International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-12

https://doi.org/10.70917/ijcisim-2026-3135
Article

A Study on the Cross-cultural Conversion Mechanism and Compensation Strategies of Conceptual Metaphor in the English Translation of Chinese Classical Literature

Guannan Xiao [1], Florence Kuek [1,*] and Pik Shy FAN [1]

[1] Universiti Malaya, Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, 50603, Malaysia
* Correspondence author: Nancyflo@163.com

Abstract: This paper proposes an end-to-end metaphorical vocabulary recognition model named CBAM, which combines character-based and vocabulary-based representation methods to automatically obtain lexical information. The BiLSTM network is used to capture the context semantics of the target word, and the multi-head attention mechanism is introduced to enhance the model's perception of metaphors. The similarity network and Softmax classifier are used to determine the metaphorical nature of the vocabulary. In the translation task, an English-to-English neural machine translation method for metaphorical vocabulary is proposed, integrating the recognized metaphorical vocabulary information into the model's encoder to help the model learn deep semantic information. Through the gating mechanism, the translation information of the vocabulary is controlled for input. Experimental results show that the metaphorical vocabulary recognition model outperforms the best comparison model in all parts of speech and in the case of only verbs, with an F-score increase of 5.0% and 5.5% respectively. The classical literature English-to-English neural machine translation model achieved a BLEU4 evaluation index value of 45.78%. This paper further summarizes the compensation strategies for the conceptual metaphors in the translation of Chinese classical literature, providing methodological references for the cross-language dissemination of classical literature.

Keywords: CBAM; Neural Machine Translation; Metaphorical Vocabulary Recognition; Compensation Strategies; Chinese Classical Literature English Translation

1. Introduction

In the global context, the interest of the English-speaking world in Chinese classical literature has never waned. Whether it is poetry, prose, historical narratives or drama texts, the task of translation is not merely to transfer the literal meaning to another language; rather, it is on the crest of the waves of transcending time, culture, and aesthetic traditions, allowing readers to experience the temperament, emotions and thoughts of the original text in a different language environment [1-4].

However, translating Chinese classical literature into English is by no means an easy task. It faces numerous challenges. There are significant differences between the grammatical structures, lexical semantics, and expression methods of the Chinese and English languages. Chinese emphasizes semantic coherence, with relatively loose sentence structures, expressing complete thoughts through the interplay of words and sentences; while English emphasizes syntactic coherence, emphasizing the rigor and logic of sentence structures, and using various conjunctions and grammatical rules to construct sentences [5-8]. Chinese words often carry rich cultural connotations and multiple meanings. A simple word may contain profound historical and cultural backgrounds as well as multiple semantic layers, and it is difficult to find an exact equivalent word in English for translation [9-11]. The unique rhythm and cadence of classical literature, such as the even and uneven tones, rhyming, etc., are difficult to be perfectly replicated in English. These differences bring great difficulties to the translation of Chinese poetry into

English. How to accurately convey this vague beauty in the English translation becomes one of the key issues in the translation of Chinese poetry into English [12-14]. In this context, the research on the cross-cultural transformation mechanism and compensation strategies of conceptual metaphors in the English translation of Chinese classical literature has become increasingly significant, which can further promote the development of translation theory towards a more diversified and refined direction.

To effectively identify metaphorical expressions in classical literature, this paper proposes a metaphor detection model. This model obtains character word embeddings by integrating ELMo word vectors and GloVe word vectors, inputs the embedding information into the model, and uses deep neural networks to mine the contextual relationships between words, uses similarity networks and Softmax classifiers to detect the differences between the literal meaning and the contextual meaning of the target word, in order to achieve metaphor recognition. The neural machine translation model incorporates the metaphorical vocabulary information from the source language side and improves the encoder. The gating mechanism is introduced to control the entry of word vector dimensions, giving the model the right to autonomously select information. Three types of compensation strategies including literal translation, revision translation and interpretive translation are summarized. Comparative experiments are designed to verify the usability of the metaphor detection model and the neural machine translation model respectively.

2. Identification of metaphorical words incorporating character information

Metaphor is a product of conceptualization, and its essence is that an experience in one cognitive domain is used to understand an experience in another cognitive domain. In the cross-cultural translation process of classical Chinese literature into English, the identification of metaphorical words is the prerequisite and cognitive foundation for the smooth conversion of conceptual metaphors.

