to cut molecular words, resulting in a shared word list size of 110,000. The model does not use manual labeling in the pre-training phase, but automatically constructs self-supervised labels from these input texts. Specifically, it is pretrained by two tasks:

(1) Masked Language Modeling Task (MLM): given any sentence, the model randomly masks out 15% of the words using a mask, and then inputs the masked sentence into the model and predicts the words corresponding to the masked tokens.

(2) Next Sentence Prediction (NSP): during pre-training, the model takes as input two randomly masked sentences spliced together using the “[SEP]” marker. The model must predict whether these two sentences are adjacent in the original text.

3.1.3. The XLM-R model

In order to explore the performance of BERT in cross-language scenarios, the cross-language pre-trained language model XLM was born. XLM uses BPE encoding without changing the BERT architecture, cuts the text of multiple languages into finer-grained subwords, and takes advantage of the grammatical similarity between the word formation laws of a single language and that of the same language family, which greatly reduces the number of word lists and mitigates the problem of excessive number of unregistered words in the inference. However, the number of training corpus of different languages is not consistent, which will lead to the problem of unbalanced weights of words in each language when constructing BPE fusion word lists, so when constructing BPE fusion word lists, it is necessary to resample the training data with the resampling probability:

$$q_i = \frac{p_i^{\alpha}}{\sum_{j=1}^{N} p_j^{\alpha}} \tag{8}$$

$$p_i = \frac{n_i}{\sum_{k=1}^{N} n_k} \tag{9}$$

n_i denotes the corpus number of the i nd language, p_i denotes the corpus share of the i th language, which is smoothed to get the final sampling probability q_i , where the smoothing coefficient a is taken as 0.5. The BPE word list constructed by training corpus resampling ensures that the low-resource languages take up a certain proportion of the word list construction without affecting the status of the high-frequency languages in the word list.

The XLM-R model is an improved version of the XLM model. Compared to XLM, XLM-R no longer relies on a bilingual parallel corpus and uses larger training data and more languages.The main body of the model in XLM-R is still the Transformer, and the training target is the multilingual MLM.Unlike XLM, XLM-R uses a larger shared vocabulary list and increases the overall model to 550 million references. In addition, XLM-R introduces a new parameter-change strategy that involves sampling low-resource languages during training and vocabulary construction.

3.1.4. mBART model

Unlike XLM and mBART, which only utilize the encoder structure, mBART is based on the complete Transformer structure and is pre-trained with the help of a sequence-to-sequence noise reduction self-encoding task on a reptilian corpus containing 25 languages (CC25). mBART employs two forms of noise addition:

(1) Randomly selecting 35% of text segments based on a Poisson distribution and replacing them with masked tokens.

(2) Using positional offsets to disrupt the order of subwords in the original sentence.

In the training phase, the loss of the model can be expressed as:

$$L_{\theta} = \sum \log P(X \mid g(X); \theta) \tag{10}$$

Where X denotes the original input text and $g(X)$ denotes noise addition to the input text.

3.2. Translation Quality Estimation Model

This paper focuses on sentence-level translation quality estimation in English language and literature. The task of sentence-level translation quality estimation is the process of scoring the whole sentence as a unit. The scoring process is as follows: the data included in the training are the translations obtained by inputting the source language and the translation system, the human post-edited translations obtained by the translation practitioners who manually edited the translations according to the text in the source language, and the HTER values obtained by scripting the machine translations and the human post-edited translations.The specific details of the HTER values are obtained by the TERCOM script by comparing the machine translations and the human post-edited translations, the number of substitution, insertion, deletion and moving operations, and the number of operations are compared with the machine translations and human post-edited translations. The specific details of the HTER value are obtained through the TERCOM script by comparing the machine translation and the manual post-editing translation to get the number of replacement, insertion, deletion, and movement operations, adding up the number of operations to get the total number of operations, and dividing it by the number of words in the manual post-editing translation. The formula is shown below:

$$HTER = \frac{\#\text{insert} + \#\text{delete} + \#\text{substitute} + \#\text{shift}}{\#\text{human reference words}} \tag{11}$$

It can be guided from the formula that HTER is a value between 0 and 1. The closer the value is to 1, it means that the higher percentage of post-edited translations edited is obtained, implying that the quality of translations is poor. On the contrary, the closer the value is to 0, the lower the proportion of post-edited translations edited obtained, implying a better quality of translations.

