

Research Article

English-Chinese Machine Translation Based on Transfer Learning and Chinese-English Corpus

Bo Xu 

School of Foreign Studies, Huanggang Normal University, Huanggang 438000, China

Correspondence should be addressed to Bo Xu; xu_bo@hgnu.edu.cn

Received 8 June 2022; Revised 29 August 2022; Accepted 3 September 2022; Published 27 September 2022

Academic Editor: Jun Ye

Copyright © 2022 Bo Xu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes an English-Chinese machine translation research method based on transfer learning. First, it expounds the theory of neural machine translation and transfer learning and related technologies. Neural machine translation is discussed, the advantages and disadvantages of various models are introduced, and the transformer neural machine translation model framework is selected. For low-resource Chinese-English parallel corpus and Tibetan-Chinese parallel corpus, 30 million Chinese-English parallel corpora, 100,000 Chinese-English low-resource parallel corpora, and 100,000 Tibetan-Chinese parallel corpora were used to pretrain the transformer machine translation architecture. The decoders are all composed of 6 identical hidden layers, the initialization of the model parameters is done by the transformer uniform distribution, and the model training uses Adam as the optimizer. In the model transfer part, the parameters with the better effect of the pretrained model are transferred to the low-resource Chinese-English and Tibetan-Chinese machine translation model training, so as to achieve the purpose of knowledge transfer. The results show that the model transfer learning of low-resource Chinese-English parallel corpus improves the translation system's translation by 3.97 BLEU values compared with the translation system without transfer learning at 0.34 BLEU values. Model transfer learning on low-resource Tibetan-Chinese parallel corpus increases the BLEU value by 2.64 BLEU compared to the translation system without transfer learning. The neural machine translation system that uses BPE technology for preprocessing plus model transfer learning is compared to the translation system that only performs transfer learning and shows an improved 0.26 BLEU value. It is verified that the transfer learning method proposed in this paper has a certain improvement in the effect of low-resource Chinese-English and Tibetan-Chinese neural machine translation models.

1. Introduction

As an essential element of human communication, with the development of time and international trade, people from all over the world communicate and cooperate more, people in more countries have more relationships, and the need for seamless communication and understanding has become very important. Using machine translation technology to solve language barriers is a valuable tool for solving human-to-human communication problems, and translators have always been the focus of attention among scientists. In today's age of intellectual property, machine translation has become possible, thanks to advances in software technology, the emergence of new algorithms, and improvements in computer performance. Machine translation has been used extensively in modern translation work, and its role and

impact are unpredictable. Some experts even predict that it will replace human translation in the future. Google online translation is a machine translation tool that can indeed help translators solve certain problems. Although it plays a different role in the translation of different texts, it still needs to be edited manually to varying degrees. In order to improve the quality and efficiency of machine translation and reduce the involvement of human translation, this paper focuses on English-Chinese translation and proposes a more efficient machine translation model.

2. Related Works

Song, Q. and others said that China's research in the field of machine translation began in 1957. It is the fourth country in the world to start machine translation research after the

United States, Germany, and the Soviet Union. In 1958, the first machine translation experiment was carried out on domestic 104 large general-purpose digital computers, successfully translating 20 different types of Russian sentences into Chinese [1]. Later, the Harbin Institute of Technology, the China Institute of Science and Information Technology, the South China Institute of Technology, and other organizations also set up the machine. Translate research teams and conduct research on English-Chinese or Russian-Chinese translators. Zhang, X, and others said that the phrases based on the Tibetan-Chinese translator are based on the features of Tibetan morphology and grammar. This article focuses on the use of Tibetan coding conversions and Tibetan automated word distribution in the system. In addition, this article contains important requests and guidelines for research and translation technology related to China, including the integration of the Tibetan Corps and the automatic language distribution of both equal languages [2]. Forty and others claim that “research on the Tibetan-Chinese neural network translator” is the first time that a neural translator has been used for the Tibetan-Chinese translator area [3]. Jin and others said the work used end-to-end models based on cyclic neural networks and monitoring networks, and that migration education was used to start modeling. Good for solving the problem of insufficient data for small models. In-depth case study [4], Ren et al. said that the model method has an improvement of 3 BLEU values compared to phrase-based Tibetan-Chinese machine translation in practical experiments, and this work is indispensable for the study of low-resource neural machine translation. It can be seen from the abovementioned literature that the research units of low-resource neural machine translation in China are mainly concentrated in universities including Northwest University for Nationalities, Tibet University, Harbin Institute of Technology, Inner Mongolia University of Technology, and other universities [5]. Xia and others said that in addition, some state organs and enterprises are also actively developing such software and providing relevant services. For example, China National Language Translation Bureau and Yayi Network Technology Co., Ltd. provide machine translation services between Tibetan-Chinese, Mongolian-Chinese, Uyghur-Chinese, and other language pairs [6]. Ming, N. et al. said that although these system services can temporarily alleviate the society’s demand for machine translation for low-resource languages, these systems are all based on statistical models, compared to the English-Chinese provided by NetEase, Baidu, Google, Sogou, and other enterprises. English-German and other online translation services, there is still a big gap in translation quality [7]. He and others said that how to use a neural network to improve the translation quality of low resource language pairs and minimize this gap is the common concern and focus of current machine translation researchers related to low resource language [8]. The main disadvantage of the academics’ work is that the network has already adapted to the overall input of 300-1000-dimensional vectors before it starts producing outputs. Therefore, some scholars have proposed the so-called attention mechanism. Ji, B. et al said that the attention mechanism

