Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

A Language Model-based Approach to Context Analysis in Business English Translation

Xiaojing Li[1,] ✉

[1] *Applied Foreign Language and International Education Department, Luohe Vocational Technology College, Luohe, Henan, 462000, China*

ABSTRACT

In this paper, we use a large language model for business English translation and context analysis, and propose an adaptive parameter unfreezing method based on the quantization difference between adjacent layers within the decoder to fine-tune the layers of the language model related to the translation task, and to understand the behavior of the model in the relevant layers. Then the method of combining different encoders is proposed as a dual encoding-decoding framework on top of the traditional encoding-decoding framework, which is applied to the task of context analysis in business English translation. The fine-tuning method in this paper significantly improves the text translation quality of the language model, especially in the English-X tri-lingualization, which improves the COMET and BLEU metrics by 3.22 and 2.58 points respectively. In addition, the dual encoding-decoding model proposed in this paper is applicable to the task of contextual analysis in business English translation, which significantly improves the performance of contextual analysis in business English, and the F1 value on the HIT-CDTB dataset is improved by 11.60% compared with that of Rutherford's model. The experiment proves that the proposed method of text has made progress in the research of the task of analyzing textual contextual relations in business English.

Keywords: Big Language Modeling, BLEU Indicators, Machine Translation, Context Analysis

1. Introduction

Along with the rapid development of economic globalization, English, as an international communication language, plays an increasingly stronger role in international business communication and plays a decisive role in the success or failure of business activities [1-2]. Business English is mainly used in economic and trade activities between Chinese and Western countries, respecting the differences between countries, using civilized, decent, sincere and polite language, reflecting good professional ethics and personal cultivation. In different cultural contexts, there are great differences

✉ Corresponding author.
E-mail address: mslijing1617@163.com (X. Li).
Received 25 July 2024; Revised 15 September 2024; Accepted 10 January 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-044

in human ideology and behavior, and people's understanding of the same things are also very different [3-5]. Chinese people emphasize abstract thinking and emotional thinking, but Western people emphasize logical thinking and rational thinking [6]. As a result, the Chinese and Westerners will have a great difference in understanding, and the difference in thinking and culture will make business English translation more difficult, which will often lead to misunderstanding and poor understanding of words in translation, thus making business activities suffer great obstacles and difficulties [7-10]. In order to promote the rapid development of China's economy, business English professionals should ensure the accuracy and scientificity of business translation to realize the normal and orderly development of business activities [11-12]. If they want to translate business English accurately and appropriately, they should have a comprehensive understanding of the politics, customs, habits and history of the countries corresponding to the translated language, and understand the social and cultural differences between each country [13-14]. At the same time, business English requires that translators should have a comprehensive understanding of business and trade terminology, the characteristics of business English, a broad international vision, reasonable knowledge of international business, and accurate and skillful translation skills [15-16].

Business English text has obvious professional characteristics, with a large number of new words, with the help of corpus technology can be standardized into programmed expressions, but also can be composed of different types of specialized corpus, to help the translation practice to carry out smoothly. Literature [17] shows that bilingual parallel corpus can provide rich teaching materials for translation teaching, improve business English writing skills by constructing business English writing corpus, and provide convenient translation means for translation practice. Literature [18] explores the potential features of Chinese-English business translated texts in terms of pragmatic collocation, and examines the two translation commonalities of simplification and clarity through the corpus, and the study shows that, compared with the native business English, the translated texts utilize a large number of free combinations and collocations, which reflect the simple and clear translation features. Literature [19] adopts a corpus-based approach to analyze the differences and similarities of discourse connectives in different contexts, and by compiling and processing the frequency of discourse connectives in the corpus, it can provide reference for translation teaching or practice.

Based on the premise of continuous updating of business English knowledge and combining different cross-cultural dimensions, the business English translation corpus can be updated with the times, thus further enhancing the accuracy of cross-cultural business English translation. Literature [20] fully considered the cultural differences and audience psychology in the translation process of city publicity, and proposed a corpus combining EM algorithms to reconstruct the cross-cultural translated text, so that the translated text is easy to be accepted by readers through more accurate translation. Literature [21] respectively constructed a corpus containing website information of foreign-invested enterprises and local enterprises, cross-culturally interpreted the enterprise data from Hofstede's cultural dimension, and put forward suggestions for website construction adapted to the local culture, realizing the effectiveness of cross-cultural business communication. Literature [22] studied the cross-cultural translation strategy of CSR reports and proposed a parallel corpus based on English and Italian sources, based on which the English CSR reports were translated into a special discourse form suitable for the Italian culture, which played a significant role in conveying the corporate values and enhancing the corporate image. Literature [23] examined the use of metadiscourse nouns in Chinese and American CSR reports, and the results of the corpus analysis showed that the differences between the United States and China in high and low contexts resulted in a higher frequency of metadiscourse nouns with interactions in the American discourse, which revealed the cultural differences in translating business English in different contexts.

This paper investigates a language model-based approach to neural machine translation, and proposes a fine-tuning method to optimize the performance of a business English translation task in order to construct a language model that goes beyond traditional neural machine translation. The

method fine-tunes the unfreezing parameters through singular value decomposition to determine the structure and properties associated with the particular layer to be unfrozen. Then an encoder is added to the original encoding one decoding framework so that each source language corresponds to its respective encoder one by one, while the attention mechanism for multilingual input is explored. Relevant experiments are designed to verify the performance of the large language model and its fine-tuning method on business English translation, and the attention model is applied to practice to verify the effect of its contextual analysis.