Cross-language pre-training models have strong characterization ability on cross-language tasks, and both the pre-training method and pre-training corpus lead to cross-language pre-training models suitable for handling natural language processing tasks. Cross-language pretraining models are categorized into models with Transformer's Encoder as the structure such as mBERT and XLM-R and models with Transformer's full Encoder-Decoder as the structure such as mBART. Therefore different approaches are used to construct translation quality estimation models for different model structures.

3.2.1. Encoder-based architecture

For the structure of translation quality estimation model based on encoder structure can be structured in two ways, BiEncoder structure and CrossEncoder structure.

The specific modeling details of the method are described first. For the Bi-Encoder structure, n and $n^{'}$ denote the sequence lengths of the input source text, machine translated text, i.e., the number of tokens after disambiguation, respectively. After input to the cross-language pre-trained language model, for each input token, after the cross-language pre-trained model, the last layer of hidden vector from the output is obtained h_D , and average pooling is done on this hidden vector. The sentence limited representation of the source language text, machine translated text is obtained u, v , the process is shown in Eq:

$$u = \text{mean}(E(src)) = \frac{1}{n}\sum_{t=1}^{n} h_D(src)_t \tag{12}$$

$$v = mean(E(mt)) = \frac{1}{n^{'}}\sum_{t=1}^{n^{'}} h_D(mt)_t \tag{13}$$

Next, the cosine similarity was calculated by putting u, v and the result obtained from the calculation was used as the HTER score:

$$output = \cos(u, v) \tag{14}$$

Among them:

$$\cos(u, v) = \frac{u \cdot v}{\| u \| \| v \|} = \frac{\sum_{i=1}^{n} u_i \times v_i}{\sqrt{\sum_{i=1}^{n} (u_i)^2} \times \sqrt{\sum_{i=1}^{n} (v_i)^2}} \tag{15}$$

Finally, since the sentence-level translation quality estimation task serves as a regression task, this model uses the mean square error loss function (MSE):

$$MSE(x,\tilde{x}) = \frac{1}{n}\sum_{i=1}^{n}(x-\tilde{x})^2 \tag{16}$$

$$L_{sentence\_level} = MSE(HTER, output) \tag{17}$$

For the Cross-Encoder structure, the source text src and the machine translation mt are spliced into src<s>mt as input to the model. Then the <CLS>token of the hidden state of the last layer obtained after the model is taken out, and the corresponding hidden vector representations of the source text and the machine translation are obtained u. Next, u is input into a layer of Fully Connected Layer (FC), and sigmoid is used as the activation function, and the final output representations of the outputs which take the values between 0 and 1 are obtained:

$$output = sigmoid(FC(u)) \tag{18}$$

Among them:

$$sigmoid(x) = \frac{1}{1+e^{-x}} R \to (0,1) \tag{19}$$

Finally, Cross Encoder also uses mean square error (MSE) as a loss function:

$$L_{\text{sentence level}} = MSE(HTER, output) \tag{20}$$

3.2.2. Encoder-Decoder Based

In this paper, an mBART-based translation quality estimation model applicable to sentence-level QE tasks in English language and literature is proposed. Firstly, the source text src is input to encoder, the machine translated text tgt is input to decoder, and the <eos> token of the last layer of the hidden state output from decoder is taken, and the sentence vector representation of the source language text and the machine translated text is obtained u. Then, similar to the cross-language pre-training model based on the Transformer Encoder, the HTER prediction is obtained by taking u as the input, the predicted valueoutput of HTER is obtained by going through one fully connected layer (FC) and sigmoid activation function, respectively:

$$output = sigmoid(FC(u)) \tag{21}$$

Finally, Cross Encoder also uses mean square error (MSE) as a loss function:

$$L_{\text{sentence level}} = MSE(HTER, output) \tag{22}$$

4. Experimental analysis of the model

4.1. Experimental data set

In this section, the translation quality estimation task is experimented on two language pair directions, English-Chinese and English-German. Since the English language and literature translation quality estimation model proposed in this section is divided into a pre-training phase and a translation quality estimation phase, two different datasets need to be used.

For the pre-training corpus, parallel corpora as similar as possible to the English language and literature translation quality estimation task in the Chinese-English and English-German directions are selected. Among them, the Chinese-English parallel corpus comes from the CCMT2020 Chinese-English news domain machine translation task and the parallel sentence pairs consisting of source language sentences and human post-edited translations for the CCMT2020 translation quality estimation task, while the English-German parallel corpus comes from the parallel sentence pairs consisting of source language sentences and human post-edited translations for the WMT17 English-German news machine translation task, the WMT2017 translation quality estimation task , the TED2020 v1 dataset, and the QED dataset. Parallel sentence pairs with sentence length greater than 70 are filtered out, and the length ratio between bilingual sentence pairs is controlled to lie between 1/3 and 3 to ensure the quality of the parallel corpus.