gives the network the ability to reconsider all input words and use this information when generating new words [9]. The previous architecture is redesigned with a convolutional neural network (CNN), which processes all input words together, so it makes the training and reasoning process faster. That same year, Google subversively proposed a neural translator model that left the entire cyclic neural network and convolutional neural network. Lin, L., and others said that the model would also use “encoder-decoder” as the basis for the model. In the structure, multiple head listening methods and feed-forward neural networks are used to design the encoder and decoder structure. The model has achieved impressive results in working as a translator for several languages. One way is to plan for the integration of language structures (LMs) that are learned in the NMT system, that is, only speech data (s). Experimental results show that integration of monolingual corpus can improve translation problems (Turkish-English) and translation problems (Chinese writing in English) [10]. The principle of English-Chinese machine translation is shown in Figure 1.

3. Method

As one of the research centers in natural language processing technology, translation aims to enable computers to correctly understand and translate natural language like people. The development of machine translation technology has been closely associated with the development of computer technology, information theory, linguistics, and other disciplines. It aims to transform data into information that is relevant through communication and decision-making. Among them, translators play an important role in translation. This is a research analysis of how to quickly identify a translator model with high accuracy and robustness [11]. The basic framework of machine translation is shown in Figure 2.

Generally, machine translation can be regarded as the transformation from one sequence to another. Machine translation is a widely recognized and useful example of sequence-to-sequence models, and allows us to demonstrate the difficulties encountered in trying to solve these problems using many intuitive examples. The encoder sequentially encodes the text and deletes the language information in the distribution representation, and then the decoder converts the information representation into a presentation in other languages, as shown in Figure 3.

Figure 3 shows the relationship between the encoder and decoder. First, through the encoder, the source language sequence “ x_1 , x_2 , and x_T ” is encoded by the encoder to generate a vector C representation, and then the vector is sent to the decoder as an input, and the decoder decodes this vector into the target language sequence [12]. During the target language, sequence generation is performed word by word, when a certain word is generated, it depends on the historical information of the previously generated target language until the end of the sentence is generated. Recurrent neural network (RNN): a recurrent neural network that takes sequence data as input, performs recursion in the evolution direction of the sequence and connects all nodes in

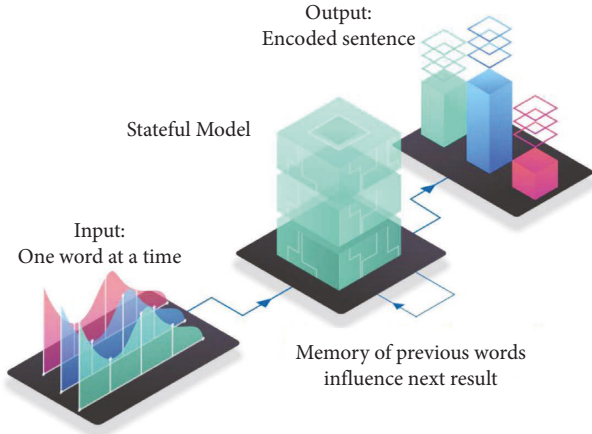


FIGURE 1: Principles of English-Chinese translation.

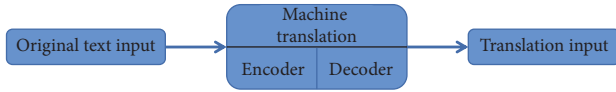


FIGURE 2: Basic framework of machine translation.

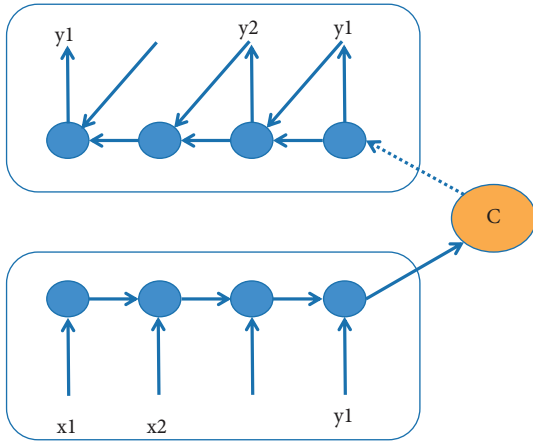


FIGURE 3: Example of encoder and decoder.