2.1. Character-based Word Representation

The first module of the CBAM model is the embedding layer, which consists of multiple word embedding representations. Firstly, the ELMo word vectors and GloVe word vectors are concatenated and then further processed [15]. This approach has achieved good performance in multiple natural language processing tasks. However, in order for the model to learn explicit lexical, syntactic, and orthographic information, which is necessary for better performing metaphor tasks, this paper adds character-level word embedding representations. A one-dimensional convolutional neural network is used to calculate the character-level word representation information. Assuming that the word at position t is composed of characters $\left[c_1, c_2, \cdots, c_n\right]$, where l is the length of the word and d is the dimension of the character embedding, and it does not exceed the character vocabulary. Let $C^t \in R^{d\times l}$ represent the character-level embedding matrix of word t, which is formed by convolving a filter $H \in R^{d\times w}$ with width w.

$$f^t = \tanh\left(C^t * H + b\right), f^t \in R^{l-w+1} \tag{1}$$

Next, using the maximum pooling based on the length of f, an output of the filter is obtained.

$$y^t = \max_{1\le j\le l-w+1}\left\{f_j^t\right\} \tag{2}$$

Using multiple filters of different widths and connecting the outputs of each filter together, the character-level word vector representation of the word t is finally obtained. Assuming that h is the number of filters, $y_1, y_2, \cdots, y_n$ are the outputs of the filters, then the character-level word vector representation of the word t is $c_t = \left[y_1^t, y_2^t, \cdots, y_n^t\right]$. This paper uses a total of 4 filters, with input dimensions of all 50, output dimensions of 25, 50, 75, and 100 respectively, and convolution kernel sizes of 1, 2, 3, and 4. Finally, a 250-dimensional character-level word vector representation is obtained. The GloVe word vector g_t and c_t are connected and run in a single-layer Highway network.

$$a_t = \left[g_t, c_t\right] \tag{3}$$

$$t = \sigma\left(W_T a_t + b_T\right) \tag{4}$$

$$z_t = t \otimes g\left(W_H a_t + b_H\right) + \left(1-t\right) \otimes a_t \tag{5}$$

Among them, z_t and a_t have the same dimension in construction, W_H and W_T are square matrices, t is the transformation gate, and $(1-t)$ is called the carry gate. The Highway network is used to select the dimensions that are modified and directly passed to the output. Therefore, this network can not only connect GloVe word vectors and character-based word vectors, but also connect other word vectors such as ELMo and visual information. By adjusting the vector dimensions of the Highway network connections, it can better learn context semantics and lexical information. This paper conducted experiments on the effect of Highway network connecting different word vectors, but the method of connecting GloVe word vectors and character-based word vectors requires fewer parameters and performs better.

2.2. Bidirectional Long Short-Term Memory Network

The second module in the CBAM model is the BiLSTM network layer [16]. The nodes between the hidden layers of the recurrent neural network are all interconnected and interact with each other. The output of the hidden layer in the previous time period and the input of the entire embedding layer constitute the input of the hidden layer at a specific moment. That is, in this task, the hidden state of the target word is calculated based on the hidden states of its adjacent words (the hidden states of the words before and after the target word need to be calculated independently).

Metaphors are usually formed through subject-verb-object combinations, so bidirectional long short-term memory networks are of great help in better obtaining the contextual semantics of the target word. In the CBAM model, the input of the Bi LSTM network layer is the connection matrix of multiple word vectors. The BiLSTM network layer calculates the forward and backward hidden layer information through training. The forward $\overrightarrow{LSTM}$ generates a set of hidden states from left to right $\vec{h}=\left\{\vec{h}_1,\vec{h}_2,\cdots,\vec{h}_n\right\}$, while the backward $\overleftarrow{LSTM}$ also generates a set of hidden states from right to left $\overleftarrow{h}=\left\{\overleftarrow{h}_1,\overleftarrow{h}_2,\cdots,\overleftarrow{h}_n\right\}$. Finally, the hidden state of each word is obtained by encoding its surrounding context and the word itself:

$$\vec{h}=\overrightarrow{LSTM}\left(\left[c_1,c_2,\cdots,c_n\right]\right) \tag{6}$$

$$\overleftarrow{h}=\overleftarrow{LSTM}\left(\left[c_1,c_2,\cdots,c_n\right]\right) \tag{7}$$

$$h_t=f_{BiLSTM}\left(W_t,\vec{h}_{t-1},\overleftarrow{h}_{t+1}\right) \tag{8}$$

Here, W_t represents the word embedding information of the word.

2.3. Multi-head Attention Mechanism

This paper employs a multi-head context attention mechanism to calculate the context information of the target word [17]. Unlike the multi-head self-attention mechanism, the multi-head context attention mechanism computes the context representation by setting a window of size n to attend to the target word. This network layer takes the hidden state of the target word as input and outputs its context hidden state. The context representation based on multi-head attention is the multi-head attention context representation with a window size of 3.