2. Machine translation tasks for large language models and their fine-tuning methods

2.1. Large Language Model

Large language models usually use massive amounts of unlabeled text for unsupervised or semi-supervised learning. Most of their structures are based on the Transformer [24] architecture, so they are highly generalizable and can achieve relatively good results in numerous benchmark tests.

Recent studies have shown that modelability can be significantly improved by increasing the parameters and data size of large pre-trained language models Accompanied by the birth of the emergence effect, three major capabilities have emerged from large language models, namely, In-context Learning, Instruction Fine-tuning, and Thought Chaining.

1) In-context learning. The core idea of In-context learning is to learn by analogy.In-context learning combines the question to be queried (i.e., the input that needs to be labeled for prediction) with a set of contextual examples (relevant cases) to form an input with hints, which is then fed into the language model for prediction.

2) Instruction tuning. Unlike In-context learning which stimulates the complementary ability of a large language model, the main role of Instruction tuning is to stimulate the comprehension ability of a large language model, by giving more obvious instructions to make the model make more correct actions.

3) Chain-of-thought. Chain-of-thought breaks down a logical reasoning problem into multiple steps on the outside of the model, step by step, so that the generated result has a clearer pulse. The method is very simple, mainly giving some examples and encouraging the big model to explain the reasoning process, so that the reasoning of the process is also carried out when outputting the result, which leads to a better result.

Big Language Models are usually trained using large-scale textual data from different domains, which makes them perform well for general language understanding. However, for domain-specific tasks (e.g., medical, legal, or technical domains), the distribution of the input data may differ from that used by the model in the pre-training phase. At this point, this distributional bias needs to be reduced by fine-tuning on the data for the target task, which can make the model more adapted to the characteristics of the data in real applications. The number of parameters of a large language model is huge, and the price paid for full parameter fine-tuning is extremely high, so some special methods are needed for efficient fine-tuning:

(1) Adapter fine-tuning. The Adapter module [25] adds several sets of trainable parameters to the large model for each task and keeps the original network parameters fixed, the structure of the Adapter module is shown in Figure 1.

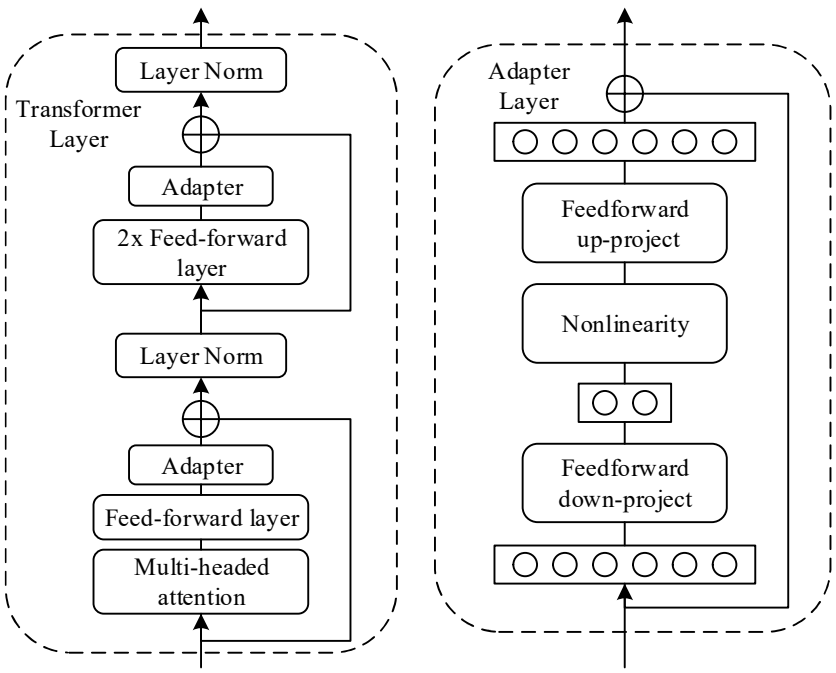

Fig. 1. Adapter structure diagram

Adding two Adapter modules to each Tranformer layer ensures efficient training by holding the parameters of the original pre-trained model constant during training, and only fine-tuning the added Adapter structure and Layer Norm layer. There is a residual structure in this module to re-add the inputs of the Adapter to the outputs, which ensures that even if the parameters of the Adapter are initialized close to 0 at the beginning, the result is close to a constant mapping due to the presence of the residual structure, thus ensuring the effectiveness of the training.

(2) Low Rank Adapter (LoRA). The goal of LoRA is to indirectly train some dense layers in a neural network by optimizing the rank decomposition matrix of the dense layer changes during the adaptation process, while keeping the pre-trained weights constant. The theoretical basis for this is that models are over-parameterized, they have smaller intrinsic dimensions, and the model relies heavily on this low intrinsic dimension to do task adaptation.