a chain. Cyclic neural network is mainly used to process sequence data, especially for variable length sequence data. Its core part is a directed graph. Chained elements in directed graph expansion are called recurrent units [13, 14]. RNN can be thought of as more than one application on the same neural network, and each neural network will transfer data to another location. Where $x = \{x_1, x_2, \dots, x_t\}$ represent different data lengths. At point t , the hidden state h_t is changed by the following formula, as shown in the following formula:

$$h_t = f(h_{t-1}, x_t). \quad (1)$$

f is a bugless operation, U is the weight matrix input to the hidden layer, V is the weight matrix from the hidden layer to the output layer, y is the target sequence to be achieved by the model, L is the loss function, and W is the weight matrix from the hidden layer to the hidden layer. The

time series t is in the range $[1, t]$ and the input x is mapped to the output o by a recurrent neural network. The entire network is transformed by the following model, as shown in the following formula:

$$\begin{aligned} a_t &= Eh_{t-1} + Ux_t + b, \\ h_t &= \tanh(a_t), \\ o_t &= Vh_t + c, \\ \hat{y} &= \text{soft max}(o_t). \end{aligned} \quad (2)$$

The cyclic neural network unifies the length of the input vector of the input sequence of different lengths, and the same parameters and transformation energy can be used at any time point, which is required for operation files of different lengths. In addition, RNN can theoretically capture any precedent the idea of RNN is to use serialized information. In traditional neural networks, we assume that all inputs and outputs are independent of each other, but for many tasks, this assumption is problematic. For example, if you want to predict the next word in a sentence, you need to know which words come before it. LSTM differs from RNN. RNN can only have short memory due to gradient loss. The long-short-term memory network combines short-term memory with long-term memory by introducing gate control, which solves the problem of gradient disappearance to a certain extent. Short-term memory networks ... through directional control." is grammatically unclear. Please rephrase the sentence for clarity and correctness." LSTM is composed of three gate control units, namely, input gate, forget gate, and output gate. The input gate controls the input of the network, the forget gate controls the memory unit, and the output gate controls the output of the network [15]. The memory information at time t is used to save important information. Just like a record book, it saves the knowledge points learned in the past [16]. To control the content of forgetting the cell state of the previous layer, use sigmoid as the activation function, X_t of this sequence as the input, and then according to the h_{t-1} of the previous sequence, get the content of the cell state of the previous layer that needs to be removed and which needs to be retained. It should be noted that the input is in the form of a vector. We expect that the output value of the forget gate is mostly 0 or 1, that is, each value in the vector is completely forgotten or completely reserved, so we choose the sigmoid function as the activation function (6) shows.

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f). \quad (3)$$

The input gate determines how much of the input x_t of the current network is retained to the current time ct . This process is divided into two steps, the first is to use the input gate containing the sigmoid layer to decide what new information is added to the cell state; after determining the new data to be added, it is necessary to convert the new data into a format that can be added to the cell state. Thus, the second step is to use the tanh function to create a new candidate vector, as shown in the following formula:

$$\begin{aligned} i_t &= \sigma(W_f * [h_{t-1}, x_t] + b_f) \\ \tilde{C}_t &= \tanh(W_C * [h_{t-1}, x_t] + b_C) \end{aligned} \quad (4)$$

After the data are checked by the port and the port input, the state of the Ct-1 cell can be adjusted to Ct. As shown in the example below, where $f_i \times C_{t-1}$ represents the information you want to delete and $i \times C_t$ represents the new information, as shown in the following formula:

$$C_t = f_t * C_{t-1+i} * \tilde{C}_t. \quad (5)$$

How much of the control panel Ct is sent to the current output value ht to LSTM. That is, selectively release the contents of the cell state preservation. Like the updated two parts of the front door, the output gate also needs to use the sigmoid activation function to determine which information needs to be output. The cell state is processed through the tanh layer, multiply the two to get the information we want to output. We then use the tanh activation function to make the contents of the cell state and divide it into two sections to get what we want to release, as shown in the following formula:

$$\begin{aligned} ot &= \sigma(W_o[ht - 1, xt] + b_o), \\ ht &= (ot * \tanh(Ct)). \end{aligned} \quad (6)$$

GRU (gated recurrent unit) model is a type of RNN model. Like LSTM, it can intercept relationships with long-distance connections and reduce the likelihood of disappearing or breaking. At the same time, the structure and calculation are simpler than LSTM. GRU merges the forget gate and output gate into an "update gate," which has a very good effect. Therefore, it is also a network structure of very manifold at present. To solve the problem of gradient vanishing and gradient explosion, the method and structure are shown in the following formula:

$$\begin{aligned} Zt &= \sigma(Wz * [ht - 1, xt]), \\ rt &= \sigma(Wz * [ht - 1, xt]), \\ ht &= \tanh(W * [rt * ht - 1, xt]), \\ ht &= (1 - zt) * ht - 1 + zt * \tilde{h}_t, \end{aligned} \quad (7)$$