2.4. Similar Networks

In previous work, the weighted cosine similarity network was used to determine the similarity between two word vectors in a phrase. Therefore, in this paper, metaphor is identified by calculating the similarity between the literal semantics and the contextual semantics. Firstly, all words are embedded and projected into the same dimension as the hidden layer of the BiLSTM.

$$x_t=\left[z_t,e_t,v_t,p_t\right] \tag{9}$$

$$\tilde{x}_t=\tanh\left(W_z x_t\right) \tag{10}$$

This step has two purposes: (1) to reduce the size in order to increase the computing speed of the model; (2) to perform vector space mapping. Since the word embedding information of each input has different dimensions, this method enables the network to learn more vector information related to metaphors. Next, the element-wise multiplication is used to connect $\tilde{x}_t$ and h_t:

$$m_t = \tilde{x}_t \otimes h_t \tag{11}$$

m_t serves as the input for the next network layer:

$$u_t = \tanh\left(W_u m_t\right) \tag{12}$$

Next, u_t is used as the input for the Softmax classifier to predict the metaphorical nature of the target word.

$$p\left(\tilde{y}_t \middle| u_t\right) = \sigma\left(W_y u_t + b\right) \tag{13}$$

2.5. Experimental Results and Analysis

Based on previous studies, this paper evaluates the proposed model using recall rate, precision rate and the F value. The F value is the harmonic mean of recall rate and precision rate, calculated as $F = 2 \times P \times R / (P + R)$. For the metaphor detection dataset, this paper has built a Chinese classical literature English translation metaphor annotation dataset (CCLE-Metaphor), which selects classic English translations of "Dream of the Red Chamber", "The Analects of Confucius" and Tang poetry. Two annotators with a background in cognitive linguistics independently annotated the conceptual metaphors in the original text and followed the corresponding processing methods in the English translations. Finally, a parallel corpus containing approximately 5,000 annotated samples was formed. The dataset is classified by text type into poetry, novels, essays and operas. This paper also reports the performance of the model in the four types (poetry, novels, essays and operas) of data.

Based on previous metaphor detection studies, this paper evaluates the performance of the multi-attention model on the CCLE-Metaphor metaphor detection dataset and trains and tests the model using the training set and test set respectively. Additionally, this paper also verifies the experimental results of the model in all word classes and only with verbs.

Five recent metaphor detection models were adopted as the baseline for the experiment: Di-LSTM, CNN-BiLSTM, BiLSTM-CRF, WordNet-CRF, and Token-Level.

The detailed introduction of the models is as follows:

(1) Di-LSTM: A deep neural structure Di-LSTM contrast network for metaphor detection was proposed, and the experimental results verified that the contrast features are helpful for identifying metaphors.

(2) CNN-BiLSTM: A metaphor detection method combining CNN and BiLSTM was proposed, extracting local features of words and context features respectively.

(3) BiLSTM-CRF: A metaphor detection method based on a hybrid model of BiLSTM and CRF was proposed, and this method uses fewer features than the state-of-the-art sequence model.

(4) WordNet-CRF: A method using a CRF-based classifier to detect metaphors in the text was proposed, and this method effectively improves the recall rate of metaphor detection by using WordNet.

(5) Token-Level: A new supervised metaphor detection method was proposed, combining pre-trained word embeddings and neural networks to identify text metaphors.

This paper compares the performance of the proposed models with the baselines on the test dataset, and also uses the verb metaphor data with the largest quantity in the dataset and reports the experimental results of the model on verbs.

The experimental results are shown in Table 1. The experimental results of the metaphor recognition model integrating character information of this paper outperforms the baseline model. Compared with the best baseline in all word classes and only with verbs, the proposed model's F-score increases by 5.0% and 5.5% respectively. The experimental results show that the multi-head attention model can significantly improve the performance of metaphor detection. Compared with previous methods, the proposed model can more fully extract text features. This method provides a new perspective for feature extraction. Overall, these experimental results show that the proposed model outperforms the baseline model and proves the effectiveness of the proposed model.

Table 1. Experimental results

#Model	# All words/%			# Verbs/%		
	Precision	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**
Di-LSTM	51.2	64.2	57.0	52.8	70.7	60.5
CNN-BiLSTM	60.6	70.1	65.0	60.2	76.4	67.4
BiLSTM-CRF	64.3	45.7	53.4	67.4	51.6	58.5
WordNet-CRF	63.4	58.5	60.9	-	-	-
Token-Level	58.8	53.4	56.0	-	-	-
This method	67.9	71.8	70.0	70.1	75.9	72.9

To further illustrate the effectiveness of the model proposed in this paper, this section reports the performance of the model on four types of data.

The experimental results are shown in Table 2. The model proposed in this paper achieved good results on all four types of data sets. Specifically, the model proposed in this paper achieved the best performance on the prose data (with an F-score of 76.55%). This is because prose data has a more formal writing style, stronger logic, and more complete expression, making it easier to identify metaphors.

The recognition performance for poetry data was the lowest (with an F-score of 62.62%), because the language of poetry is highly condensed, with jumping imagery, extensive use of ellipsis, inversion, and unfamiliar combinations. Verb metaphors often rely on external cultural background, and there are insufficient context clues, making it difficult to identify metaphors in poetry. Although the type of poetry data has a certain impact on the model's performance, compared with the previous optimal model (Token-Level), the performance of the proposed model has significantly improved.

For novel data, the text expression is smooth and formal, making it easier to identify metaphors. Similar to prose data, the data of the type of opera has a large amount of omission, resulting in poor metaphor recognition results. The above results prove that the model proposed in this paper can better obtain the semantics of different types of data and effectively improve the performance of the metaphor detection task.