For the translation quality estimation datasets, the sentence-level and word-level datasets related to translation quality estimation in the direction of CCMT2019 Chinese ↔ English (CN-EN) and

WMT2017 English ↔ German (EN-DE) are selected in this paper, respectively.

4.2. Experimental evaluation indicators

Spearman's correlation coefficient (Spearman), Pearson's correlation coefficient (Pearson), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used as the four indexes to measure the performance of a model for estimating the quality of translation. Spearman's correlation coefficient and Pearson's correlation coefficient are used as the main indicators, the higher the value of the two, the better the performance of the model for estimating the quality of English language and literature translation, and the lower the value of the two, the better the performance of the model, with the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) as the reference indicators.

4.3. Experimental results

4.3.1. Quality of English-German translations

On the WMT2017 English-German sentence-level task, the classical Predictor-Estimator as well as the Bilingual Expert model were used as the two baseline models for this experiment. The two models gained top positions on all subtasks related to translation quality estimation on WMT17 and WMT18, respectively. In addition, in order to better compare the model performance, the source code provided by Bilingual Expert is reproduced in this paper, and a pre-training corpus that is consistent with this model is selected to train the predictor of Bilingual Expert.

The experimental results of translation quality estimation from English to German and from German to English are shown in Figures 1 and 2. Where, (1) ~ (7) denote Predictor-Estimator paper result, Bilingual Expert paper result, Bilingual Expert with our implement, Our Model with BERT pretrained, Our Model with mBERT pretrained, and Our Model with mBERT pretrained, respectively. , Our Model with mBERT pretrained, Our Model with XLM-R pretrained, Our Model with mBART pretrained.

From the experimental results, it is found that the translation quality estimation models based on pretrained language models end up with good performance when BERT, mBERT, XLM-R, and mBART are chosen as the models in the pretraining stage. Among them, the models with XLM-R and mBART are chosen to perform better, outperforming the Predictor-Estimator baseline model and the Bilingual Expert model trained with the same corpus in this paper on both WMT2017 EN-DE and DE-EN tasks. In addition, the mBART-based pre-trained model in this paper gets the best experimental results, with Spearman and Pearson of 0.729 and 0.686 on the WMT2017 EN-DE dataset, and 0.654 and 0.727 on the DE-EN dataset, which both outperform the baseline model by 0 to 4 percentage points.