where rt represents a gateway reset, which is used to determine the level of forgetfulness of previous data. zt means to change the door. The update gate acts like the forget gate and input gate in LSTM. It determines, which data to forget and which new data to add to the neural structure. Each word will be represented as a real vector. This corresponds to a representation model of words. This section mainly introduces the difference between the traditional word representation model and the word representation model based on a real number vector. One hot coding is a traditional word representation method. A hot coding represents a word as a 0-1 vector of uppercase letters, where only the corresponding product for the word is 1, and all other objects are zero. For example, suppose a dictionary contains 10,000 words and numbers. Then each word can be represented as a 10k dimensional one-hot vector. Using *Python*

is an a-explanatory, object-oriented, dynamic data type advanced programming language to solve problems. Only the dimension corresponding to the number is 1, and the other dimensions are 0. The advantage of one hot coding is that the form is simple and easy to calculate, and this representation has a good correspondence with the dictionary, so each code can be interpreted. However, one hot coding regards words as mutually orthogonal vectors. This results in no correlation between all words. Single-thermal encoding is often used to handle features that do not have size relationships between categories. As long as they are different words, they are completely different under one hot coding [17]. For example, one might expect words like "table" and "chair" to have some similarity, but the one-hot encoding treats them as two words with 0 similarities. A distributed representation is used in neural language models. In the neural language model, each word is no longer a completely orthogonal 0-1 vector, but a point in a multi-dimensional real number space, which is embodied as a real number vector. In many cases, this distributed representation of words is also called word embedding. The distributed representation of words can be viewed as a point in Euclidean space, so the relationship between words can also be characterized by the geometric properties of the space. Different words can be represented on a 512-dimensional space. Under this representation, there is a certain connection between "table" and "chair." The traditional machine learning method of natural language processing firstly trains a model for a specific language in a large number of parallel corpora and then applies the machine translation model to the translation task of the specific language. Compared with transfer learning, its basic conditions are no longer required. First, training materials and test data for machine learning standards should be distributed independently and independently; second, the balance in the body used for exercise must be measured and performed to achieve good results. The concept of transformational education allows for the use of existing data to train neural network models and transfer the learned experience to neural network models with less training corpus so that training materials can be reduced. And training time can be reduced. In general machine learning, for various positions, it is necessary to write various registration documents related to the training for attaining their independent standards. Compared to these ideas, learning changes can be a good model in the context of small data [18, 19]. Transfer learning stores the knowledge acquired by training model A and applies it to new tasks. The figure shows the training of model B to achieve the purpose of improving the performance of model B. The transfer learning strategy is very suitable for tasks that lack of existing labeled data. In addition to a small number of languages with rich parallel corpus data resources (such as Chinese, English, and German), the problem of lack of corpus resources in many languages is common, and there is not enough labeled data. The introduction of transfer learning will effectively alleviate this difficulty. Domain-specific machine translation systems are in high demand, while general-purpose machine translation systems have a limited range of applications.

TABLE 1: Comparison of inductive, direct push, and unsupervised transfer learning.

Transfer learning style	Source domain labeling	Labeling of target domain
Inductive transfer learning	Yes	Little
Direct push transfer learning	Yes	No
Unsupervised transfer learning	No	No

Generic systems are generally less performant and therefore important for domain-specific machine translation development [20]. Domain-specific adaptation is a key problem in machine translation. The goal is to study the specific domain of the model. As we all know, special reconstructive models (news, speech, medicine, literature, etc.) have more accuracy in neurological pathogens under the same name. Specifically, when the training data are distributed unbiasedly on the target domain, the final model will be compared against the test data during training on the dev set. Domain adaptation usually includes terminology, domain, and style adaptation. However, if the training data comes from a different source of purpose, the performance will be reduced accordingly. To build well-performing machine learning (ML) models, the model must be trained and tested against data from the same target distribution. For example, when the training data come from news articles and the test domain is specific to the medical domain, the translation performance will be unsatisfactory. We often have a large number of out-of-domain parallel statements. The challenge of training domain-specific models is to improve translation performance in the target domain given only a small amount of additional in-domain data. This can be accomplished by modifying the structure with special data (also called continuous training). Domain adaptation has been used successfully in computing and neural translation. In a typical neural machine translation domain adaptation setting, we first train the parent model on a resource-rich out-of-domain parallel corpus. On the basis of the general model, the training corpus is converted into an in-domain corpus and the parent model is fine-tuned. We can think of domain adaptation as transfer learning from an out-of-domain parent model to a domain-specific child model [21]. However, in real scenes such as online translation engines, the domain of sentences is not given. Guessing the domain of input sentences is very important for correct translation. In order to solve the problem of lack of data in the domain, the domain of a single sentence in the training data can be classified, and then the training sentences close to the target domain can be searched and selected. Inductive migration: the learning tasks of the source domain and the target domain are different but related. The labeled data of the target domain are available, but the labeled data of the source domain are not necessarily available. According to whether the label data of the source domain are available, it can be further divided into multitask learning (labeled data available) and self-learning (labeled data not available). Direct push transfer: when the target task and the source task are the same, the target domain data are unlabeled, but there is a large amount of available labeled data in the source domain. In this case, it is assumed that the tags of the same instance are the same across different

domains, meaning that the whole case of the same instance does not depend on the author. Unsupervised migration: the registry and destination functions are different but related, and there is no script, as shown in Table 1.