Table 2. The performance of the model on four types of data

	Model	**Precision/%**	**Recall/%**	**F-score/%**
Poetry	Di-LSTM	41.26	59.64	48.78
	CNN-BiLSTM	48.30	69.22	56.90
	Token-Leve	58.14	45.21	50.87
	This method	57.63	68.55	62.62
Novel	Di-LSTM	56.66	59.06	57.84
	CNN-BiLSTM	66.38	64.70	65.53
	Token-Leve	64.30	61.87	63.06
	This method	68.55	72.24	70.35
Prose	Di-LSTM	64.06	68.30	66.11
	CNN-BiLSTM	72.45	74.57	73.49
	Token-Leve	58.42	54.05	56.15
	This method	73.88	79.42	76.55
Opera	Di-LSTM	34.60	72.40	46.82
	CNN-BiLSTM	45.31	71.20	55.38
	Token-Leve	65.90	47.78	55.40
	This method	63.14	69.76	66.29

3. Classical Literature English Translation for Metaphorical Vocabulary by Neural Machine Translation

After identifying metaphorical vocabulary, how to accurately convert these metaphorical expressions, which embody cross-cultural cognitive tensions, into the target language is the core component of the conceptual metaphor translation mechanism in the English translation of classical literature.

This translation model is divided into three modules: the preprocessing module, the metaphorical vocabulary recognition and representation module, and the encoder-decoder module. The preprocessing module's function is to segment the input text, remove long and short sentence pairs, remove sentence pairs with excessive differences in source language and target language lengths, and remove empty texts, in order to ensure that the model can be trained normally; the metaphorical vocabulary recognition and representation module's function is to identify the metaphorical words in the source language sentences in Chinese classical literature, and then use the real meaning of these words in the context together with the corresponding target translation words to represent the metaphorical words, in order to help the model translate the metaphorical words correctly; the encoder-decoder module's function is to use the encoder to encode the source language and then use the decoder to decode and produce the target translation corresponding to the source language. This paper uses the Transformer model as the encoder-decoder, and improves the encoder based on this. The word vectors involved in the metaphorical vocabulary recognition and representation module use the Word2vec word vectors. The model is shown in Figure 1.

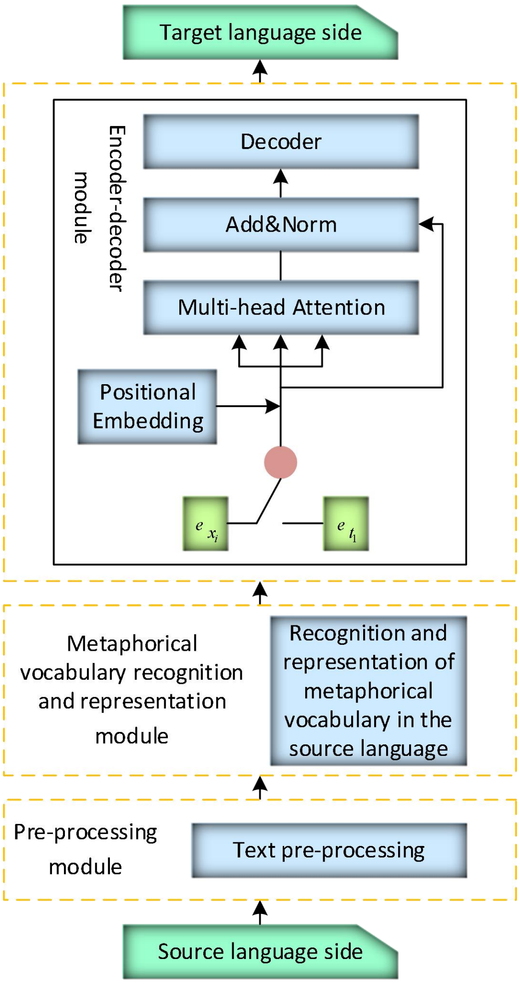

Figure 1. Machine translation model

3.1. Preprocessing Module

(1) Use the jieba word segmentation tool to segment the Chinese in the parallel corpus. Separate the words with spaces. Since English itself is divided into words by spaces, no word segmentation is needed for English.

(2) Delete the texts in the training corpus that are too long, too short, or empty to ensure the smooth progress of model training.

(3) Delete the sentence pairs in the corpus where the lengths of the source language and the target language differ too much.

