

# Design of a Smart Teaching English Translation System Based on Big Data Machine Learning

Chunye Zhang, Harbin University of Science and Technology, China\*

Tianyue Yu, Harbin University of Science and Technology, China

Yingqi Gao, Harbin University of Science and Technology, China

Mau Luen Tham, Universiti Tunku Abdul, Malaysia

## ABSTRACT

In the context of artificial intelligence, the use of machine translation in English reading classroom teaching is a more common learning method. In traditional teaching methods, machine translation is more convenient and faster than human translation, but it often deviates from the original text in terms of grammar and sentence pattern. Based on the perspective of English reading class, this paper compares traditional and machine translation, and discusses the future development trend and influence mechanism of the current situation of using machine translation in English reading class under the effect of artificial intelligence.

## KEYWORDS

Artificial Intelligence, English Translation, Intelligent Robot, Machine Learning, Smart Teaching

## INTRODUCTION

With the enhancement of China's comprehensive national strength and international competitiveness and the deepening of trade and cultural exchanges with countries around the world, English, as the most widely used language, has become a bridge between China and other countries. Therefore, it is important to learn English well (Wairagkar et al., 2021). In the actual process of teaching English, reading is the most important means for people to obtain information today (Church et al., 2021). The acquisition of new knowledge, the improvement of thinking ability, and the enhancement of social adaptability are closely related to reading (Gröls, 2022). In the teaching of English reading, students will inevitably encounter unfamiliar words and sentence patterns in the learning process, and the function of translation is essential here (Nguyen et al., 2022). It is not difficult to find that machine translation is quietly changing the way people learn (Arumugam & Kumar, 2021). From international conferences to daily life, Chinese-English language conversion is essential (Kang, 2021). With the rise of Artificial Intelligence, there is

DOI: 10.4018/IJWLTT.330144

\*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

a gradual shift towards machine translation in the English classroom (Bondaryk et al., 2021). Some scholars have indicated that “simultaneous interpretation profession will be replaced by machine translation” (Dämmer, 2022, p.20). In the actual English teaching classroom, the use of machine translation and the use of computers to convert Chinese into English are the main ways students choose to learn English (Kailasam, 2022).

At first, deep learning technology was applied in the field of image processing, which can extract the most relevant features of the target from many images without manual settings. Later, it was applied to the field of machine translation to produce machine translation based on neural networks, which processes complex information by simulating the “hierarchical learning” of the human brain. Compared with traditional statistics-based machine translation methods, the quality of neural machine translation (NMT) has been significantly improved, and it has become the core technology of commercial machine translation systems. In addition, thanks to the rapid development of computer technology and artificial intelligence technology, neural machine translation methods are also making continuous progress and self-innovation to generate higher-level translations. This progress has included the initial methods based on a recurrent neural network (RNN) to transformer-based methods to the current popular cross-language-based pre-training model methods, such as mBART, T5, and mBERT models (Eludiora et al, 2021). Although the quality of the models is constantly improving, the structure, parameter, quantity, and data size of the model are correspondingly becoming larger and more complex, and the machine translation method has also changed from the original experience-based knowledge-driven approach to the data-driven approach (Wen, 2020). Therefore, large-scale and high-quality bilingual parallel corpora are the basis for improving the performance of the model (Chen, 2022).

Machine translation uses “direct translation,” “syntax conversion,” “semantic conversion,” and other technical means to make the target language approach the source language as much as possible and try to maintain its fluency and accuracy (Song, 2021). At present, machine translation can be divided into the following three categories: rule-based machine translation, statistics-based machine translation, and neural network-based machine translation (Liu, 2021). In the early 1990s, due to the success of deep learning and neural networks in the field of artificial intelligence, machine translation introduced these new technologies, so statistical machine translation developed into neural machine translation (Xia, 2020). The main characteristics of the artificial neural network are the distributed storage of information and the parallelization of information processing (Izaz, 2022). It adopts the connectionism method and has the abilities of self-organization and self-learning, which enable people to use machines to process information in new ways and solves some problems that are difficult to solve with the traditional symbolism method (Samadi & Jond, 2021). Currently, most of the translation tasks are low resource tasks (Zhang, 2020). Due to the lack of a corpus to provide sufficient knowledge for the model, no matter what structure and training method the model adopts, it is difficult to significantly improve the quality of the translation (Doltsinis et al., 2020).

This paper starts from the need to improve the quality of low resource language machine translation. First, it summarizes some problems of low resource tasks at this stage, and expounds some mainstream low resource neural machine translation methods, such as transfer learning methods and pre-training methods (Yuan, 2020). Taking advantage of the rich language materials of high resource languages, the problem of low translation quality of low resource languages can be solved by means of two transfer learning and word list integration. Secondly, from the perspective of decreasing generalization ability during model domain migration (Kun et al., 2021), the model was fine-tuned twice to improve the generalization ability of low resource language translation models. This paper aims to provide some reference and help for the follow-up exploration and research in the field of machine translation.