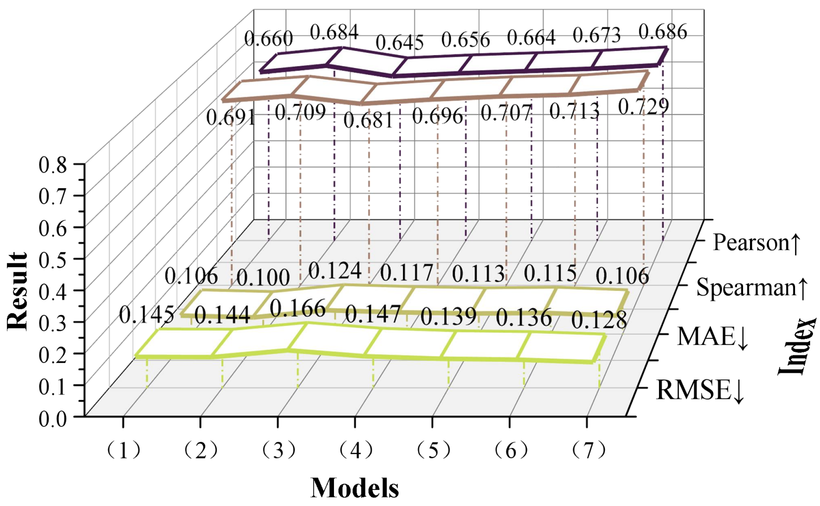

Figure 1. The quality estimate result about English translating into German

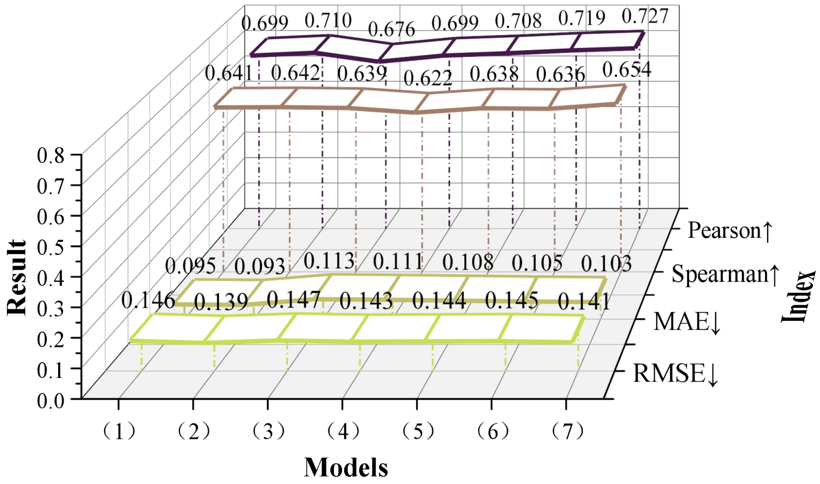

Figure 2. The quality estimate result about German translating into English

4.3.2. Quality of Chinese-English translation

The results of the translation quality estimation experiment for English into Chinese and the translation quality estimation experiment for Chinese into English are shown in Figures 3 and 4. Among them, (1)~(5) represent Bilingual Expert with our implement, Our Model with BERT pretrained, Our Model with mBERT pretrained, Our Model with XLM-R pretrained, and Our Model with mBART pretrained, respectively. Since, the Bilingual Expert model has not published modeling results under this dataset, the Bilingual Expert model reproduced in this paper based on the same pre-training corpus is used as the baseline model.

The experimental results show that the mBART model proposed in this paper also has excellent results on the CCMT2019 Chinese-English sentence-level task, and the translation quality estimation model based on the four pre-trained models outperforms the baseline model on the CCMT2019 Chinese-English task. In addition, the English language and literature translation quality estimation model based on the mBART pre-training in this paper achieves the best results among all the models, with Spearman values exceeding the baseline model by 13.3 percentage points and 5.6 percentage points in the CCMT2019 EN-ZH and ZH-EN datasets, respectively.

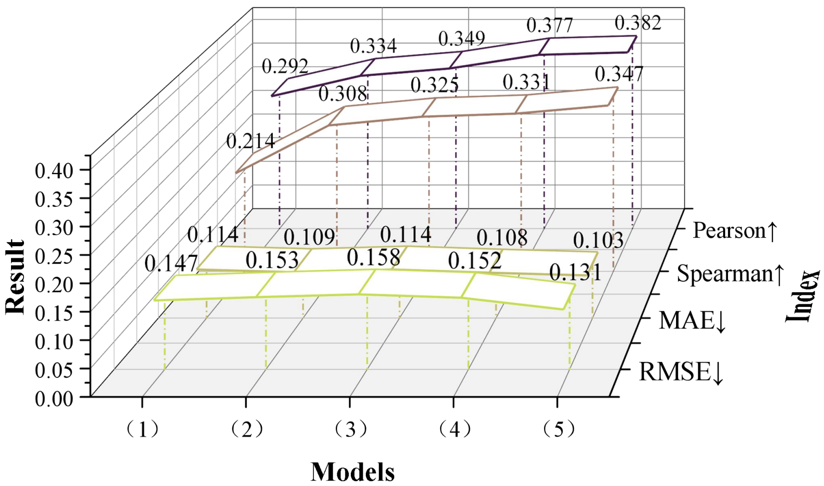

Figure 3. The quality estimate result about English translating into Chinese

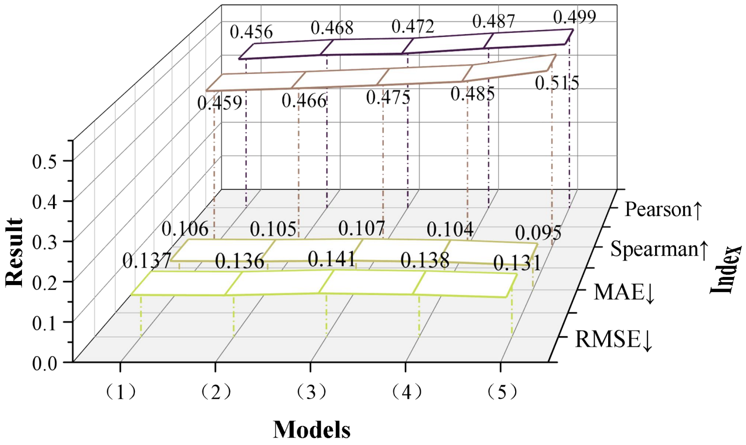

Figure 4. The quality estimate result about Chinese translating into English

4.4. Other experiments

The experiments in Section 4.3 show that the English language and literature translation quality estimation model based on the mBART pre-trained language model proposed in this paper outperforms other pre-trained language models on the sentence-level QE task. The performance of the model under different factors is explored next.