2.2. Machine Translation Evaluation Criteria

The development of machine translation has led to the advancement of evaluation metrics to ensure reliable assessment of translation quality. Instead of extending the more conventional BLEU [26] metric, this study adopts SacreBLEU as the main evaluation metric to measure the machine translation ability of the model, which takes the value range of 0-1. In order to better compare the experimental results, the results of SacreBLEU are usually translated into the form of a percentage, i.e., the range becomes 1-100.SacreBLEU is based on the BLEU metric. A modified form of SacreBLEU, it improves the consistency and reproducibility of the assessment by using a fixed tokenization method and a reference translation. The formula is as follows:

$$SacreBLEU = BP \times \exp\left(\sum_{n=1}^{N} \omega_n^N \log p_n\right) \tag{1}$$

where p_n is the n-gram precision, which expresses the proportion of n-grams (consecutive phrases of length n) that correctly match the reference translation in the machine translation output. ω_n is the n-gram weight.

BP is the length penalty term, which is used to penalize translations that are too short, because too short outputs may improve n-gram accuracy but do not necessarily reflect high quality translations. It is calculated as if the length of the machine translation output is less than the length of the reference translation:

$$BP = \exp\left(1 - \frac{\text{Reference length}}{\text{Output length}}\right) \tag{2}$$

$BP = 1$ if the length of the machine translation output is greater than or equal to the length of the reference translation.

2.3. Task-related layer fine-tuning methodology proposed

Although large language models enable translation, when performing translation in a decoder-only architecture, the definition and context of cross-language interactions tend to become ambiguous and variable, making it difficult for the model to define cross-language interaction rules. In this paper, we propose an adaptive parameter unfreezing strategy based on quantized differences between neighboring layers within the decoder. The method consists of two steps. The first step is to determine the specific layers to be unfrozen by SVD analysis, and the second step is to fine-tune the unfrozen parameters.

Let $A^{(l)}$ denote the attentional weight of layer l in the decoder, whose singular value decomposition (SVD) can be formulated as:

$$A^{(l)} = U^{(l)}\Sigma^{(l)}\left(V^{(l)}\right)^T \tag{3}$$

where $U^{(l)}$, $\Sigma^{(l)}$ and $\mathrm{V}^{(l)}$ represent the left singular vector, singular value and right singular vector of the attentional weights $A^{(l)}$ at layer l, respectively. Given that the singular values $\Sigma^{(l)}$ capture the energy or importance of different components in the attention mechanism, the difference $D(l, l+1)$ between two neighboring layers l and $l+1$ can be quantified by comparing their respective singular values. The metric for such quantification can be defined as:

$$D(l, l+1) = \left\|\Sigma^{(l)} - \Sigma^{(l+1)}\right\|_F \tag{4}$$

Here $\|\cdot\|_F$ denotes the Frobenius paradigm. To assess the significance of the difference, in this paper we evaluate $D(l, l+1)$ on the validation set.Formally, we define $\bar{D}(l, l+1)$ as the average difference of the n examples between layers l and $l+1$.

$$\bar{D}(l, l+1) = \frac{1}{n}\sum_{i=1}^{n} D_i(l, l+1) \tag{5}$$

where $D_i(l, l+1)$ is the difference between layers l and $l+1$ in the ind example. Layer l to be thawed is identified by the following:

$$l^* = \arg\max_l \bar{D}(l, l+1) \tag{6}$$

For a given translation of a source language into multiple target languages, we evaluate the interlayer differences individually for each target language, and thus determine the layers to unfreeze for that particular language pair. Formally, let L denote the set of target languages, and for each target language $lang \in L$, we compute its interlayer variance $\bar{D}(l, l+1)$ and determine the layers to unfreeze l^*_{lang} :

$$l^*_{lang} = \arg\max_l \bar{D}\left(l, l+1\right) \tag{7}$$

This multilingual adaptive layer unfreezing strategy is expected to further enhance the performance and generalization ability when performing translation tasks between different target languages.

3. Neurocontextual semantic analysis of multilingual to semantic expressions

3.1. Dual-code binding

In the monolingual-to-semantic-expression encoding-decoding framework, the number of network layers for both the encoder and the decoder is 3. First, the encoder takes a natural language sentence from the source as an input and outputs its corresponding implicit-layer state which is passed to the decoder as an input to the decoder immediately afterward, and there exists an attentional model between the encoder and the decoder.

A general RNN reads the input sequence $x=(x_1,\ldots,x_m)$ in front-to-back order, and the encoder reads both in front-to-back and back-to-front order. By running the RNN in both directions, a word x in the sequence contains two implicit layer states in both directions, forward and backward, which are finally spliced together to obtain the encoder output. This output includes both the information of the word in front of x and the information of the word behind.

Based on the above encoding and decoding framework, this paper adds an encoder so that each source language has its own corresponding encoder and shares a decoder. x and z denote the two different inputs at the source, y denotes the output at the target, hx and hz denote the implicit layer states of the two encoders, and s_i denotes the implicit layer state of the decoder at the current moment. The two encoders correspond to two attention scores cx and cz, and the two attention scores are combined to obtain c_i, which is used as an input to compute the implicit layer state of the decoder at the current moment.