The main idea of instance-based transfer learning is to reduce the difference between the source domain and the target domain by changing the existing form of the samples, which is mainly suitable for situation where the similarity between the source domain and the target domain is high. The main idea of migration based on feature representation is to find a better feature representation, minimize the difference between domains and the error of classification and regression, and make the source domain and target domain show similar properties in a certain feature space through feature transformation, which can be applied to the case, where the similarity between domains is not too high or even dissimilar, it can be divided into supervised and unsupervised situations [22]. The transfer method based on model parameters assumes that the models on related tasks can share some parameters from the perspective of the model, so as to share some parameters between the source domain model and the target domain model to achieve the effect of transfer learning. The relationship-based transfer achieves the effect of transfer learning by establishing a map of the correlation knowledge between two domains. It does not assume that the data in each domain is independent and identically distributed, but transforms the relationship between the data from the source domain is migrated to the target domain, as shown in Table 2.

Isomorphic transfer learning: its source domain and target domain have the same feature space, that is, its feature dimension is the same, but its feature distribution is different, see Table 3 for details. The realization of isomorphic transfer learning needs to solve the problem of domain adaptive learning. Commonly used methods include example weighted domain adaptive learning, feature representation domain adaptive learning, parameter and feature decomposition domain adaptive learning, multisource domain adaptive learning, and heterogeneous learning. Transfer learning: the feature space, feature dimension, and feature distribution of the source and target domains are different. Therefore, realizing heterogeneous transfer learning needs to solve the problem of feature space alignment first, and then solve the problem of domain adaptive learning, which is more complicated than homogeneous transfer learning.

4. Experiment and Analysis

The NMT model represents a sentence as a long vector in a sentence, but the long vector does not represent the entire

TABLE 2: Comparison of four kinds of transfer learning based on examples, features, models, and relationships.

Transfer learning style	Characteristic
Case-based transfer learning	Give the source domain instance a certain weight and reuse it
Feature-based transfer learning	Reducing the gap between source domain and target domain based on feature transformation
Model-based transfer learning	Find out the shared parameters between the source domain and the target domain network model
Relationship-based transfer learning	Mining the relationship similarity between different fields

TABLE 3: Comparison between isomorphic transfer learning and heterogeneous transfer learning.

Transfer learning style	Source domain and target domain feature space	Problems to be solved
Isomorphic transfer learning	Identical	Domain adaptive learning
Heterogeneous transfer learning	Different	Adaptive learning of its + domain in feature space

semantic data of the sentence. The NMT-based monitoring process first encodes the sentence-by-sentence vector sequence, and then dynamically searches for relevant information related to word creation through the monitoring process while developing languages, which NMT's capabilities make much better. This document provides Chinese-English and English-Chinese translation standards required by Klein. First, two preschool courses (A and B) focused on large-scale Sino-English parallel corpora and more than one Anglo-Chinese parallel corpora; second, during the training of the Sino-English NMT model, the encoder parameters of the Sino-English translation standard are started by the encoder parameters of the Sino-English standard, and the decoder parameters of the Sino-English standard translation unit are started by the decoder. English-Chinese model endless is an excellent model to achieve the final TINMT_CV model (C) starting with the Sino-English parallel corpus [23]. It can be assumed that the final result (BLEU value) of the Sino-English and Tibetan-Chinese neural translation using the transition model is better than the traditional translation without exchange of knowledge. Based on the extensive training and adaptation of the Sino-English model, BLEU's rate of improvement at an early stage was faster than the standard translation that relied on Sino-English and Tibetan-Chinese text for the neural network. Chinese-English resource training is used when using the BLEU value of 25 as the target for the joint corps. With standard modifications, standard translation can be completed in 20,000 steps, but without standard translation, standard translation can be learned in 80,000 steps. When the BLEU value reaches 40, the target point of Tibetan sugar is less when combining corpus training. However, with standard definition, interpretation can be completed in 50,000 steps, as shown in Figure 4.