## RELATED WORK

### Overview of Machine Translation

Research on machine translation has been carried out for nearly 70 years. In 1954, the machine translation experiment conducted by Georgetown University and IBM was regarded as a milestone in this field (Liu et al, 2021). Looking back on the development of machine translation, it can be roughly divided into three stages (Green et al., 2016). The first stage is rule-based machine translation (RBMT), which requires the help of bilingual dictionaries and linguistic rules of each language (Lei et al., 2020). Linguists design conversion rules between different languages according to the characteristics of different languages, and then computer scientists convert the rules into executable programs (Li, 2021). Because this method requires linguists to write many rules, natural language itself has endless rules and changes, and with the increase of rules, the interference and conflict between rules also increase gradually. This method soon ran into a bottleneck and was eventually eliminated by history.

The second stage is example-based machine translation (EBMT). After the rule-based method met the bottleneck, in 1984, Long Tail of Kyoto University tried to get rid of the constraint of rules and proposed an instance-based machine translation method. This method first saves multiple corresponding translations, then directly compares the sentences to be translated with the saved corresponding translations in the process of translation, finds out the differences, and replaces them. This method was the first to propose the concept of a parallel corpus. The more sentences in the corpus, the more accurate the translation results will be.

With the birth of the case-based machine translation method, the third stage of the research field was born soon after: the statistical-based machine translation method (SMT). This method does not involve the formulation of rules or the establishment of bilingual corresponding dictionaries, and it calculates the translation result with the highest probability completely according to the statistical results. IBM put forward the prototype of this method in the 1990s and constantly improved on its basis.

Whether it is the EBMT of translation templates obtained from translation examples or the SMT, they belong to the category of empirical methods. They all use computers to learn how to translate from many actual translation sentence pairs. From the perspective of the specific application of machine translation, based on the diversity of machine translation tools, teachers should pay attention to selecting the appropriate translation machine according to the specific teaching requirements for the course in order to maximize the auxiliary role of the translation machine in English reading.

### Neural Network Machine Translation

In the second decade of the 21st century, deep learning has developed rapidly, which also affects the direction of machine translation research. At first, the neural network was only used as a substitute for some components of statistical machine translation, and the model structure did not change. Until Sutskever (Sutskever et al., 2014) put forward an end-to-end neural network machine translation method, which is completely based on the neural network structure, greatly simplifying the complexity of feature design and model. This has received extensive attention and rapid development in recent years. Statistical machine translation can only focus on some words or phrases in the translation process, which may produce local optimal solutions in the translation results. Although the logic of this translation is like the characteristics of paying attention to some words in human translation, it lacks the important step of reading sentences thoroughly before human translation. In contrast, the neural network machine translation method realizes this process by converting all words in the sentence into a fixed length vector to complete the “read through” step before translation and to obtain a more fluent predicted translation. Many experiments show that the performance of machine translation based on the neural network has surpassed that of statistical machine translation, and it is the mainstream machine translation method at present. The languages involved in the evaluation of neural network machine translation in the above research include English, French, German, and

Spanish, among others, but there is a lack of English-Chinese translation. Compared with the translation between English and French or English and German, the translation between English and Chinese is more difficult, and it is also one of the translation language pairs with the lowest performance of neural network machine translation system. In the research of English-Chinese machine translation, Chinese to English translation is the main direction. According to the test results, the quality of English to Chinese translation from the neural network machine translation system is inferior to that of Chinese to English translation.

The machine translation method based on neural network discards the components in statistical machine translation models, such as the word alignment model and the language model, and uses an end-to-end structure, which is divided into two parts: encoder and decoder. The encoder first encodes the source language into a group of vectors and then the decoder decodes the target language from the group of vectors.

The outstanding neural network machine translation performance mainly depends on big data, deep network models, and complex computing. Firstly, large-scale training data is needed to ensure the performance of the neural network machine translation systems. Secondly, in order to effectively learn, the hidden layers in the deep neural network structure must reach a certain number. Finally, neural network machine translation also relies on complex computing. The personal computer is no longer sufficiently competent for the training task of the deep learning model.

With the development of deep learning technology, neural machine translation has experienced the development of cyclic neural network structure, convolution neural network structure, and network structure based on self-attention mechanism. Compared with statistical machine translation, neural machine translation is based on the coder-decoder framework and does not need artificial features or other prior domain knowledge. It realizes automatic coding and decoding through the neural network. On the other hand, neural machine translation is data driven. The neural machine translation system based on deep learning technology has many parameters in its model structure. The performance of the translation system depends heavily on high-quality parallel corpus, so it does not perform well in translation tasks in low-resource areas. Parallel corpora are difficult to obtain, and high-quality parallel corpora are costly. For almost any language, monolingual corpora are very easy to obtain. Therefore, based on a limited bilingual corpus, using a monolingual corpus to enhance data and improve translation performance has been an important research direction in the field of neural machine translation for a long time.

In the context of AI, translation teaching needs more technical means to develop higher level translation abilities more quickly and effectively. Translation teaching in colleges and universities has also started to set up post-translation editing courses (Kobyakova & Shvachko, 2016) to expand new areas of translation services and train professional post editors. In translation teaching, courses focused on learning about tools and software should be added. Skilled use of translation software has gradually become one of the skills that high-end translators must master.

## **METHODS**

### **Virtual Template Assumption Based on the Back Propagation Process of the Machine Learning Model**

The machine learning model is a “grey box” that is obtained by data-driven training with a known structure. In the process of machine learning, the model gradually selects effective information from the original 3D perceptual data and produces meaningful output. This effective information is difficult to understand because of its high dimension and abstract characteristics. This paper will analyse it through the virtual template hypothesis.