3.2. Metaphorical Word Recognition and Representation Module

By integrating the metaphorical vocabulary recognition model with character information, the source language corpus with labels can be obtained, where "1" indicates that the vocabulary is a metaphorical one, and "0" indicates that it is not. The metaphorical vocabulary X_2 and its context vocabulary sequence $\{X_1, X_3, X_4, X_5 \cdots X_n\}$ can be obtained through the output of the metaphorical vocabulary recognition model. If the source language is Chinese, the relevant vocabulary sequence $\{h_1, h_2 \cdots h_n\}$ of Chinese metaphorical vocabulary is obtained from the language knowledge base HowNet; if the source language is English, the relevant vocabulary sequence $\{h_1, h_2 \cdots h_n\}$ is obtained from the English-Chinese bilingual dictionary and HowNet. Then, the average of the word vectors of the context vocabulary can be obtained to form the context representation vector $context$. The cosine similarity calculation is performed between $context$ and the word vectors of the vocabulary in the related sequence $\{h_1, h_2 \cdots h_n\}$ of the metaphorical vocabulary X_2 and its related vocabulary. The vocabulary x_i with the highest similarity score to $context$ is the meaning representation of the metaphorical vocabulary X_2 in the context. By using the vocabulary x_i and the English-Chinese bilingual dictionary, the translation t_i corresponding to x_i in the target language can be obtained. Then, the word vectors e_{x_i} and e_{t_i} corresponding to x_i and t_i are used to jointly represent the metaphorical vocabulary X_2.

3.3. Codec Module

(1) Transformer

This section briefly describes the basic encoder-decoder used by the model in this paper. The Transformer consists of an encoder and a decoder. The most distinctive feature of this neural machine translation system is based on the self-attention mechanism [18]. Unlike traditional neural machine translation models, the encoder and decoder of the Transformer are composed of a feedforward neural network and an innovative attention mechanism. The authors call this mechanism "multi-head attention mechanism". Multiple pointwise attention mechanisms can be calculated by computing the three vectors Q, K, V, which constitutes the "multi-head attention mechanism". The pointwise attention mechanism is shown in Equation (14):

$$Attention(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right) \tag{14}$$

Among them, d_k represents the dimension of the vector K. Next, the calculation results of multiple dot attention mechanisms are concatenated together and then sent into the feedforward neural network. As shown in Formulas (15) and (16):

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^v\right) \tag{15}$$

$$MultiHead(Q,K,V) = Concat(head_1 \cdots head_h)W^O \tag{16}$$

Among them, $W_i^Q \in R^{d_{model} \times d_k}$, $W_i^K \in R^{d_{model} \times d_k}$, $W_i^V \in R^{d_{model} \times d_v}$, and $W^O \in R^{hd_v \times d_{model}}$, where h represents the number of heads in the "multi-head attention mechanism".

The feedforward neural network consists of two layers of fully connected network layers, and the activation function used is RELU. The definition of the feedforward neural network is as shown in formula (17):

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{17}$$

Among them, W_1, W_2, b_1, and b_2 represent the parameters of the model.

The Transformer incorporates positional information into the encoder and decoder word vectors to address the issue of the loss of input sequence order information caused by the absence of use of recurrent and convolutional neural networks. The definition of positional information is as shown in Formulas (18) and (19):

$$PE(pos, 2i) = \sin\left(pos / 10000^{2i/d_{model}}\right) \tag{18}$$

$$PE(pos, 2i+1) = \cos\left(pos / 10000^{2i/d_{model}}\right) \tag{19}$$

Here, pos represents the position of the word in the sentence, and i represents the dimension of the vector.

(2) Integrating metaphorical vocabulary information into Transformer

To address the issue of inaccurate translation of metaphorical words by neural networks in the source sentence, this paper will use the vocabulary x_i representing the meaning of the metaphorical word in the current context and the translation t_i of x_i in the target end to jointly represent the metaphorical word. Since t_i is obtained through a bilingual dictionary, it is prone to incorrect representation. This is because using a bilingual dictionary can easily cause the target end translation t_i to be the same in different source end contexts, ignoring the context information of the source end vocabulary, thereby easily causing incorrect translation problems in the target end. Therefore, this paper introduces a gating mechanism to control the vector dimension of the target end translation t_i entering the encoder, selecting the information conducive to translation to enter the model. The input e_i of the encoder is defined as shown in formula (20):

$$e_i = e_{x_i} + g \circ e_{t_i} \tag{20}$$

Among them, e_{x_i} and e_{t_i} are the word vectors corresponding to x_i and t_i respectively. The symbol " $\circ$ " represents element-wise multiplication, and " g " represents the gating mechanism. The definition of " g " is as shown in Formula (21):

$$g = \sigma\left(W_x e_{x_i} + W_t e_{t_i} + b\right) \tag{21}$$

Among them, σ represents the sigmoid activation function, $W_x \in R^{d\times d}$, $W_t \in R^{d\times d}$, $b \in R^{1\times d}$ are model parameters, and d represents the size of the word vector.

3.4. Experimental Results and Analysis

One of the currently widely used methods for evaluating the quality of machine translation is called Bilingual Evaluation Understudy (BLEU). This method assumes that the closer the machine-generated translation is to the manually-generated translation, the higher the translation quality. This requires the similarity between the target sentences obtained by computer machine translation and those obtained by manual translation. Usually, this is represented by dividing the number of n-grams that appear in both translation results by the total number of n-grams in the target sentence obtained by machine translation. The following is the translation:

Machine translation: You are a scientist

Manual translation: You love science

For this machine translation result, the 1-gram set is {“you”, “are”, “science”, “study”, “home”}, while the 1-gram set of the manual translation is {“you”, “love”, “science”, “study”}. The intersection of these two sets is {“you”, “science”, “study”}, so we have $p_1 = \frac{3}{5}$. Using the same calculation method, the 2-gram set of the machine translation is {“you are”, “are science”, “science study”, “study home”}, and the 2-gram set of the manual translation is {“you love”, “love science”, “science study”}. The intersection of these two sets is {“science study”}, so we have $p_2 = \frac{1}{4}$.