Due to the influence of parameter initialization before machine translation model training, the parameters of large-scale Chinese-English translation model belonging to the same translation task are introduced into the initialization of low-resource Chinese-English and Tibetan-Chinese translation models so that the model has a certain parameter basis before training, so its learning rate will be improved during retraining [24, 25]. In this document, the encoder and decoder parameters of the Chinese-English translation model are initialized together with the parameters of the Chinese encoder of the Sino-English model and the decoder of the

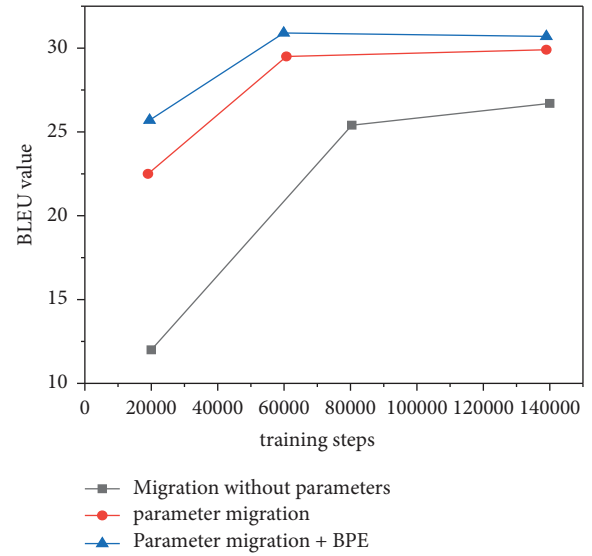


FIGURE 4: Relationship between low resource Chinese-English BLEU value and training steps.

English-Chinese model. As a basis for this, the small size of the Sino-English bilingual corpus for good training is used to achieve the Sino-English NMT standard. To improve the relationship between the encoder and the decoder received by the pretraining and to ensure that the initialization is better for good training, this article presents the training before testing. Firstly, the pivot language English is reinvigorated in the existing Chinese-English training set, and the large-scale English-Chinese parallel corpus is used to train the English-Chinese translation model; Then we use the English-Chinese translation model to retranslate the English in the English-Chinese parallel corpus, so as to obtain the Chinese-English-Chinese trilingual parallel corpus; then use the method of data enhancement 16 to increase the Chinese-English parallel corpus, improve the correlation between the model parameters, and reduce the existing noise. In this experiment, a Chinese English parallel corpus with a scale of 100,000 sentence pairs is used, of which 13,000 sentence pairs are tested and 11,000 sentence pairs are verified; 700,000 pairs of English-Chinese parallel corpora, including 5,000 pairs of test corpora and 4,000 pairs of verification corpora; There are 50 million pairs of Chinese-English parallel corpora, including 30,000 pairs of test

corpora and 10,000 pairs of verification corpora. Before the training, the experimental data are filtered for garbled code and word segmentation. In order to evaluate the effectiveness of the TINMT_CV model, the experiment selects five baseline systems Moses, transformer, CNN, NMT trans, GNMT, and the TINMT_CV model proposed in this paper. A total of 120,000 English-Chinese parallel corpora are used as training sets in the direction of English-Chinese translation. The terms used by the transformer, TINMT_CV, and NMT trans model are set to 32000, the maximum number of lines is set to 50, “transformer_ff” is set to 2048, “lab horizontal equalization” is set to 0.1, “led head” is set to 0.1, set to 2, “dropout” is set to 0.2, the number of layers is set to 2, the word embedded dimension is set to 256, “batch size” is set to 128, and the teaching value is set to 0.2. The optimizer selects Adam, with “NUM units” set to 128 and “dropout” set to 0.2. In this article, the two-dimensional high-efficiency test (BLEU) is used as a measurement tool. Table 1 shows the comparison results of the BLEU values between the baseline system and the TINMT_CV model in both English-Chinese and Chinese-English translation directions. Among them, the TINMTe is the TLNMT_CV model, which is only pretrained encoder, and the TLNMTd is the TINMT_CV model, which is only pretrained in the measurement encoder. It can be seen from the experiment that the results of the TLNMT_CV model of the Anglo-Chinese bilingual NMT are better than the basic process, of the TLNMTe bi model. Compared to Moses’ example, the EU rate increased by 1.52% for English-Chinese translations and 1.31% for Chinese-English translations. Compared with the transformer model, the BLEU value of the TLNMTe model increased by 0.38 percentage points in the direction of the English-Chinese translation and 0.44 percentage points in the area of introduction of the Chinese-English translation. By quality, U-value for the TINMT_CV model in the direction of English-Chinese translation is 0.71% higher content than the NMT trans model and 0.48% higher content in the introduction of Chinese-English translation. The TINMT_CV model is used in the direction of English-Chinese translation. The EU rate increased by 1.16 percentage points compared to standard manpower and 1.05 percentage points in the direction of Chinese-English translation. This article presents Ms. TLNMT CV method, which can guide the first error in the Chinese-English NMT encoder and decoder using large-scale Chinese-English and English-Chinese corpora and can accept Chinese-English NMT standards through small-scale Chinese-English fine-tuning training. This method can improve the performance of low-resource Chinese-English NMT. Comparative experiments also proved the effectiveness of this project. In the next step, we can explore the widespread use of Chinese-English monolingual corpus for pretraining, and the knowledge gained from pretraining of the Sino-English bilingual NMT model to improve translation efficiency. Can be integrated into the construction. In this section, large Chinese-English corps are trained 200,000 steps to achieve stable standards, 100,000 steps are trained for rare Chinese-English and Tibetan-Chinese materials, and 5,000 for comparison experiments. The BLEU value of the steps was

TABLE 4: Chinese-English translation of low resources.