The convolution information extraction function based on the parameter sharing mechanism can be expressed as Formula (1), where  $z$  is the convolution output,  $x$  is the input feature, and  $y$  is the convolution kernel weight:

$$z = y \cdot x \quad (1)$$

In the process of backpropagation, the weight training Formula of the convolution kernel is Formula (2), where,  $y$  is the current convolution kernel weight, and  $y'$  is the convolution kernel weight after iterative update,  $\alpha$  is the training rate parameter,  $dy$  is the differential of the convolution kernel weight,  $dz$  is the differential of the back propagation error, and  $x$  is the input feature participating in the convolution calculation:

$$y' = y + \alpha \cdot dy, \quad dy = dz \cdot x^T \quad (2)$$

It can be seen from Formula (2) that the update value of the convolution kernel parameter in a training is determined according to the local characteristics of the input. The nonlinear activation function can filter meaningful output according to the output value of the convolution function. Taking the widely used ReLU function as an example, if the value of the output feature  $z$  in formula (1) is less than 0, it is set to 0 by the ReLU function, and the corresponding convolution kernel weight  $y$  is not updated through the input feature  $x$ , it can be determined that the convolution kernel  $y$  does not match the input feature  $x$ . If only the input feature  $x$  matched with the convolution kernel  $y$  is considered, the convolution kernel value after the machine learning training is completed is as it shown in Formula (3), where,  $y_0$  is the random initialization value of the convolution kernel, and the second term in the Formula is the sum of the iterative process of the convolution kernel in the training process:

$$\hat{y} = y_0 + \alpha \cdot \sum dz \cdot x^T \quad (3)$$

It can be seen from Formula (3) that the convolution kernel  $y$  finally converges to the value related to the weighted average value of the matching feature  $x$ , and the weighted average value of the matching feature is the virtual template proposed in this paper.

The virtual template hypothesis of the machine learning process provides a basis for the establishment of a mathematical model of the convolution process. Considering that a group of similar local features have differences that cannot be accurately described, random variable  $X$  is used to represent similar input features. Based on the nature of the virtual template, there is a unique virtual template  $p$  corresponding to the input feature  $X$ . Since the parameters of the convolution kernel are fixed after the training of the machine learning model is completed, the corresponding virtual template  $p$  is also a certain quantity. Based on the virtual template assumption, the input features can be further decomposed into Formula (4), where,  $X$  represents the input of the convolution function,  $p$  represents the virtual template, and  $W$  represents the random variable describing the difference noise between the input feature  $X$  and the virtual template function:

$$X = p + W \quad (4)$$

In the process of three-dimensional convolution,  $p$  and  $W$  are three-dimensional tensors with the same size as the three-dimensional convolution kernel. Each element in the noise variable  $W$  is independent of each other and follows the normal distribution. The three-dimensional convolution

output can be calculated as Formula (5), where,  $y$  represents the convolution kernel weight, and  $b$  represents the offset parameter of the convolution operation:

$$Z^n = \sum_{c=0}^{c=0} \sum_{i=0}^{i=0} \sum_{j=0}^{j=0} \sum_{k=0}^{k=0} X_{cijk} \cdot y_{cijk}^n + b^n \quad (5)$$

$Z^n$  represents the output of the 3D convolution,  $c$  represents the dimension position of the feature,  $cijk$  represent the position of the element in the input feature space, and  $mmml$  represent the feature dimension and spatial scale of the variable. Since the three-dimensional convolution operation allows  $n$  convolution check input features to be used for processing, and  $n$  convolution cores are equivalent, for the convenience of explanation, only the  $n$ th convolution core is discussed here, as shown in Figure 1.

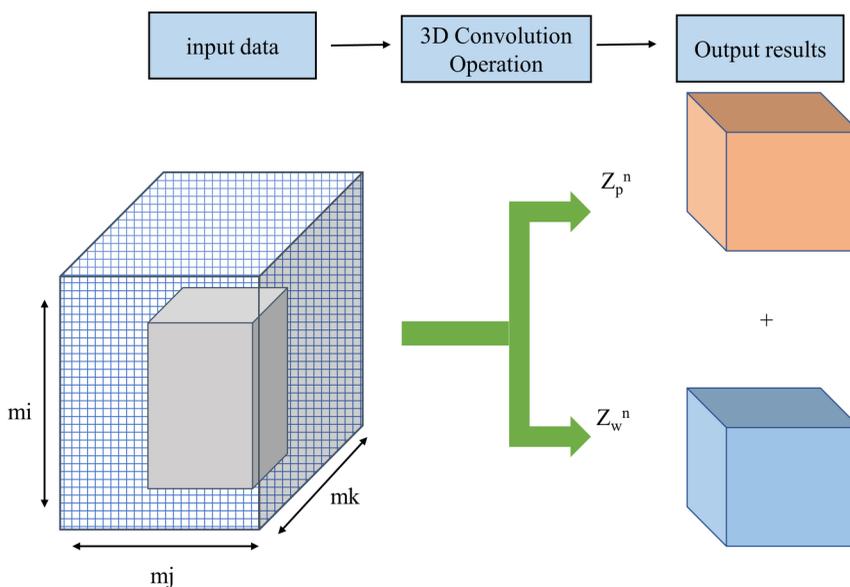
The convolution kernel  $y$  after training will generate strong response  $Z$  after convolution calculation with the matching feature  $X$ . Suppose that two similar matching features  $X_1$  and  $X_2$  operate with the convolution kernel  $ny$  to generate response outputs  $Z_1^n$  and  $Z_2^n$ , respectively, then the difference between  $Z_1$  and  $Z_2$  can be represented by the output feature difference  $Z_d^n$ , as shown in Formula (6):