The initial implied layer state of the decoding side is set as follows, after obtaining the outputs hx and hz of each encoder, hx and hz are spliced according to the last dimension of the matrix, and then the mean value is calculated according to the first dimension of the spliced matrix to obtain h. The next step is to perform a linear transformation on h using W_c, and then use a $\tanh$ nonlinear transformation on the linear transformation result to finally obtain the initial implicit layer state value s_0 at the decoding end, which is given in the following equation:

$$hx=\begin{bmatrix}\overrightarrow{hx_1} & \overrightarrow{hx_2} & \cdots & \overrightarrow{hx_m}\\ \overleftarrow{hx_1} & \overleftarrow{hx_2} & \cdots & \overleftarrow{hx_m}\end{bmatrix} \tag{8}$$

$$hz=\begin{bmatrix}\overrightarrow{hz_1} & \overrightarrow{hz_2} & \cdots & \overrightarrow{hz_n}\\ \overleftarrow{hz_1} & \overleftarrow{hz_2} & \cdots & \overleftarrow{hz_n}\end{bmatrix} \tag{9}$$

$$h=mean([hx;hz]) \tag{10}$$

$$s_0=\tanh\left(W_c\cdot h\right) \tag{11}$$

3.2. Attention Modeling for Multilingual to Semantic Expressions

In the monolingual-to-semantic-expression attention model, the moment $i-1$ implicit layer state s_{i-1}, the previous word embedding y_{i-1}, from the top level of the decoder, and the current moment attention score c_i are combined with each other to obtain the moment i implicit layer state as follows:

$$s_i=f\left(s_{i-1},y_{i-1},c_i\right) \tag{12}$$

The computation of the text vector c_i relies on $(h_1,\ldots h_m)$ sequences, which are mapped by the encoder from the input sentence, where each h_j contains information about the whole input sentence, mainly containing information about the words near the j th word in the sentence. These sequences are then weighted and summed to finally obtain c_i, which is given by the following formula:

$$c_i = \sum_{j=1}^{m} \alpha_{ij} h_j \tag{13}$$

The weight α_{ij} of each h_j in the above equation is calculated as follows:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{m} \exp(e_{ik})} \tag{14}$$

Among them:

$$e_{ij} = a(s_{i-1}, h_j) \tag{15}$$

Equation a represents the alignment model, which uses a score to reflect the degree of alignment between the words near position j in the input sentence and the output at position i. The computation of this score depends on the implicit layer state s_{i-1} and the j th sequence of the input sentence h_j. It can be seen that the implicit layer state of the previous moment is essential in computing the implicit layer state at the current moment with the input sequence h_j.

With the introduction of the attention mechanism in the model, the decoder only needs to focus on part of the information of the input sentence, which is more targeted. In order to be able to process the information from both encoders at the source end simultaneously, the attention model is modified in this paper. Referring to the formula for attention, for each of the two encoders, two text vectors are computed in this paper to correspond to them, which are cx_i and cz_i, respectively, in place of c_i in the monolingual attention model:

$$s_i = f(s_{i-1}, y_{i-1}, cx_i, cz_i) \tag{16}$$

Thus, the model proposed in this paper contains two independent sets of soft alignment sets and two text vectors cx_i and cz_i. Also the parameters in the two encoding models in the framework are completely independent of each other.

It is worth noting that while the dual encoding one decoding model is only for the scenario where two languages are used as inputs at the same time, the model in this paper can easily be extended to situations where three or more languages are used as inputs. On top of the dual-encoding-one-decoding model, each additional language at the source side corresponds to an additional encoder, at which point the multi-encoder model generates multiple encoder outputs and multiple attentional score values. Then the outputs of each encoder are combined to obtain h, which is used as the input for calculating the initial state value of the decoding side, and the results returned by each attention model are added to the f function, which is used as the input for calculating the state of the implicit layer of the decoding side at the current moment, and ultimately realize the neural-semantic analysis of the multilingual-to-semantic expressions.

4. Performance analysis of the large language model and its fine-tuning methodology

Tables 1 and 2 show the COMET and SacreBLEU scores of this paper's model for different test sets of X-English and English-X directions, respectively. For SacreBLEU, the default “13a” lexer is used for all language translations except Chinese translations which use the “zh” lexer.8 Compared with the uncorrected method, the proposed method improves COMET by an average of 1.24 points in the three X-English language orientations, and increases COMET by an average of 1.24 points in the BLEU. COMET improved by an average of 1.24 points and BLEU by 0.75 points, to a slightly lesser extent than COMET.For the three language directions of English-X, both COMET and BLEU scores improved more significantly, with an average improvement of 3.22 points for COMET and 2.58 points for BLEU. The experimental results indicate that the translations with error-prone word correction improve the quality of translation compared to the original translations, especially in languages with a lower percentage of pre-trained corpus where the performance improvement is more significant. Compared with other comparison models, the translation performance of the translation fine-tuning method exceeds that of the methods with the same parameter scale model, and in the X-English direction, this paper's method is slightly ahead of the better-performing Llama2-13B, and in the English-X language direction, the fine-tuning method is ahead of almost all the baseline methods, demonstrating the effectiveness of the translation of this paper's method, even though the performance of the original translations lags behind that of the baseline comparison model.