However, this calculation method has certain problems. For example, if the translation is "you you you you you", since the result obtained from multiple matching calculations is $p_1 = \frac{5}{5}$, this is obviously incorrect. One possible solution to this problem is to replace the occurrence count of the n-gram with the minimum value between its occurrence count in the machine translation and the manual translation, that is:

$$Count_{clip} = \min\left(Count, Max\,\mathrm{Re}\,fCount\right) \tag{22}$$

Among them, $Max\,\mathrm{Re}\,fCount$ represents the frequency of a certain n-gram word appearing in the manual translation. The calculation formula for the matching rate of the corrected n-gram word is as follows:

$$p_n = \frac{\sum_{C\in candidate}\sum_{n-gram\in C} Count_{clip}\left(n-gram\right)}{\sum_{C\in candidate}\sum_{n-gram'\in C} Count\left(n-gram'\right)} \tag{23}$$

Therefore, the calculation method of the final BLEU evaluation index is as follows:

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right) \tag{24}$$

Here, BP is the penalty factor, which is used to address the issue where the machine-generated text is short but has a high degree of matching. Let c represent the number of words in the machine-generated text, and r represent the number of words in the manually translated text. Then, we have:

$$BP = \begin{cases} 1 & , if\ c > r \\ \exp\left(1 - \frac{r}{c}\right) & , if\ c \le r \end{cases} \tag{25}$$

The weight factor $w_n = \frac{1}{2^n}$ for p_n is used. Generally, the value of N does not exceed 4.

Taking two simple sentences as examples, Figure 2 and Figure 3 show the results of training sentence A and training sentence B respectively. The x and y axes represent the words of the target sentence and the source sentence respectively. Each pixel represents the weight of the source sentence for the target sentence, which is the manifestation of the attention mechanism. The bar chart on the right ranges from 0.0 to 1.0 and represents the weight size (0.0 is blue and 1.0 is red).

Through experiments, it has been found that by using a simple sequence-to-sequence model, machine translation can achieve certain results. It can be seen that when translating a sentence, the attention is constantly concentrated on different areas of the encoded output.

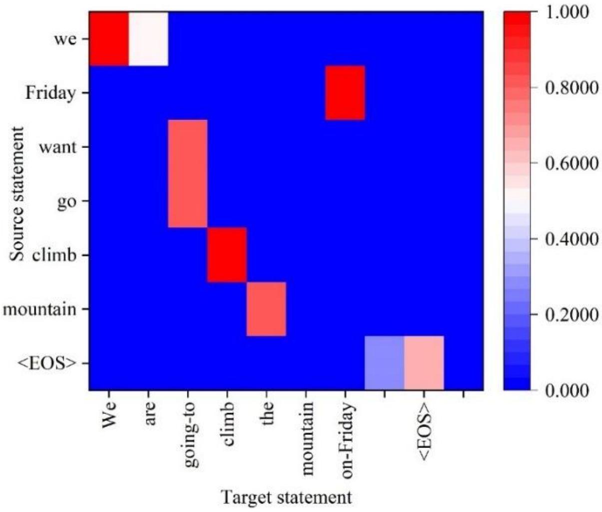

Figure 2. Training statement A

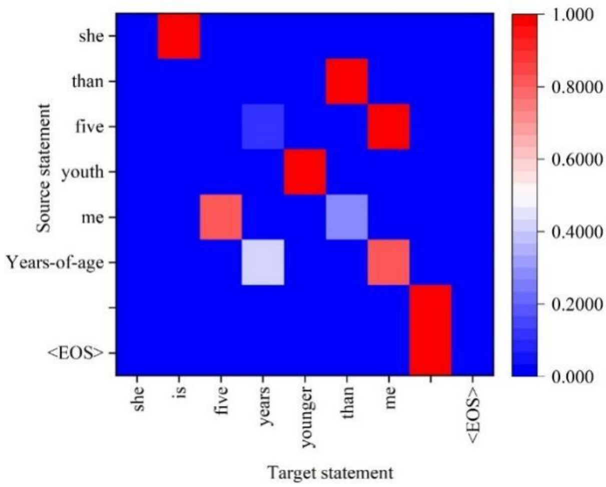

Figure 3. Training statement B

This experiment uses BLEU as the evaluation criterion to measure the accuracy rate and word error rate of the translation. The value of BLEU ranges from 0 to 1. The closer the value is to 1, the better the machine translation effect and the lower the word error rate. The dataset used in this experiment is a Chinese-English paired dataset. The predicted translated data and the reference translation data are stored in different documents. The translation quality is analyzed by calculating the values of BLEU1, BLEU2, BLEU3, and BLEU4. The Chinese-English dataset used in this experiment is sourced from the bilingual multimodal corpus of Chinese traditional culture. 400 sentences were selected for calculation, and the results are shown in Table 3.