$$Z_d^n = Z_1^n - Z_2^n \quad (6)$$

In Formula (7),  $W^1$  and  $W^2$  obey Gaussian distribution, so their expected value is 0, so the expectation of output characteristic difference  $Z_d^n$  is also 0. The variance of the output characteristic difference  $Z_d^n$  is shown:

$$Var(Z_d^n) = \sum_{c=0}^{c=0} \sum_{i=0}^{i=0} \sum_{j=0}^{j=0} \sum_{k=0}^{k=0} (Var(W_{cijk}^1) + Var(W_{cijk}^2)) \cdot (y_{cijk}^n)^2 \quad (7)$$

Figure 1. 3D convolution decomposition based on virtual template assumption



Because similar inputs produce similar outputs, the absolute value of the output characteristic difference  $Z_d^n$  is small. Although the expected value of  $Z_d^n$  is 0, the variance value will still cause similar inputs to produce significantly different outputs.

### Translation Method Based on Machine Learning

The basic idea of meta learning is to use the source task set  $\{T^1, \dots, T^K\}$  to find the initialization parameter  $\theta_0$ , and only needs a small number of training samples to learn the target task  $T_0$ . In machine translation, this is equivalent to using many high-resource language pairs to find good initial parameters and training new translation models on low-resource translation tasks from the found initial parameters, as shown in Formula (8):

$$\theta^* = \text{Learn}\left(T^0; \text{MetaLearn}\left(T^1, \dots, T^K\right)\right) \quad (8)$$

According to Formula (8), the learning process of high-resource tasks can be regarded as learning specific task parameters, and the ultimate goal is to provide initial parameters for low-resource tasks. The specific process is discussed below.

Given any initial parameter  $\theta_0$ , the prior distribution of the corresponding NMT model parameters satisfies the isotropic Gaussian distribution  $\theta \sim N(\theta_0, \beta I)$ , where  $\beta I$  represents the variance. According to the prior distribution, the learning process of a specific task is described as the logarithmic posterior probability of the model parameter that maximizes the given data, as shown in Formula (9):

$$\text{Learn}\left(D_T; \theta^0\right) = \arg \max_{\theta} L^{D_T}(\theta) = \arg \max_{\theta} \sum_{(X,Y) \in D_T} \log p(Y|X, \theta) - \beta \theta - \theta^0 \quad (9)$$

Through the training of specific high resource language pairs mentioned above, the model can simulate the translation scenarios of low resource tasks and find the corresponding parameters. Here, the objective function of meta learning is shown in Formula (10):

$$\mathcal{L}(\theta) = E_k E'_{D_{T^k}, D'_{T^k}} \left[ \sum_{(X,Y) \in D'_{T^k}} \log p(Y|X; \text{Learn}(D_{T^k}; \theta)) \right] \quad (10)$$

In Formula (10),  $k$  represents a meta learning scenario, and  $D$  and  $D'$  obey uniform distribution in data set  $T$ . It can be seen that the SGD method is used to approximate the maximum meta learning objective function. For each meta learning scenario, randomly sample a high resource task  $T$ , then sample two batches of training samples  $D$  and  $D'$  from the selected task, use the training samples of the first batch to train specific task model parameters, and use the training samples of the second batch to evaluate the model. When the learning rate is  $\mu$ , the updating process of specific task model parameters is shown in Formula (11):

$$\theta'_k = \text{Learn}\left(D_{T^k}; \theta\right) = \theta - \mu \nabla_{\theta} \mathcal{L}^{D_{T^k}}(\theta) \quad (11)$$

In this paper, the parameter  $\theta$  obtained from the training of the first batch is used for the evaluation of the second batch dataset  $D$ , and the calculated gradient  $\theta$  is used as the meta gradient to update the meta parameter. Then the updating process of the meta parameter is shown in Formula (12):

$$\theta \leftarrow \theta - \mu' \sum_k \nabla_{\theta} \mathcal{L}^{D_{T^k}}(\theta'_k) \quad (12)$$

However, although the meta learning method alleviates negative transfer and other problems to a certain extent compared with ordinary transfer learning and also shows better potential and robustness to new tasks, the updating of meta parameters is limited by the two-ladder degree in specific task training, resulting in a large consumption of training resources. The reason is that when meta learning method is adopted, given the training set *D r a n n* and test set *D e s t* corresponding to a specific task, the maximum likelihood estimation method is used to learn model parameters, as shown in Formula (13):

$$Learn(D_{train^i}; \theta') = \theta - \alpha \nabla_{\theta} Loss_{t_i}^0(nmt_{\theta}^i) \quad (13)$$

