A Study on Chinese-English Machine Translation Based on Transfer Learning and Neural Networks

Congli Li

School of Foreign Languages, Zhanjiang University of Science and Technology, Zhanjiang 524094, China

Correspondence should be addressed to Congli Li; dnuria915@mustangs.deltacollege.edu

Received 27 January 2022; Revised 16 February 2022; Accepted 22 February 2022; Published 18 March 2022

Academic Editor: Zhiguo Qu

Copyright © 2022 Congli Li. 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.

The existing Chinese-English machine translation has problems such as inaccurate word translation and difficult translation of long sentences. To this end, this paper proposes a new machine translation model based on bidirectional Chinese-English translation incorporating translation knowledge and transfer learning, and the components of this model include a recurrent neural network-based translation quality assessment model and a self-focused network-based model. The experimental results demonstrate that our method works better on the dataset of machine translation quality assessment task for Chinese-English translation with more information, and the Pearson correlation coefficient of its quality assessment feature vector (such as word prediction vector representation) is higher.

1. Introduction

Language, as the primary means of human communication, is one of the most important tools of communication. With the development of the development of the times and economic globalization, people around the world are communicating and cooperating more and more frequently. The language barrier is becoming more and more serious and obvious, and the need for seamless communication and understanding is becoming crucial. The need for seamless communication and understanding has become crucial. Machine translation has been a hot topic of interest for researchers as an effective means to solve the cross-language communication barrier. As an effective means to address the barriers to cross-language communication, machine translation has been a hot topic of interest for researchers. With the great progress of deep learning research, neural network-based neural machine translation (NMT) has gradually emerged [1–3]. At present, many Internet companies provide machine translation services, such as Baidu, Netease, Google, and Sogou. Although the quality of machine translation still has a certain degree of improvement compared with that of professional translators, there is still a certain gap between the quality of machine translation and professional translators’ translation, but in the scenarios where the quality of translation is relatively low for specific languages in specific fields or for specific languages in specific fields or in scenarios with relatively low requirements for translation quality, machine translation can effectively complete the task and can satisfy some people’s needs to a certain extent. Neural machine translation is a method that uses deep-learning neural networks to obtain the mapping relationships between natural languages. NMT uses a state vector connecting the encoder and decoder to describe the semantic equivalence of relationships. Sentences of a language are vectorized and passed through the network layer by layer, transformed into an expression that the computer can understand, and then translated into a form that the computer can understand. After the sentences of one language are vectorized and passed through the network layer by layer, they are transformed into a form of expression that the computer can “understand,” and then translated into a form that the computer can “understand.” The translation method of “understanding language and generating translation” is realized. The biggest advantage of this translation method is that the translation is smooth, more grammatical norms, and easy to understand. It is a “leap
forward” in quality compared to previous translation technologies. According to Sun et al. [4], there are 7097 living languages in the world. The proposal of neural machine translation provides a faster and more accurate translation method for machine translation. However, most language pairs have only a few hundred to thousands of parallel sentences. The lack of data is a serious problem for training a suitable machine translation system, because both neural machine translation (NMT) and statistical machine translation (SMT) are highly dependent on data. The data dependency of both NMT and SMT is high.

As mentioned above, the performance of current neural machine translation models is heavily dependent on the quality of the parallel corpus and size. However, the current high-quality and large-scale parallel corpus is very limited for most language pairs; the performance of neural machine translation models in these scenarios is somewhat limited. Therefore, it is important to investigate how to use limited corpus resources to train and improve neural machine translation models.

Therefore, this paper conducts a series of research on the existing neural machine translation methods with limited parallel corpus resources. Specifically, we first train the model for the Chinese-English parallel corpus with abundant parallel corpus resources. Then, the best model parameters of the neural machine translation model are investigated and trained. Model is generalized, and the trained model parameters are transferred to the parallel corpus with scarce. Finally, we investigate how to expand the corpus by using the existing language resources, combining with data transfer techniques, and applying them to a single corpus to help the training of neural machine translation models.

This study can give theoretical and data support to the teaching of mutual translation, translation technology, and engineering applications for people in low-resource languages to regions and provide new ideas to improve the research in the direction of low-resource neural machine translation. Meanwhile, the research on machine translation in low-resource language regions can not only promote the language cultural and technical exchanges but also will effectively preserve the language and culture of low-resource language regions, promote our multilingual and cross-platform technology research, help promote the popularization problems that are often encountered.

2. Related Work

With the explosion of deep learning, Internet companies have started to join the research of machine translation and begin to provide related services. Baidu translation currently achieves mutual translation of 28 languages; Tencent launches the translation robot “Translator”; the neural machine translation model developed by Alibaba is officially online. In [5], a phrase-based Tibetan-Chinese statistical machine translation system is built for English lexical and syntactic characteristics, and the implementation methods of English coding conversion and English automatic word separation in the system are described in the paper. The paper in [6] is the first application of neural machine translation to the field of Tibetan-Chinese machine translation. In this work, the authors used an end-to-end model based on recurrent neural networks and attention networks and used transfer learning to initialize the model parameters to alleviate the data scarcity problem of small sample data in deep learning. The first end-to-end NMT system was developed by [7]. They used a long short-term memory (LSTM) cell recurrent neural network (RNN) model. The so-called attention mechanism was proposed by [2]. The attention mechanism gives the network the ability to reconsider all input words and use this information when generating new words. [8] redesigned the previous architecture with a convolutional neural network (CNN) that processes all input words together, thus making the training and inference process faster. [9] effectively integrated language models (LMs) trained only on monolingual data (target language) into the NMT system, and experimental results showed that integrating monolingual corpora can improve translation problems (Turkish English) and domain-constrained translation problems (Chinese-English SMS chat) in translation systems.