4.4.1. Influence of Translated Sentence Types

Based on the CCMT2019 dataset, the utterance types of English-Chinese translations are set to be declarative, special usage, interrogative, juxtaposed compound, and exclamatory sentences to test the proposed model to assess the quality of translations in English language and literature. The effect of translation sentence types on the assessment performance of the proposed model is shown in Figure 5. The results of Spearman, Pearson, MAE and RMSE under different sentence types are 0.713~0.728, 0.652~0.664, 0.098~0.112, 0.133~0.148, respectively.The differences between the assessment results of the proposed model and the actual situation are relatively small under different translation sentence types, and only the results of the assessment of translation quality of the special usage sentences are slightly low In the case of the proposed model, there is no significant effect of declarative sentences, interrogative sentences, juxtaposed compound sentences and exclamatory sentences on the assessment effect of the proposed model.

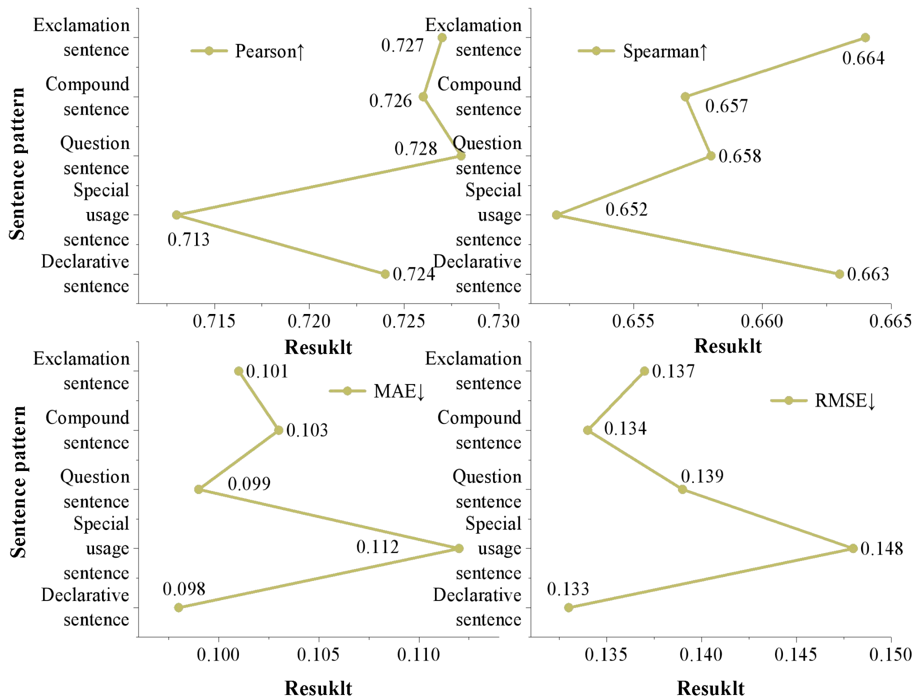

Figure 5. The effect of the translation pattern on the performance of the model

4.4.2. Effect of number of statements

The effect of the number of utterances on the model assessment performance is shown in Figure 6. The number of utterances does not have a significant effect on the assessment performance of the proposed model, and when the number of utterances is increased from 500 to 8,000, the Pearson of the English Language and Literature Translation Quality Assessment Model is increased from 0.715 to 0.729, the Spearman is increased from 0.650 to 0.664, and the MAE and the RMSE are always below 0.16, which shows that the proposed English Language and Literature Translation Quality Assessment Model has a good performance in using it.