$\theta$  represents the model parameters,  $\alpha$  represents the learning rate, and  $n m t$  represents the corresponding machine translation model.  $L o s s t 0$  represents the loss function corresponding to the 0th batch of data in task  $t$ . In order to generalize the model parameters to different new tasks, initialize the parameter  $\theta$  in each iteration step, use the training set *D r a n n* of  $K$  specific tasks to calculate the gradient and obtain the new parameter  $\theta^*$ , and then use the test set *D e s t* to update the meta parameter  $\theta$  for  $K$  tasks. Task specific model parameters  $\theta^*$  and meta parameters  $\theta$  are shown in Formula (14) and Formula (15):

$$\theta^* = \arg \min_{\theta} \sum_{t_i \sim p(t)} Loss_{t_i}^1(nmt_{\theta}^i) \quad (14)$$

$$\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{t_i \sim p(t)} Loss_{t_i}^1 \left( nmt_{\theta - \alpha \nabla_{\theta} Loss_{t_i}^0(nmt_{\theta}^i)} \right) \quad (15)$$

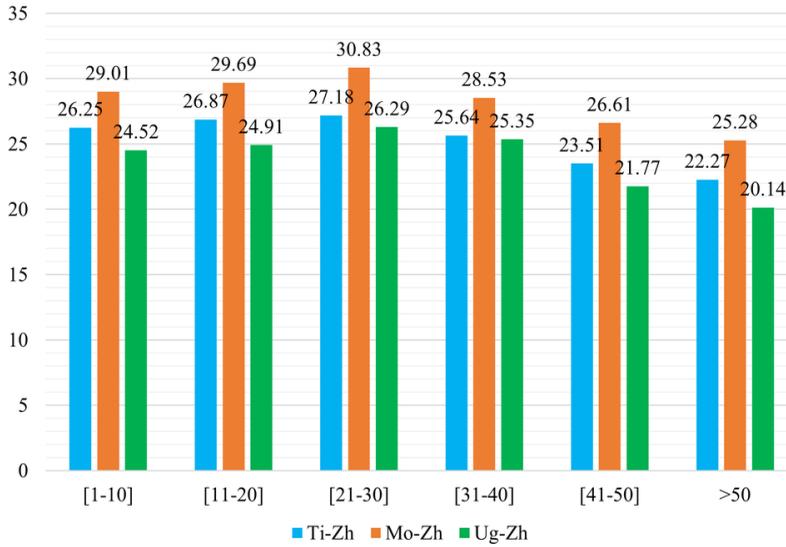
$\beta$  represents the meta learning rate, and other variables are consistent with the above. As shown in Formula (15), meta learning is regarded as a parameter updating process of loop nesting. Specific task parameters are learned in the inner loop, and then new batches of data are sampled from the same training data set in the outer loop to calculate the model's meta gradient. The meta parameters are updated iteratively with the current meta gradient direction as the update direction. It can be seen that this method updates the meta parameters by calculating the two-ladder degree, which consumes a lot of computing resources.

## RESULTS AND DISCUSSION

### Effect of Meta Learning Methods on the Quality of Machine Translation Models

Manual evaluation of translation quality is high, but translation takes a long time and is expensive. Therefore, Papineni et al. proposed the method of Bilingual Evaluation Understudy (BLEU)—that is, the method of automatic evaluation of machine translation quality (Papineni et al., 2002). BLEU evaluates the quality of translation from the perspectives of adequacy, smoothness, and accuracy. The results of translation are compared with those of human translation. The higher the BLEU value, the better the translation, and vice versa. The examples in this paper include sentence length analysis, visual word alignment analysis, and translation analysis. Figure 2 shows the BLEU scores corresponding to various methods under different sentence lengths.

Figure 2. BLEU scores of different translation tasks

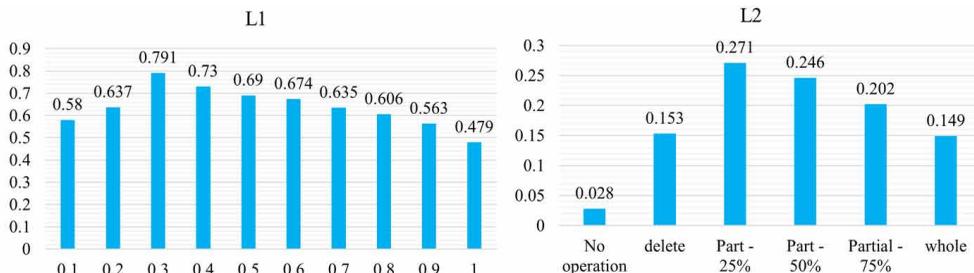


When the sentence length is between 21-30, BLEU scores are the highest. When the sentence length is greater than 50 words, BLEU scores show a downward trend. In this paper, we observe the mapping quality of the proposed cross-language word embedding through the attention heat map and the word embedding map in vector space. The alignment accuracy of other words is more than 50%, which shows that MAWE can restore its original semantics and parts of speech better than CSLS when learning verbs, nouns, adverbs and other words of parts of speech.

For our gradient-based word flipping experiment, the most common methods are typography confrontation and punctuation confrontation. Among them, the typo confrontation is to replace, delete, or insert the constituent characters of the input word, thereby changing its original semantics, while the punctuation confrontation is to replace or delete the punctuation in the training set, forcing the model to predict the real punctuation according to the context. Figure 3 shows the proportion of typographical replacement of words and the impact of punctuation on  $L_2$  loss.