Model framework	BLEU
Transformer (not migrated)	27.39
Transformer (migration)	31.36
Transformer + BPE	31.68

TABLE 5: Low resource Tibetan-Chinese translation.

Model framework	BLEU
Transformer (not migrated)	46.02
Transformer (migration)	48.53
Transformer + BPE	48.92

recorded. Table 4 compares the benefits of educational change based on the function of the Chinese neurotransmitter. The Table shows the training test results under the English material resources in 10 W, as shown in Tables 4 and 5.

The comparison results of machine translation models are shown in the table mentioned above. It can be seen that the model transfer learning of low resource Chinese and English parallel corpora improves the translation of the nontransfer learning translation system by 3.97 BLEU values, and the translation of the translation system pretreated with BPE technology improves the translation of the translation system by 0.34 BLEU values compared with the translation system with only transfer learning. The model transfer learning of low resource Tibetan-Chinese parallel corpus improves the translation value of the nontransfer learning translation system by 2.64 BLEU values, and the neural machine translation system with BPE technology preprocessing and model transfer learning improves the translation value by 0.26 BLEU values compared with the translation system with only transfer learning. NMT is a typical encoding and decoding structure, in which the encoder reads the entire sentence sequence and encodes it to obtain the vector table of the sentence. The decoder uses the sentence vector obtained by the encoder as the target input and generates the words of the target language word by word. Sequence transfer learning can transfer the parameters learned by the model to similar tasks, and use the parameters obtained from high-resource translation tasks to improve the performance of low-resource translation tasks, thereby reducing the translation task’s dependence on parallel data, but fixed-length vectors cannot be used. Fully express the semantic information of the sentence in the source language. However, the semantic information of a sentence cannot be fully expressed in the source language using a fixed-length vector. The NMT-based monitoring process first encodes sentence by sentence into vector sequences, and then dynamically searches for contextual information related to word generation through the language development monitoring process, which greatly enhances the capabilities of NMT.

5. Conclusion

With the application of artificial intelligence and deep learning technology in more and more fields, machine

translation, as an important part of natural language processing, frequently appears in people's daily life applications, which is of great research value. At this stage, the mainstream machine translation methods have turned from traditional statistical methods to deep neural network methods. The main work is divided into the following parts: by reading Chinese and foreign literature related to machine translation, consulting reference materials, learning neural machine translation technology, fully understanding the main technologies proposed by academia and industry in the field of neural machine translation, and the application of these technologies, compare the proposed background, application scenarios, advantages and disadvantages of each model, learn various machine translation models according to the introduction of the references, and fully understand the multiangle knowledge of machine translation. Through the research on various neural machine translation methods, it is found that when using pretraining deep learning technology to initialize the model, obtaining a high-quality pretraining model greatly affects the translation effect of the neural machine translation model, because pretraining is a pretrained and saved network that was previously trained on a large dataset, we can use the pretrained model as a feature extraction device for transfer learning. When the features learned by the pretraining model are easy to generalize, transfer learning can get better results. When using deep learning technology to deal with text translation problems, it is first necessary to convert the text into word vectors. The traditional recurrent neural network word vectors can only represent the frequency of occurrence of different words and the co-occurrence relationship between words, although the co-occurrence relationship is to a certain extent, it reflects the correlation between words, it still cannot accurately reflect the contextual relationship, which affects the accuracy of the algorithm for text translation. To solve this problem, this paper uses a model-based transfer method. First, the Chinese-English parallel with sufficient training data are used. The corpus task trains the transformer machine translation model, and then the model parameters are transferred to the model training of low-resource Chinese-English and Chinese parallel corpora. In this process, the idea of model transfer is used, that is, the parameters of the machine translation model trained on the massively parallel corpus are transferred to the training of the low-resource neural machine translation model, thereby improving the accuracy of the low-resource neural machine translation. On the other hand, the traditional recurrent neural network structures RNN, LSTM, and GRU have complex structures, many model parameters, cannot process data in parallel, and are difficult to train. Therefore, this paper uses the transformer model based on the attention mechanism for model training, which speeds up the training speed, and improves the translation effect. Then, experiments are used to demonstrate that the proposed low-resource neural machine translation method based on model transfer has higher

translation accuracy than the untransferred neural machine translation method.

Data Availability

The data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This work was supported by the First-Class Course Foundation of Huanggang Normal University (no. 2020CK07).