Table 1. COMET scores and SacreBLEU scores of X-English test set (%)

Method	Chinese-English		German-English		Russian-English		Average	
	COMET	BLEU	COMET	BLEU	COMET	BLEU	COMET	BLEU
ParroT-7B	75.94	20.16	83.02	29.76	--	--	--	--
SWIE-7B	76.48	21.26	83.02	30.46	--	--	--	--
Llama2-7B	75.17	18.71	82.71	30.5	82.8	35.63	80.23	28.28
ParroT-13B	76.73	21.69	83.63	31.3	--	--	--	--
BigTranslate-13B	74.28	14.18	80.73	23.33	77.81	26.83	77.61	21.45
Llama2-13B	78.09	21.79	83.02	31.03	82.91	36.49	81.34	29.77
This method	78.91	23.37	83.74	31.7	83.66	36.49	82.10	30.52

Table 2. COMET scores and SacreBLEU scores of English-X test set (%)

Method	English-Chinese		English-German		English-Russian		Average	
	COMET	BLEU	COMET	BLEU	COMET	BLEU	COMET	BLEU
ParroT-7B	80.33	30.3	81.6	26.1	--	--	--	--
SWIE-7B	80.59	31.2	82.38	27.2	--	--	--	--
Llama2-7B	71.96	17.19	76.48	19.04	73.33	16.13	73.92	17.45
ParroT-13B	81.05	31.66	82.59	28.09	--	--	--	--
BigTranslate-13B	81.31	28.55	78.82	21.43	78.19	17.66	79.44	22.55
Llama2-13B	79.69	30.02	75.57	13.69	63.86	0.6	73.04	14.77
This method	81.43	28.66	83.06	25.33	83.48	21.4	82.66	25.13

We first focus on the mitigation of off-target translation and hallucination by our proposed method for the large language model, which we evaluate and analyze from two aspects using the off-target translation rate (OTR) and the proportion of sentences with chrF2<10 in the test set, respectively.The results for the X-English and English-X language directions are shown in Tables 3 and 4, respectively.The off-target translation problem is dominant in both the off-target translation and hallucination, and the all the language directions of the The off-target rate and the proportion of sentences with chrF2<10 decreased, with the average off-target rate for X-English decreasing from 10.41% to 1.54% and the average proportion of sentences with chrF2<10 decreasing from 7.55% to 0.81%. The average off-target rate in the three language directions of English-X was reduced from 37.85% to 2.77%, and the average proportion of chrF2<10 was reduced from 26.61% to 5.89%, among which the off-target translations, which were the most serious, were reduced from 52.92% to 4.37% in English-Chinese. The experimental results demonstrated the effective suppression of off-target translation and hallucinations by the fine-tuning method.

Table 3. OTR and translations with chrF2<10 of X-English test set (%)

Method	Chinese-English		German-English		Russian-English		Average	
	OTR	chrF2<10	OTR	chrF2<10	OTR	chrF2<10	OTR	chrF2<10
Llama2-7B	23.85	21.48	3.44	0.50	3.94	0.66	10.41	7.55
This method	1.58	1.71	0.81	0.45	2.24	0.28	1.54	0.81

Table 4. OTR and translations with chrF2<10 of English-X test set (%)

Method	English-Chinese		English-German		English-Russian		Average	
	OTR	chrF2<10	OTR	chrF2<10	OTR	chrF2<10	OTR	chrF2<10
Llama2-7B	52.92	49.41	29.76	2	30.88	28.41	37.85	26.61
This method	4.37	13.96	1.22	0.59	2.72	3.11	2.77	5.89

The best corrective positions of the translation error-prone words on the Chinese-English test set were statistically analyzed. Figure 2 shows the distribution of the length of the translations on the test set. The length of the translations is widely distributed (sentences exceeding 100 words are not shown in the figure), mainly in the range of 1-80 words, with an average of 28 words per sentence.

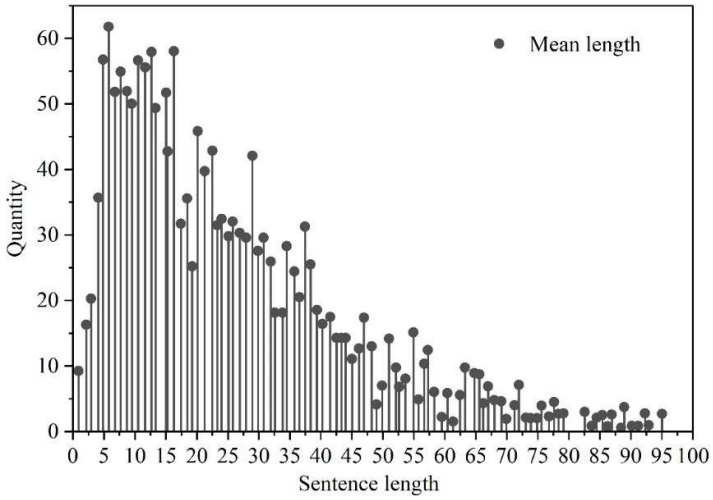

Fig. 2. Translation length distribution

Figure 3 shows the statistics of the proportion of the number of times each position of the translated text in the test set was corrected to the total number of times it was corrected. The higher the corrected position was, the higher the proportion of corrections made, and the proportion of corrections made when the first word of the translated text was an error-prone word amounted to 23.5%, which is about twice as much as the proportion of corrections made for the second word. The results show that correcting the error-prone words in the front of the translation has a greater impact on the overall decoding results, and the translation is generally best corrected in a more forward position.