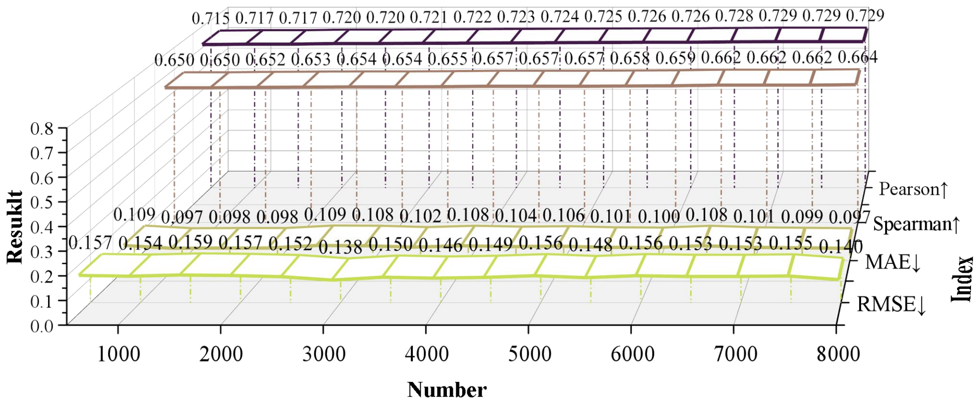

Figure 6. The effect of the statement number on the performance of the model

In order to further analyze the effectiveness of the proposed model, it is compared with Predictor-Estimator and Bilingual Expert methods to get the Pearson comparison of the three models, and the Pearson comparison of the three methods under different number of utterances is shown in Figure 7. No matter the number of utterances is 500~4000 or 4000~8000, the Pearson of the proposed English Language and Literature Translation Quality Assessment Model is much higher than the other two methods, and the accuracy is up to 0.733, and it is more stable, which is of certain application value.

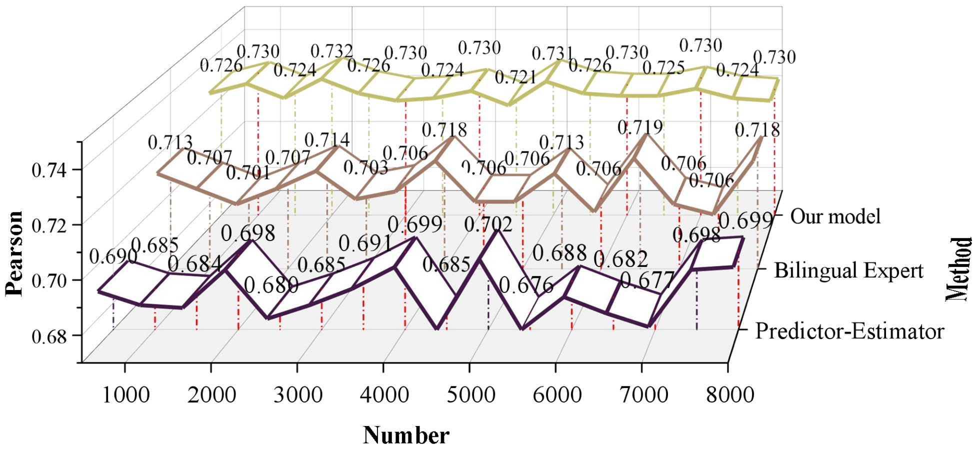

Figure 7. Pearson comparison of three methods of different statement quantity

In order to better analyze the application value of the proposed model, the quality evaluation efficiencies of the three models are compared again, and the quality evaluation efficiencies of the three methods under different number of utterances are shown in Figure 8. While guaranteeing the accuracy of quality assessment, the translation quality assessment efficiency of the English Language and Literature Translation Quality Assessment Model is much higher than that of the other two methods, with the highest quality assessment efficiency above 96%, which can prove the feasibility of the proposed model to a certain extent.

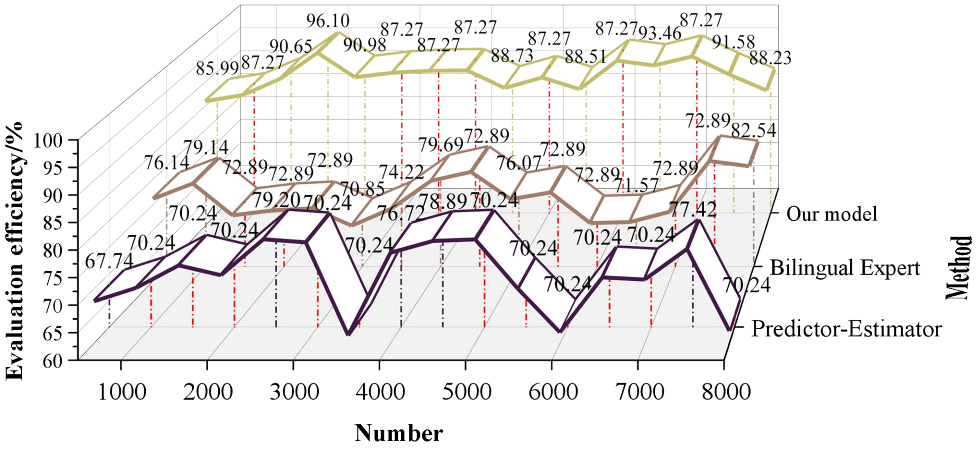

Figure 8. The quality evaluation efficiency of the three methods of different statement quantity

5. Strategies for dealing with the translation of English literature

The translation quality estimation model of mBART, a cross-language pre-training model proposed in the study, can be used in the translation evaluation of English language and literature, and the translation of English literature can not just be a simple “direct translation”, but also strengthen the treatment of the artistic language, which not only accurately expresses the meaning of the original text, but also embodies the characteristics of the culture, so this chapter discusses the translation processing strategies in English language and literature. Therefore, this chapter discusses the translation processing strategy in English language and literature.