When the number of wrong characters accounts for 30% of the total number of characters in a word, the loss function increases significantly, reaching 0.791. When the words are partially deleted and the deletion ratio is 25% of the total punctuation, the corresponding  $L_2$  loss increases significantly, reaching 0.271. It can be concluded that after retraining the model with the confrontation samples

Figure 3. Impact of typos and punctuation on the model



generated above, the model can give certain feedback and parameter optimization according to the confrontation attack to learn the non-robust features.

## Simulation Experiment

The neural network and statistical pattern recognition method are selected to compare with the multi feature fusion method proposed in the paper. The recognition accuracy, recognition result recall, and automatic recognition effectiveness of these methods are shown in Figure 4.

Analysis of the data shown in Figure 4 reveals that the experimental results of this indicator reflect the use effect of the three methods. The recognition accuracy of the multi-feature fusion method is relatively high, and it can identify and extract the information of most translation errors. Compared with this method, the other two methods can only recognize a small amount of translation error information after use and cannot perform high-precision analysis and recognition on the experimental group information.

In many experiments, the recognition accuracy of the multi-feature fusion method is higher than that of the other two methods. Therefore, it can be determined that the multi-feature fusion method has a high use value. In order to verify the effectiveness of this method, the original model transformer, the improved model A transformer, the maximum likelihood training method, and the adversarial training method are arranged and combined, and compared with other baseline methods. There are six methods, including four combined methods and two other methods. All experimental results will be explained around these methods. In Figure 5, the existing baseline method and the method in this paper are plotted. On the test data set newstest2017, the change curve of BLEU score, the evaluation indicator of the generated translation with the sentence length is plotted; BLEU index and confusion index are currently the mainstream indicators for machine translation evaluation.

In order to observe the performance of each model in sentences of different lengths, with reference to the practice of Bahdanau et al., this paper divides the sentences in the test set into 6 groups according to the sentence length and evaluates the BLEU value of each group of sentences (Bahdanau et al., 2014). The test results of different sentence lengths are shown in Figure 6. Through observation, it can be seen that the CN-NMT model proposed in this study, which integrates language concept category information, performs better than the baseline model RNN Search and the comparison