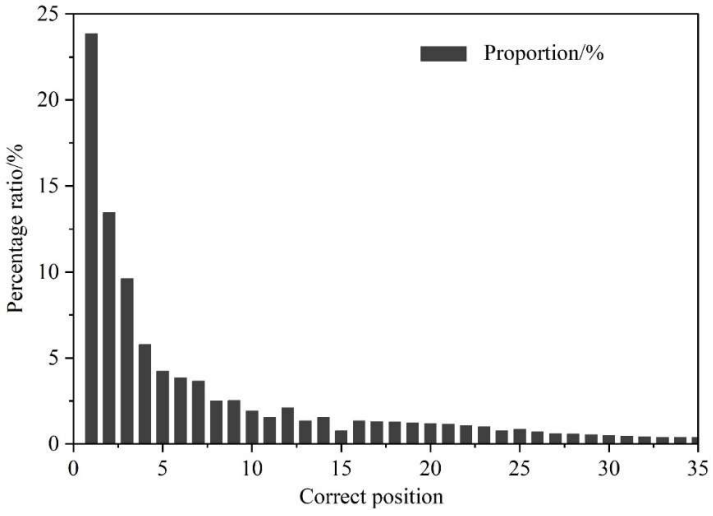

Fig. 3. The number of corrective times is the percentage of total corrected times

Figure 4 shows the cumulative percentage of each correction position. When the error-prone word correction mechanism is used to judge only the first 5 words of the translation whether they are error-prone words or not, it can improve the quality of the translation by about 40%. When only the first 15 words are judged to be error-prone words, the quality of translation can be improved by about 80%. When judging the first 30 words, which is close to the average length of the translation, it can improve the quality of 95% of the translation. The results show that the larger the window for the error-prone word correction mechanism to search for the best correction position, the higher the probability of finding a better translation, but the gain achieved gradually decreases, and when the size of the search window reaches the average length of the translation, it is almost possible to find the best correction position of the translation.

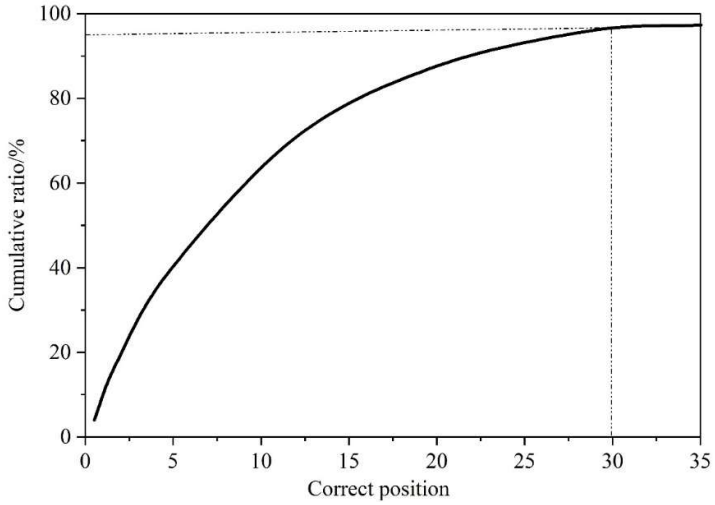

Fig. 4. The cumulative proportion of each corrective position

5. Results of the semantic analysis

According to the different classification granularity, the inter-sentence relations of HIT-CDTB's chapters can be classified into 4, 19, 32 and 38 categories, i.e., the data can be classified into 4 categories according to the main categories, 19 categories according to the second level categories, and so on. Considering the data imbalance problem of the model, it is necessary to highlight the imbalance phenomenon in order to test the model's effectiveness in solving the data imbalance problem, but also not to make the model too much affected by the imbalance problem, so the classification results of the 19 categories are used as an example for the analysis of the model's effectiveness. Table 5 shows the experimental results of 19 classes, the calculation time is the time of one batch of model training, and the test results are evaluated by accuracy, precision, recall and F1 value.

Among all the neural network model structures tested in the experiments, the attention model proposed in this paper performs the best overall, with an improvement of 11.60% in F1 value compared to the earlier long and short-term memory network model with one hidden layer, and the model results are stronger than the recurrent neural network baseline model. The recurrent neural network baseline model uses the attention mechanism to compute the feature focus, so it is stronger than the long-short-term memory network model without the attention mechanism. It can be seen that the attention mechanism has an absolute advantage in dealing with the task of analyzing the semantic relations of long text and can improve the overall performance of the model.

The recall indicator reflects whether the model emphasizes all the category features in a balanced way, and the overall recall of the model tends to be lower when there is an imbalance in the distribution of categories in the dataset. The recall values of the three baseline models are close to each other, indicating that the baseline model does not have a strong ability to deal with the data imbalance problem. The recall of this paper's model is significantly better than that of the baseline model, indicating that the improved loss function in the model effectively mitigates the problem of data category distribution imbalance, and this improved method is also applicable in other category-imbalanced business English text analysis tasks.

Meanwhile, it can be concluded from the computation time of all the experimental models that although the model structure of this paper's model is more complex, the computation time consumed is comparable to that of other complex models. This is because the model in this paper takes a variety of measures in the training process to avoid consuming too much computational resources, such as the setting of the loss rate, and the batch size setting directly affects the model training time. However, compared to convolutional neural networks and other single model structures, recurrent neural networks do consume a large amount of arithmetic power during the training process, and how to reduce the computational complexity is a problem that needs to be solved for hybrid models.