In the table, "gram" represents the number of words. When "gram" is 4, the BLEU value is relatively lower than when it is 1. However, when the number of words is large, the accuracy rate is high. Therefore, BLEU4 is chosen as the final evaluation value for this experiment. From the value of BLEU4, it can be seen that the model's BLEU4 value has increased by 1.45% after using the attention mechanism, and the evaluation value of BLEU4 is relatively high, indicating that the predicted translation effect is relatively ideal.

Table 3. BLEU value contrast

Model	BLEU-1gram	BLEU-2gram	BLEU-3gram	BLEU-4gram
With attention model	0.5498	0.5235	0.4926	0.4578
No attention model	0.5401	0.5104	0.4836	0.4433

4. Compensation Strategies for Conceptual Metaphor in the English Translation of Chinese Classical Literature

Although after a long period of differentiation, humans began to exhibit significant differences in their characteristics, in the material world, the social development stages experienced by humans are basically the same. During the process of social development, human thinking and emotions, apart from forming different languages and cultures, also contain certain commonalities. Therefore, Chinese and English have extremely similar degrees of compatibility in some figurative expressions. For such highly translatable expressions, the common practice is to adopt the direct translation method that retains the image. The direct translation method means that the original text can find a basically the same or completely identical expression in the translation, even if it is translated literally, it can convey all the information of the words without changing the image of the original text. There are also some sentences that, although there are no ready-made words for translation in the process of translation, can still adopt the figurative direct translation. For example, the proverb "Silence is golden" can be translated as "Silence is golden." "Out of sight, out of mind" can be translated as "Out of sight, out of mind." No matter what expression is used in the original text to present the image, as long as the readers can have the same associations as the original readers when reading the translation, the figurative direct translation method should be adopted. From these figurative direct translation sentences, it is not difficult to see that the cultures of China and the West have gradually merged over a long period of interaction. Many proverbs, idioms or expressions have similarities. For this limited type of translation, most will adopt the direct translation method.

The language systems of English and Chinese are different. In terms of expression, both China and the West have their own unique ways of thinking. This difference determines that in English-Chinese translation, it is impossible for all things to be translated using the direct translation method. Most expressions still need to be translated through appropriate image transformation. Only in this way can the translation gap be reduced and the effect of the original text and the translation be equivalent. For example, in Chinese, the expression "Love me, love my dog" means liking someone and including the love for the dog on his roof, metaphorically meaning loving someone and caring about the people or things related to him. In English, the word "dog" will be replaced with a common animal and translated as "Love me, love my dog." These idioms or expressions with cultural connotations have different images formed in the feelings of Chinese and Western readers, but within the range of associations, they are very similar. If these words have the same effect in the readers of both cultures, the strategy of converting the image for translation can be adopted to achieve equivalent translation between Chinese and English.

The differences in Chinese and Western cultures lead to the uniqueness of each culture. With the gradual differences in social systems, customs, folkways and religious beliefs, many languages of China and the West have formed their own unique expressions over the years. These expressions cannot be translated using the direct translation or image transformation methods. If they need to be translated, they can only be completely discarded of the original image and translated through the translator's understanding of the text for interpretation, in order to reduce the sense of difference in reading. In Chinese, both idioms and proverbs are difficult to be translated. Usually, the original image needs to be abandoned and the meaning explained, such as "Severe man loses his horse, but knows it is good fortune." "The ship is already built" and "Love money as if it were life" are all idioms that have been accumulated in Chinese culture over thousands of years. If translated directly, it is bound to cause misunderstandings for the readers. However, adopting the image-abandoning interpretation strategy, although it cannot vividly interpret the image meaning of the original text, it explains the deep meaning of the idiom and does not cause a significant understanding deviation in the reading process. Although the readers of the translation cannot feel the author's creative emotions as the original readers do, they can understand the appreciation of the author's creation by the original readers, thereby obtaining a new reading enjoyment.

5. Conclusion

This paper proposes a metaphorical vocabulary recognition model that integrates character information and applies it to the neural machine translation task of classical literature English translation for metaphorical vocabulary. Finally, it discusses the compensation strategies for conceptual metaphors.

(1) The article fully considers the rich information contained in characters and words, and designs a metaphor recognition model CBAM that integrates character information and the attention mechanism. This model uses BiLSTM and multi-head attention mechanism to obtain the context semantics of the target sentence, and uses the Softmax classifier for metaphor recognition and outputs the results. The performance of the model in all word classes and only in verb cases is superior to the best baseline, with the F-score increased by 5.0% and 5.5% respectively. This shows that the method in this paper enriches the word representation of the target and improves the accuracy of metaphor recognition in the neural machine translation of Chinese classical literature.