Figure 4. Identification accuracy

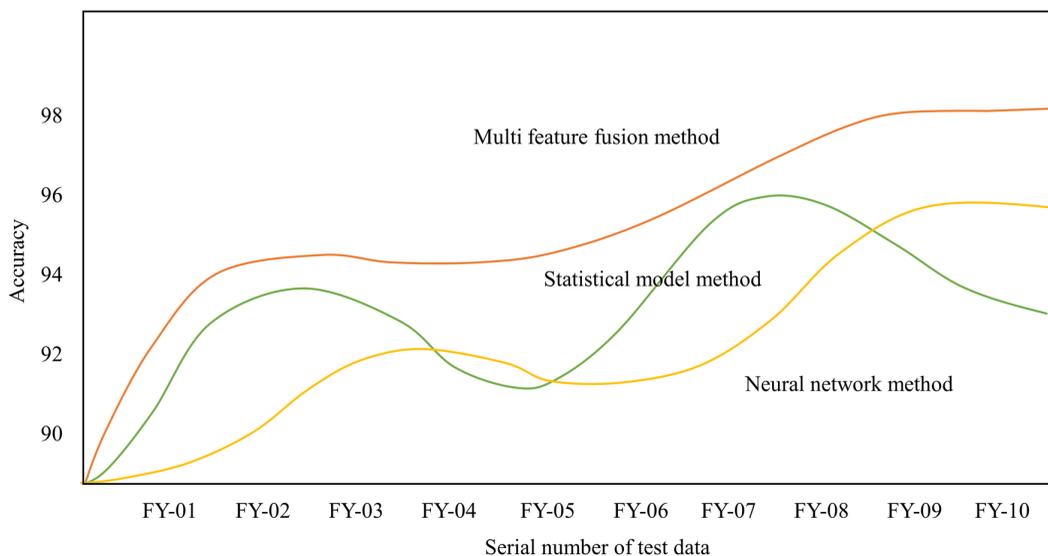


Figure 5. Comparison of BLEU scores of different models with sentence length

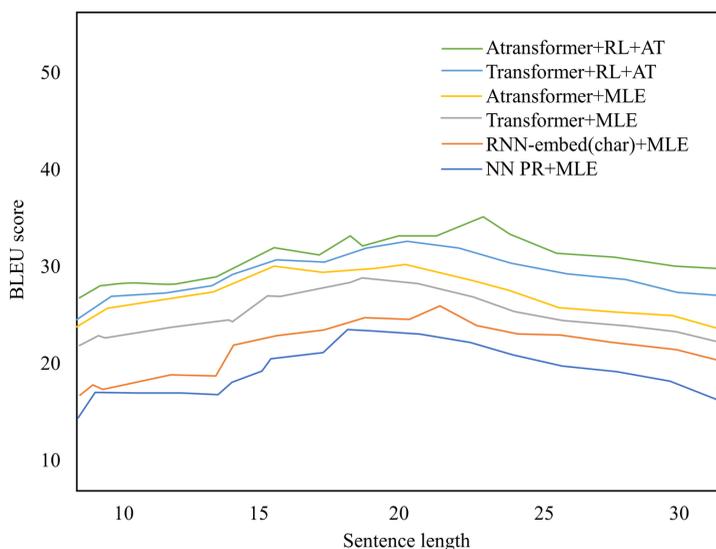
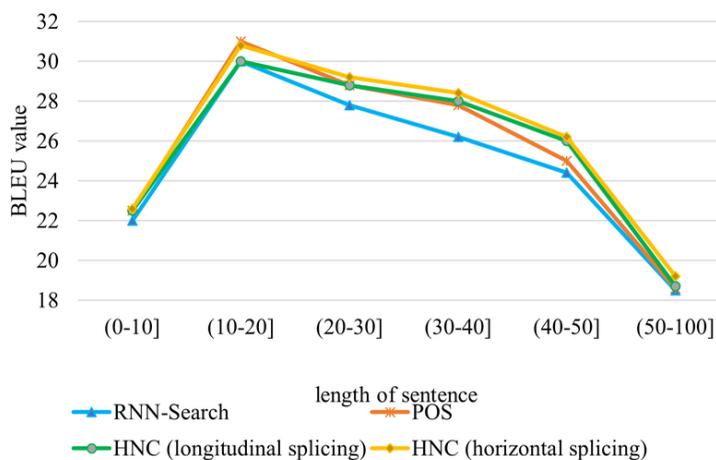


Figure 6. Performance comparison of models in different sentence length intervals



model POS-NMT in all length intervals. Figure 6 also shows that, once the sentence length exceeds 20 words, with the increase of the sentence length the BLEU values of all models have a downward trend, which is consistent with the research. However, we can still find that the decline rate of the CN-NMT model is slower. Once the sentence length exceeds 50 words, the translation performance of each model drops sharply. The most likely reason is that in order to speed up the model training, the sentence length of the source language end of the NMT model is limited to 50 in the experiment, and the sentence length of the source language end of the horizontal fusion concept category label is set to 100. The trained model does not have the ability to handle long sentences. In a word, the experimental results show that the neural machine translation method incorporating conceptual category information can effectively improve the quality of translation, whether in a short sentence or long sentence translation.

Many challenges arise in the rapid development of neural machine translation.

Among many challenges, this paper has carried out research work in data sparsity and model improvement. This paper analyses the dilemma of diversity and high quality of parallel corpus in resource-rich and low-resource scenarios and combines low-frequency word replacement methods and reverse translation methods to achieve complementarity. In this method, the pseudo-parallel corpus generated by the reverse translation method is further enhanced by replacing low-frequency words, and the grammar error correction module is added additionally in low-resource scenarios to reduce grammar errors. This paper validates the model from the perspective of sentence length, visual word alignment, and translation. The experimental results show that the neural machine translation method based on deep learning can effectively improve the translation quality in both short and long-sentence translation.

This paper applies the pre-training model to the translation task. In fact, with the rapid development of deep learning technology, there have been more pre-training models with more model parameters and larger training scales, such as BERT, GPT, ERNIE, and NEZHA. These pre-training models have achieved good results in many downstream tasks in the field of natural language processing but have few applications in the field of machine translation. Future work could consider combining the pre-training model with the machine translation task more closely and carrying out more in-depth research.

## **ACKNOWLEDGMENT**

The authors would like to show sincere thanks to those techniques who have contributed to this research.

## **DATA AVAILABILITY**

The figures used to support the findings of this study are included in the article.

## **CONFLICTS OF INTEREST**

The authors declare that they have no conflicts of interest.

## **FUNDING STATEMENT**

This work was supported by the 2022 National College Student Entrepreneurship Training Program Favors You - Smart Pet Life Platform Based on Big Data Enabling. Project No. 202210214057X.