Table 5. HIT-CDTB of 19 experiment

Model	Calculate time/s	Accuracy /%	Precision/%	Recall rate/%	F1/%
Rutherford	6.3	64.58	67.08	61.25	64.24
Ronnqvist	9.9	69.89	76.59	60.59	68.15
NNMA-BERT	16.5	70.25	79.19	60.11	69.12
CNN-BERT	15.9	75.36	79.48	68.97	74.56
This model	15.2	76.08	80.59	72.15	75.84

The performance of the model in this paper was also tested on the CDTB dataset and compared with the baseline. In this experiment, we only tested the "interpretation", "connection", "extension" and "causality" relationships according to the data selection method of the baseline model on the CDTB dataset, which are the four most common relationships in the CDTB test set, and the data volume of other relationships is not enough for classification. The experiments report the accuracy, micro-averaged F1 values and macro-averaged F1 values of the model test for one batch of time of model training, and the experimental results are shown in Table 6. The model's nine-class classification accuracy on CDTB is reported in the corresponding literature on adaptive convolutional methods, and their experimental results are displayed as a reference in Table 6.

On the CDTB dataset, the micro-averaged F1 value of this paper's model is slightly better than the other baseline models and 1.09% higher than the NNMA-BERT model, but the advantage is not obvious. This is because when performing fine-grained hierarchical classification, the tree structure can calculate the match between business English text features and categories layer by layer, but when there are only four categories, the tree structure doesn't work, and the advantage of the model is only generated by the introduction of the BERT pre-training method. This problem can also be reflected from the value of F1, which is indirectly calculated from the recall, so both F1 and recall reflect the importance of the model to all categories, because the recognition task of the CDTB dataset on the four categories has almost no category imbalance in the data, so the distinction between the F1 values of several models is not obvious.

Table 6. The results of the results of the CDTB classification

Model	Calculate time/s	Accuracy /%	Microscopic mean F1/%	Macroscopic mean F1/%
Rutherford	-	70.45	-	-

Ronnqvist	-	73.21	65.32	53.45
NNMA-BERT	-	-	68.05	54.46
CNN-BERT	-	67.25	-	-
This model	13.5	75.36	69.14	54.23

Unlike English, Chinese does not have explicit marking of tenses and clauses, the relationship between sentences is conveyed through semantics, and the associative words between Chinese sentences are more flexible and diverse, so the recognition of implicit inter-sentence relations in Chinese is correspondingly difficult.

The model in this paper performs well in the process of business English translation test, but there are many errors in the recognition of fine-grained inter-sentence relationships in Chinese semantic analysis experiments, the model judges correctly in the process of identifying inter-sentence semantic relations in Chinese, but the fine-grained classification is wrong, according to the sample analysis of different categories, due to the large sample size in the "explanation" category, the model attaches importance to the characteristics of this category, and the relationship between two Chinese sentences is not easy to distinguish even for humans. In this paper, the model pays more attention to large categories than small categories, and the problem of category imbalance is not completely solved, which affects the model's analysis of the semantics of Chinese text. Therefore, the application of the model in Chinese semantic analysis needs to be further improved and explored.

6. Conclusion

Aiming at the difficulties of text context analysis in business English translation, this paper proposes a text context analysis model and its fine-tuning method based on the large language model.

The experimental results of the performance test show that the fine-tuning method of the task-related layer in this paper can improve the text translation quality of the big language model, and the COMET and BLEU of this paper's model in the three linguistic directions of X-English are improved by 1.24 and 0.75 points respectively. In the three language directions of English-X, COMET and BLEU improved by 3.22 and 2.58 scores on average respectively, indicating that this paper's model can not only show some potential in the business English translation task, but also fulfill the translation tasks of many different languages.

In this paper, we conducted several sets of comparative experiments on the business English contextual relationship analysis model based on the attention mechanism, and the model proposed in this paper had the best overall performance on the HIT-CDTB dataset, and the F1 value of this paper's model was improved by 11.60% compared with that of Rutherford's model. On the CDTB dataset, the micro-average F1 value of this paper's model improves by 1.09% compared with the NNMA-BERT model. According to the performance of this paper's model on the above two datasets, it shows that the structure of this paper's model is well-designed and has a good ability to analyze the context of business English. Although this paper's model performs better on the task of business English translation context analysis, the problem of category imbalance has not been completely solved, which affects this paper's model in analyzing the more complex Chinese text context. Therefore, there is still some room for improvement in the application of this model in Chinese context analysis.

Funding

This work was supported by Research on the Development Path of "Dual-element" Textbooks for Higher Vocational English in Accordance with the New Standards - Taking "Business English Communication and Customer Service" as an Example (No. WYJZW-2023IN0002), which is sponsored by the Teaching Steering Committee for Foreign Language Majors in Vocational Colleges of the Ministry of Education.

References

[1] Chidlow, A., Plakoyiannaki, E., & Welch, C. (2014). Translation in cross-language international business research: Beyond equivalence. *Journal of International Business Studies*, 45, 562-582.