References

- [1] Q. Song, N. Zhang, and H. Liang, "Review of the Chinese internet philanthropy research (2006-2020): analysis based on citespace," *The China Nonprofit Review*, vol. 13, no. 1&2, p. 4, 2021.
- [2] X. Zhang, "Research on the language characteristics of agricultural English and its translation strategies," *Journal of Higher Education Research*, vol. 3, no. 1, pp. 5-8, 2022.
- [3] L. Forti, "Review of lu (2017): a corpus study of collocation in Chinese learner English," *International Journal of Learner Corpus Research*, vol. 8, no. 1, pp. 144-149, 2022.
- [4] L. Jin, "Research on pronunciation accuracy detection of English Chinese consecutive interpretation in English intelligent speech translation terminal," *International Journal of Speech Technology*, vol. 5, pp. 1-8, 2021.
- [5] R. Qing-dao-er-ji, Y. L. Su, and W. W. Liu, "Research on the lstm Mongolian and Chinese machine translation based on morpheme encoding," *Neural Computing & Applications*, vol. 32, no. 1, pp. 41-49, 2020.
- [6] D. Xia, "Research on Chinese-English translation compensation based on cognitive linguistics," *International Journal of Social Science and Education Research*, vol. 3, no. 4, pp. 235-242, 2020.
- [7] N. Ming, "Characteristics and trends of English testing research in China: visual analysis based on citespace," *Region - Educational Research and Reviews*, vol. 3, no. 3, pp. 1-6, 2021.
- [8] W. He and L. Chen, "A Research of Neural Style Transfer on Line Structure Based on Sequence to Sequence Learning," *IEEE Access*, vol. 8, no. 99, pp. 112309-112322, 2020.
- [9] B. Ji, Z. Zhang, X. Duan, M. Zhang, B. Chen, and W. Luo, "Cross-lingual pre-training based transfer for zero-shot neural machine translation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 1, pp. 115-122, 2020.
- [10] A. Kumar, A. Pratap, A. K. Singh, and S. Saha, "Addressing domain shift in neural machine translation via reinforcement learning," *Expert Systems with Applications*, vol. 201, Article ID 117039, 2022.
- [11] X. Wu, Y. Xia, J. Zhu, L. Wu, S. Xie, and T. Qin, "A study of BERT for context-aware neural machine translation," *Machine Learning*, vol. 111, no. 3, pp. 917-935, 2022.
- [12] K. Xu, "Needs analysis of Chinese English majors in egg courses," *Journal of Language Teaching and Research*, vol. 12, no. 3, pp. 452-465, 2021.

- [13] X. Qiao and R. Hao, "Move-step analyze of the introduction in Chinese and English accounting journals: based on analysis of accounting research and journal of accounting and economics (2016-2018)," *Open Journal of Modern Linguistics*, vol. 9, no. 6, pp. 540–550, 2019.
- [14] J. Luo and D. Li, "Universals in machine translation?:a corpus-based study of Chinese-English translations bywechat translate," *International Journal of Corpus Linguistics*, vol. 27, no. 1, pp. 31–58, 2022.
- [15] N. Peng, "Research on the effectiveness of English online learning based on neural network," *Neural Computing & Applications*, vol. 34, no. 4, pp. 2543–2554, 2021.
- [16] X. Zhao, "Research on the English translation of Chinese public signs based on face theory," *OALib*, vol. 8, pp. 1–7, 2021.
- [17] S. Huang, "A contrastive analysis of English novel the notebook and its Chinese translation from the perspective of rewriting theory," *Advances in Social Sciences Research Journal*, vol. 7, no. 12, pp. 320–332, 2020.
- [18] B. Zhang, "Analysis of web-based college English teaching from 2000 to 2017 in China," *English Education Journal of English Teaching and Research*, vol. 5, no. 1, pp. 1–12, 2020.
- [19] X. Geng, L. Wang, X. Wang et al., "Learning to refine source representations for neural machine translation," *International Journal of Machine Learning and Cybernetics*, vol. 13, no. 8, pp. 2199–2212, 2022.
- [20] F. Xu, X. Zhang, Z. Xin and Alan Yang, and A. Yang, "Investigation on the Chinese text sentiment analysis based on convolutional neural networks in deep learning," *Computers, Materials & Continua*, vol. 58, no. 3, pp. 697–709, 2019.
- [21] X. D. Li and H. H. Cao, "Research on vr-supported flipped classroom based on blended learning — a case study in learning English through news," *International Journal of Information and Education Technology*, vol. 10, no. 2, pp. 104–109, 2020.
- [22] A. Poncelas, G. M. D. B. Wenniger, and A. Way, "Improved feature decay algorithms for statistical machine translation," *Natural Language Engineering*, vol. 28, no. 1, pp. 71–91, 2022.
- [23] Y. Chang, "A research proposal on applying Chinese phonetic system in teaching pronunciation of English words to older Chinese efl adult learners," *Journal of Higher Education Research*, vol. 3, no. 1, pp. 21–25, 2022.
- [24] M. Ramos, "Teaching English for medical translation: a corpus- based approach," *Language Teaching Research Quarterly*, vol. 8, no. 2, pp. 25–40, 2020.
- [25] R. Baruah, R. K. Mundotiya, and A. K. Singh, "Low resource neural machine translation: Assamese to/from other indo-aryan (indic) languages," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 1, pp. 1–32, 2022.