(2) The neural machine translation model for the translation of metaphorical vocabulary proposed in this paper uses the metaphor recognition method to identify metaphorical vocabulary in the source sentences of Chinese classical literature. It uses the gating mechanism to control the input dimension of the word vector of metaphorical vocabulary, and finally completes the model training through the encoder-decoder. The model using the attention mechanism in this paper has an increase of 1.45% in the BLEU4 evaluation index, and the evaluation value reaches 45.78%, indicating that the translation effect is relatively ideal. The neural machine translation model for the translation of metaphorical vocabulary proposed in this paper improves the translation level of Chinese classical literature translation.

(3) At the compensation strategy level, this paper proposes three different compensation paths. For sentences with high image expression compatibility between Chinese and English, a direct translation strategy that retains the image is adopted, so that the readers of the translation and the original text readers have the same associations. For expressions that cannot be directly translated due to differences in language systems and thinking methods, a modified translation strategy that transforms the image is adopted. For idioms and proverbs carrying unique cultural connotations, the original image is discarded for interpretation, to ensure the comprehensibility of the translation.

References

1. Xia, C., & Jing, Z. (2018). English translation of classical Chinese poetry. Orbis Litterarum, 73(4), 361-373.
2. Shu, H. (2018). A Model for English Translation of Chinese Classics. Language & Semiotic Studies, 4(4).
3. Xiangtao, F. A. N. (2026). Investigating the Translation of Chinese Classics: Problems and Methods. Taylor & Francis.
4. Qi, R., & Roberts, M. (2020). Classical Chinese Literature in Translation: Texts, Paratexts and Contexts. Translation Horizons, 10(2), 1-6.
5. Pan, W. (2025). Translating into/out of one’s mother tongue: on the feasibility of translating Chinese classics into english by native Chinese translators. Translation and Interpreting Studies, 20(2), 302-316.
6. Li, F. (2015, December). Current Situation and Results on English Translation Research for Chinese Cultural Classics. In 2015 3rd International Conference on Education, Management, Arts, Economics and Social Science (pp. 984-990). Atlantis Press.
7. Liu, B., & Xie, J. (2022). On the Translation of Architectural Culture in Chinese Classics from the Perspective of Cultural Translation Theory. Transactions on Comparative Education, 4(1).
8. Li, R. (2024). The Dilemma and Choices of Indirect Translation: The Transformation of Chinese Classical Novels in Europe from the Early Translations of Yu Jiaoli. Comparative Literature: East & West, 8(1), 63-74.

9. Qiufen, W., Amini, M., & Tan, D. A. L. (2025). Strategies, errors, and challenges in translating culture-specific items in Chinese-English literary works: A systematic review. Jurnal Arbitrer, 12(2), 259-273.
10. Amini, M., Li, J., Siaw, J. X., & Kasuma, S. A. A. (2023). Exploring Challenges in Translating Li Bai's Classical Chinese Poems to English. International Online Journal of Language, Communication, and Humanities, 6(1), 48-60.
11. Zhi, L. I. (2017). The Study of English Translations of Chinese Classics in China (1997–2016): Recent Development and Future Prospects. US-China Foreign Language, 15(2), 121-128.
12. Yang, Y. (2025). Analysis and Countermeasures of Mistranslation in English Translation of Chinese Classics——A Case Study of Ming and Qing Novels. Mediterranean Archaeology and Archaeometry, 25(2).
13. Xiao-chun, G. U. O. (2020). Discussion on the Translation of Chinese Classics. Journal of Literature and Art Studies, 10(10), 895-902.
14. Lu, R. (2025). English Translation Studies: Chinese Literary Works. Arts, Culture and Language, 1(3).
15. Abhishek Bajpai & Shalinee Sahu. (2026). Lightweight and Interpretable Hybrid Deep Network for Potato Leaf Disease Detection Using CBAM and Vision Transformer. Potato Research,69(1),21-21.https://doi.org/10.1007/S11540-025-09981-8.
16. Shunfu Lin,Bing Zhao,Yinfeng Zhan,Junsu Yu,Xiaoyan Bian & Dongdong Li. (2024). Non-intrusive residential load identification based on load feature matrix and CBAM-BiLSTM algorithm. Frontiers in Energy Research,12,1443700-1443700.https://doi.org/10.3389/FENRG.2024.1443700.
17. Yingming Cai,Hengkai Li,Kunming Liu,GuanShi Wang & ZhiYu Zhang. (2026). MSDCC-Net: A Fine-Scale Remote Sensing Extraction Method for Bare Surface Land in Rare Earth Mining Areas Based on Multi-Scale Attention Mechanisms. Land Degradation & Development,37(12),8170-8183.https://doi.org/10.1002/LDR.70647.
18. Muhammad Rashid,Junfeng Wang & Sulman Ahmed. (2026). Enhancing industrial human pose estimation through a hybrid Transformer–CNN architecture. The Journal of Supercomputing,82(10),542-542. https://doi.org/10.1007/S11227-026-08689-X.