5.1. Fidelity to the original text

Only on the basis of respecting the original and being faithful to the original text can the accuracy of translation be better guaranteed, and the meaning of the original text as well as the thoughts and feelings contained in it can be presented to the readers in the best possible way. Based on the original text, we can flexibly choose reasonable translation methods to translate, and skillfully process and interpret the artistic language in English literature, so that the translated language is full of artistry and can accurately express the meaning, thoughts, culture and other contents of the original text.

5.2. Integration of the context of creation

English literature created in different social environments and times also has a strong sense of the times, English literature has become a carrier to reflect and embody the social and cultural environment and the development of the times to a certain extent, and each English literary work has its own unique creative background. Therefore, in the translation of English literature, the translator should have a deep understanding of the creative background of the original work and follow the principle of combining the creative background in the treatment of artistic language. Only in this way can we fully understand and realize the original creator's life state, social experience and ideological condition, etc., so that we can do a good job in translation and more perfectly interpret the connotation of the original work.

5.3. Fluidity

The treatment of artistic language in English literary translation should be as smooth as possible, complete in meaning, natural in sentence articulation, logical and organized, coherent in discourse, able to accurately convey the meaning of the original text, and the translated translation should be compatible with the reader's aesthetic interest.

5.4. Rational planning

Translators should follow the principle of rational planning whether in the translation of English literary works or in the treatment of artistic language. This will reasonably transform the artistic language of English literature into a language that meets the logical thinking, reading habits and comprehension ability of the majority of readers, so that the readers can read the translated English literature as smoothly and naturally as they read their own literature, and not only understand it, but also deeply understand the ideological and cultural connotations of the original work, and feel the emotional world of the creator of the original work.

6. Conclusion

Translation quality estimation is a kind of automatic evaluation technique for machine translation, which is widely used in translation work. In this paper, for English language and literature translation, a translation quality estimation model based on mBART, a pre-training model with Transformer encoder-decoder structure, is proposed. It is compared with mainstream cross-language pretraining models to explore the evaluation performance of the model. The test results of this paper's model in English and German translation and English and Chinese translation are better than other models, with correlation coefficient metrics 0~4% and 0~13.3% higher than those of the baseline model, and the average absolute error and root mean square error are both below 0.15. In addition, the translation sentence type and the number of utterances do not have a significant effect on the performance of the English Language and Literature translation quality assessment model, and the translation quality assessment of this paper's model is up to 96% under different numbers of utterances. Experiments show that the proposed translation quality assessment model based on the pre-training model has excellent accuracy and stability, and can be used in English language and literature translation quality assessment. The article concludes with a discussion on the processing strategies of English language and literature translation, proposing suggestions in the dimensions of fidelity to the original text, combining with the creative background, fluency and reasonable planning, with a view to making the translations clearer and more accurate, and better meeting the reading needs of the readers.

Fund Project
This paper is the final research outcome of the project of Social Science Development & Research Project in Hebei Province, titled "A Study on Translation Strategies for International Dissemination of Hebei's Local Culture from the Perspective of Eco-Translatology". (Project Number: 202402134).

References

1. Ford, B. Q., Dmitrieva, J. O., Heller, D., Chentsova-Dutton, Y., Grossmann, I., Tamir, M., ... & Mauss, I. B. (2015). Culture shapes whether the pursuit of happiness predicts higher or lower well-being. Journal of Experimental Psychology: General, 144(6), 1053.
2. Sharma, B. (2022). Spirituality and Materialistic World: Exploring the Interplay between Inner Fulfilment and External Pursuits. Multi-Dimensional Aspects of Management, 213.
3. Aniuranti, A., Faiza, D., & Wulandari, Y. (2021). Enhancing students' understanding of English grammar through literary works and exploring students' voice on their use. Scholaria: Jurnal Pendidikan Dan Kebudayaan, 11(3), 264-270.
4. Afandi, M. D., Muflihah, T., Aquariza, N. R., & Machmudah, M. (2022). Learning English through Literary Works. Indonesia Berdaya: Journal of Community Engagement, 3(4), 987-994.
5. Hussain, S. S. (2017). Translating a literary text: enigma or enterprise. International Journal of English Language & Translation Studies, 5(2), 75-78.
6. Karjagdiu, L., & Mrasori, N. (2022). The role of literary translation in the development and enrichment of national literature. Journal of Language and Linguistic Studies, 17(4).
7. Robinson, D. (2017). What kind of literature is a literary translation?. Target, 29(3), 440-463.
8. Sun, Y. (2022). Literary translation and communication. Frontiers in Communication, 7, 1073773.
9. Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2022). Progress in machine translation. Engineering, 18, 143-153.
10. Stahlberg, F. (2020). Neural machine translation: A review. Journal of Artificial Intelligence Research, 69, 343-418.
11. Poibeau, T. (2017). Machine translation. MIT Press.
12. Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: a systematic review. Language Resources and Evaluation, 56(2), 593-619.

13. Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018). Approaches to human and machine translation quality assessment. Translation quality assessment: From principles to practice, 9-38.
14. Popović, M. (2018). Error classification and analysis for machine translation quality assessment. Translation quality assessment: From principles to practice, 129-158.
15. Kasperavičienė, R., Motiejūnienė, J., & Patašienė, I. (2020). Quality assessment of machine translation output: Cognitive evaluation approach in an eye tracking experiment. Texto livre: linguagem e tecnologia, 13(2), 271-285.
16. Chatzikoumi, E. (2020). How to evaluate machine translation: A review of automated and human metrics. Natural Language Engineering, 26(2), 137-161.
17. Li, H., & Deng, J. (2022). Unreferenced English articles' translation quality-oriented automatic evaluation technology using sparse autoencoder under the background of deep learning. Plos one, 17(7), e0270308.
18. Maučec, M. S., & Donaj, G. (2019). Machine translation and the evaluation of its quality. In Recent trends in computational intelligence. IntechOpen.
19. Tian, Y., Khanna, S., & Pljonkin, A. (2021). Research on machine translation of deep neural network learning model based on ontology. Informatica, 45(5).
20. Zhao, C. (2022, December). Multi-Feature Fusion Machine Translation Quality Evaluation Based on LSTM Neural Network. In 2022 6th Asian Conference on Artificial Intelligence Technology (ACAIT) (pp. 1-5). IEEE.
21. Specia, L., Scarton, C., & Paetzold, G. H. (2018). Quality estimation for machine translation. Morgan & Claypool Publishers.
22. Han, C. (2020). Translation quality assessment: a critical methodological review. The Translator, 26(3), 257-273.
23. Mohammed, S. Y., & Aljanabi, M. (2024). Advancing Translation Quality Assessment: Integrating AI Models for Real-time Feedback. EDRAAK, 2024, 1-7.
24. Gong, M. (2024). The neural network algorithm-based quality assessment method for university English translation. Network: Computation in Neural Systems, 1-13.
25. Guzmán, F., Joty, S., Màrquez, L., & Nakov, P. (2017). Machine translation evaluation with neural networks. Computer Speech & Language, 45, 180-200.
26. Chen, Z., Tan, Y., Zhang, C., Xiang, Q., Zhang, L., Li, M., & Wang, M. (2017, September). Improving machine translation quality estimation with neural network features. In Proceedings of the Second Conference on Machine Translation (pp. 551-555).
27. Gao, Y. (2024, August). Research on Intelligent Algorithms of Deep Learning in English Translation Quality Assessment. In 2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE) (pp. 1088-1092). IEEE.
28. Kim, H., & Lee, J. H. (2016, August). Recurrent neural network based translation quality estimation. In Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers (pp. 787-792).
29. Martins, A. F., Junczys-Dowmunt, M., Kepler, F. N., Astudillo, R., Hokamp, C., & Grundkiewicz, R. (2017). Pushing the limits of translation quality estimation. Transactions of the Association for Computational Linguistics, 5, 205-218.
30. Koka, N. A., Akan, M. F., Kana'n, B. H. I., Khan, M. R., Zulfiquar, F., & Jan, N. (2023). Impact of artificial intelligence (ai) on translation quality: assessment and evaluation. Journal of Southwest Jiaotong University, 58(4).
31. Kreutzer, J., Schamoni, S., & Riezler, S. (2015, September). Quality estimation from scratch (quetch): Deep learning for word-level translation quality estimation. In Proceedings of the Tenth Workshop on Statistical Machine Translation (pp. 316-322).

32. Hu, Y. (2024). English translation evaluation method based on BP neural network. Procedia Computer Science, 243, 1049-1058.