[2] Almashhadani, M., & Almashhadani, H. A. (2023). English Translations in Project Management: Enhancing Cross-Cultural Communication and Project Success. *International Journal of Business and Management Invention*, 12(6), 291-297.

[3] Han, S. (2023). Research on Business English Translation Strategies under the Background of Cultural Differences. *International Journal of Education and Humanities*, 7(2), 130-133.

[4] Yuanyuan, C. (2022). Translation Methods of Business English in Cross-Cultural Context. *Academic Journal of Humanities & Social Sciences*, 5(1), 70-72.

[5] Liwen, Y. (2015, September). Research on the Business English Translation Methodology under the Perspective of Dynamic Cultural Equivalence. *In 2015 Conference on Informatization in Education, Management and Business* (IEMB-15) (pp. 519-523). Atlantis Press.

[6] Liao, M. H., Castillo Ortiz, P. J., Napier, J., & Newill, K. (2023). Translation and Interpretation in Cross-Cultural Business Workplace Practices. *In Intercultural Issues in the Workplace: Leadership, Communication and Trust* (pp. 193-208). Cham: Springer International Publishing.

[7] Shi, L., & Liu, Y. (2021). A study on equivalence between English translation and cross-cultural communication based on translation coordination theory. *Revista Argentina de Clínica Psicológica*, 30(1), 283.

[8] Chan, C. (2019). Stylistic Characteristics and Translation Methods of Cross-Cultural Business English. *Frontiers in Educational Research*, 2(12).

[9] Tang, Y. (2018, December). Research on Cross-cultural Factors in Business English Translation. *In 2018 3rd International Conference on Education, E-learning and Management Technology* (EEMT 2018) (pp. 525-529). Atlantis Press.

[10] Hu, W. (2021). The application of cross-cultural awareness in business English translation. *International Journal of Social Sciences in Universities*, 4(3).

[11] Gao, L. (2018). A study on the application of functional equivalence to business English EC translation. *Theory and Practice in Language Studies*, 8(7), 759-765.

[12] Chen, Z. (2024). Navigating Cross‐Cultural Challenges in Business English Translation: Strategies for Optimization. *International Journal of Social Science and Education Research*, 7(6), 41-47.

[13] Xinran, Z., & Jingwen, Z. (2023). Analyzing Business English Translation Strategies in the Context of New Media from a Cross-cultural Pragmatics Perspective. *International Journal of New Developments in Education*, 5(14).

[14] Yu, G. (2022). The transformation of cross-cultural perspective and the embodiment of translation skills in English translation. *Journal of Education and Educational Research*, 1(2), 43-45.

[15] Parizoska, J., & Rajh, I. (2017). Idiom variation in business English textbooks: A corpus-based study. *ESP Today*, 5(1), 46-67.

[16] Jaworska, S. (2017). 24 Corpora and corpus linguistic approaches to studying business language. *Handbooks of Applied Linguistics*, 583.

[17] Xu, D. (2021, December). Construction of Business English Writing Corpus Based on Data Management System. *In 2021 International Conference on Aviation Safety and Information Technology* (pp. 505-509).

[18] Feng, H., Crezee, I., & Grant, L. (2018). Form and meaning in collocations: a corpus-driven study on translation universals in Chinese-to-English business translation. *Perspectives*, 26(5), 677-690.

[19] Rezvani Kalajahi, S. A., Neufeld, S., & Nadzimah Abdullah, A. (2017). The discourse connector list: a multi-genre cross-cultural corpus analysis. *Text & Talk*, 37(3), 283-310.

[20] Li, Z., & Tang, J. (2022). Research on Cross‐cultural Text Reconstruction of Urban Publicity Translation Based on Computer Corpus. *Scientific Programming*, 2022(1), 5076637.

[21] Liu, J., & Chu, J. (2020). A Corpus Based Study on Cross Cultural Adaption of Websites of Foreign-Invested Enterprises in Heilongjiang Province. *Open Journal of Social Sciences*, 8(3), 77-86.

[22] Castagnoli, S., & Elena, M. (2019). Translating (im) personalisation in corporate discourse. A corpus-based analysis of Corporate Social Responsibility reports in English and Italian. *Lingue e Linguaggi*, 29, 205-224.

[23] Hu, C., Zhao, Z., & Lu, C. (2024). Metadiscursive nouns in corporate communication: A cross-cultural study of CEO letters in the US and Chinese corporate social responsibility reports. *English for Specific Purposes*, 76, 28-40.

[24] Aahil Khambhawala,Chi Ho Lee,Silabrata Pahari,Paul Nancarrow,Nabil Abdel Jabbar,Mahmoud M. El Halwagi & Joseph Sang Il Kwon. (2025). Advanced transformer models for structure-property relationship predictions of ionic liquid melting points. Chemical Engineering Journal158578-158578.

[25] Lu Zhou,Yiheng Chen,Xinmin Li,Yanan Li,Ning Li,Xiting Wang & Rui Zhang. (2024). A New Adapter Tuning of Large Language Model for Chinese Medical Named Entity Recognition. Applied Artificial Intelligence(1),

[26] Mozhgan Ghassemiazghandi. (2024). An Evaluation of ChatGPT's Translation Accuracy Using BLEU Score. Theory and Practice in Language Studies(4),985-994.