## REFERENCES

- Arumugam, K., & Kumar, L. A. (2021). Design and development of intelligent home rehab device for post stroke therapy with electro-hydraulic interventions. *Acupuncture & Electro-Therapeutics Research*, 47(1), 1–8. doi:10.3727/036012921X16298112109448
- Bahdanau, D., Cho, K., & Bengio, Y. (2014). *Neural machine translation by jointly learning to align and translate*. arXiv preprint arXiv:1409.0473.
- Bondaryk, L. G., Hsi, S., & Van Doren, S. (2021). Probeware for the modern era: IoT dataflow system design for secondary classrooms. *IEEE Transactions on Learning Technologies*, 14(2), 226–237. doi:10.1109/TLT.2021.3061040
- Chen, D., & Cheng, P. (2022). Development of design system for product pattern design based on Kansei engineering and BP neural network. *International Journal of Clothing Science and Technology*, 34(3), 335–346. doi:10.1108/IJCST-04-2021-0044
- Church, A., Lloyd, J., Hadsell, R., & Lepora, N. F. (2021) Optical tactile sim-to-real policy transfer via real-to-sim tactile image translation. *Proceedings of the 5th Conference on Robot Learning*, PMLR, 164, 1645–1654. doi:10.48550/arXiv.2106.08796
- Dämmer, G., Bauer, H., Neumann, R., & Major, Z. (2022). Design additive manufacturing and component testing of pneumatic rotary vane actuators for lightweight robots. *Rapid Prototyping Journal*, 28(11), 20–32. doi:10.1108/RPJ-03-2021-0052
- Doltsinis, S., Krestenitis, M., & Doulgeri, Z. A. (2020). Machine learning framework for real-time identification of successful snap-fit assemblies. *IEEE Transactions on Automation Science and Engineering*, 17(1), 513–523. doi:10.1109/TASE.2019.2932834
- Eludiora, S., & Ajibade, B. (2021) Design and implementation of English to Yoruba verb phrase machine translation system. *African NLP Workshop, 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. doi:10.48550/arXiv.2104.04125
- Green, K., Smit-Anseeuw, N., Gleason, R., & Remy, C. D. (2016) Design and control of a recovery system for legged robots. *2016 International Conference on Advanced Intelligent Mechatronics (AIM)*, 958–963. doi:10.1109/AIM.2016.7576893
- Gröls, J. R., & Castro-Dominguez, B. (2022). Intelligent mechanochemical design of co-amorphous mixtures. *Crystal Growth & Design*, 22(5), 2989–2996. doi:10.1021/acs.cgd.1c01442
- Izaz, R., Lee, H., & Kim, H. S. (2022). Mechanical fault detection based on machine learning for robotic RV reducer using electrical current signature analysis: A data-driven approach. *Journal of Computational Design and Engineering*, 9(2), 417–433. doi:10.1093/jcde/qwac015
- Kailasam, S., Achanta, S. D. M., Rama Koteswara Rao, P., Vatambeti, R., & Kayam, S. (2022). An IoT-based agriculture maintenance using pervasive computing with machine learning technique. *International Journal of Intelligent Computing and Cybernetics*, 15(2), 187–197. doi:10.1108/IJICC-06-2021-0101
- Kang, P. (2021). Design of English autonomous learning system based on digital signal processing. *Journal of Physics: Conference Series*, 1881(4), 4–20. doi:10.1088/1742-6596/1881/4/042065
- Kobyakova, I., & Shvachko, S. (2016). Teaching translation: Objectives and methods. *Advanced Education*, 9-13.
- Kun, X., Wang, Z., Zhou, Z., & Qi, W. (2021). Design of industrial internet of things system based on machine learning and artificial intelligence technology. *Journal of Intelligent & Fuzzy Systems*, 40(2), 2601–2611. doi:10.3233/JIFS-189252
- Lei, H., & Shao, W. (2020). Application of E-business English translation in multilingual system of picking robot-based on ESP theory. *Nong-ji-hua Yanjiu*, 25(6), 1442–1444.
- Li, D. (2021). Design of English text-to-speech conversion algorithm based on machine learning. *Journal of Intelligent & Fuzzy Systems*, 40(2), 2433–2444. doi:10.3233/JIFS-189316
- Liu, C., & Mu, F. (2021). Design and research of intelligent logistics robot based on STM32. *Recent Advances in Electrical & Electronic Engineering*, 14(1), 44–51. doi:10.2174/2352096513999200718002257

Liu, F. (2021). Design of Chinese-English wireless simultaneous interpretation system based on speech recognition technology. *International Journal of Antennas and Propagation*, 2021, 1–9. Advance online publication. doi:10.1155/2021/7346984

Nguyen, H. D., Tran, T. V., & Pham, X. T. (2022). Design intelligent educational chatbot for information retrieval based on integrated knowledge bases. *IAENG International Journal of Computer Science*, 5(5), 97–109. doi:10.3390/ai4010015

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (pp. 311–318). doi:10.3115/1073083.1073135

Samadi, M., & Jond, H. B. (2021). An intelligent approach for autonomous mobile robot path planning based on adaptive neuro-fuzzy inference system. *Ain Shams Engineering Journal*, 13(1), 101491. doi:10.1016/j.asej.2021.05.005

Song, X. (2021). Intelligent English translation system based on evolutionary multi-objective optimization algorithm. *Journal of Intelligent & Fuzzy Systems*, 40(4), 6327–6337. doi:10.3233/JIFS-189469

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27. Advance online publication. doi:10.48550/arXiv.1409.3215

Wairagkar, M., Lima, M. R., Bazo, D., Craig, R., Weissbart, H., Etoundi, A., Reichenbach, T., Iyengar, P., Vaswani, S., James, C., Barnaghi, P., Melhuish, C., & Vaidyanathan, R. (2021). Emotive response to a hybrid-face robot and translation to consumer social robots. *IEEE Internet of Things Journal*, 9(5), 3174–3188. doi:10.1109/JIOT.2021.3097592

Wen, H. (2020). Intelligent English translation mobile platform and recognition system based on support vector machine. *Journal of Intelligent & Fuzzy Systems*, 38(6), 7095–7106. doi:10.3233/JIFS-179788

Xia, D. (2020). Research on Chinese-English Translation Compensation Based on Cognitive Linguistics. *International Journal of Social Science and Education Research*, 3(4), 235–242. doi:10.6918/IJOSSER.202004\_3(4).0033

Yuan, L. (2020). Interactive system design of entrepreneurship education based on internet of things and machine learning. *Journal of Intelligent & Fuzzy Systems*, 13(88), 1–14. doi:10.3233/JIFS-189053

Zhang, J. (2020). Application of multimedia martial arts system in intelligent picking robot. *Nong-ji-hua Yanjiu*, 18(4), 1–26.