3. A Neural Machine Translation Approach Based on Transfer Learning

3.1. Transfer Learning Knowledge Background. In most tasks such as machine learning and deep learning, we assume that when training and testing, the data taken used to obey the same distribution and have the same feature space. However, in reality, this assumption is difficult to the following problems that are often encountered.

(1) The number of labeled training samples is limited. For example, when dealing with XY language pair information, the training resources of XY language pairs are not sufficient. At the same time, the XZ or YZ language pair associated with the XY language pair has a large number of training samples, but the YZ language pair is in a different feature space or the samples obey different distributions than the XY language pair.

(2) The data distribution can change. The data distribution is related to location, time, or other dynamic factors. As the dynamic factors change, the data distribution will also change, and the previously collected dataset will become obsolete and needs to be recollected.

In this case, knowledge transfer is a good choice, i.e., transferring knowledge from domain X to domain Y to improve the training effect of domain Y without spending a lot of time to label the data in domain Y. Transfer learning is proposed as a new learning approach to solve this problem.

As shown in Figure 1, in general machine learning, for different tasks, a large amount of task-related labeled data needs to be collected for training to obtain respective
independent models. Compared to this strategy, transfer learning can achieve relatively better models with a small amount of labeled data. Transfer learning stores the knowledge acquired by training model A and applies it to a new task, and the figure shows the training of model B for the purpose of improving the performance of model B. Figure 2 shows the schematic diagram of the transfer learning model.

The transfer learning strategy is very suitable for tasks where there is a lack of existing tagging data. At present, except for a small number of languages with abundant parallel corpus data resources (e.g., Chinese-English and English-German), the problem of lack of corpus resources prevails for many languages without sufficient tagging data, and the introduction of transfer learning will effectively alleviate this difficulty.

3.2. Domain Adaptation. Domain-specific machine translation systems are in high demand, while general-purpose machine translation systems have a limited. Generic systems usually have poor performance, so it is important to develop machine translation for specific domains [4]. Domain adaptation is a key problem in machine translation, and the goal is to investigate domain-specific studies dedicated to this model. It is well known that specific types of optimization models (news, speech, medical, literary, etc.) obtain higher accuracy for neural systems on a given domain [6, 10]. Specifically, when the training data are assigned without bias on the target domain, the final model will be compared to the test data during the training of the development set. Domain adaptation typically includes term, domain, and style adaptation. However, if the training data comes from a different domain than the target domain, the performance will be reduced accordingly. For example, when the training data is from news articles and the test domain is specific to the medical domain, the translation performance is not satisfactory. We often have a large number of out-of-domain parallel statements. The challenge of training a domain-specific model with only a small amount of additional in-domain data is to improve the translation performance in the target domain. This can be handled by fine-tuning a generic model with domain-specific data (also called continuous training).

Figure 1: Schematic of traditional machine learning model.

Figure 2: Schematic diagram of the transfer learning model.

4. A Machine Translation Quality Assessment Model Incorporating Bidirectional Translation Knowledge

In this section, we introduce the machine translation quality evaluation model integrating bidirectional translation knowledge in detail. As shown in Figure 3, the whole model is based on the "predictor-evaluator" architecture, where the input of the model is the source sentence and the system output translation to be evaluated, and the output is the predicted quality label corresponding to the system output translation to be evaluated, i.e., the HTER value.

The overall machine translation quality assessment model incorporating bidirectional translation knowledge consists of two modules, namely, the feature extraction module and the quality assessment module. Among them, the feature extraction module consists of two symmetric word prediction submodules, namely, the word prediction submodule from the source language to the target language speech direction and the reverse word prediction submodule from the target language to the source language speech direction, which are used to extract the feature vector representation corresponding to each word in the input translation of the system to be evaluated and the feature vector representation corresponding to each word in the source sentence, respectively. The quality evaluation module consists of two two-way recurrent neural networks with the same structure, which compress the feature vector representation of each word in the source sentence and the feature vector representation of each word in the system output translation into a fixed dimensional feature vector representation and measure the "quality" of the source sentence and the "quality" of the system output translation. The vector representation at the source and the vector representation at the target are then stitched together to obtain a real value through a fully connected layer of the parameter matrix, and finally, the real value is bounded between $[0, 1]$ by an S-shaped function, which is the predicted HTER value. The reason for the qualification operation here is that the value of the quality label HTER is specified between $[0, 1]$ in the machine translation quality assessment task. The specific structures of the feature extraction module and the quality evaluation module are described in detail in the next section.

4.1. Feature Extraction Module. The feature extraction module consists of two symmetric word prediction submodules,
i.e., the source-to-target direction word prediction (denoted as src-tgt word predictor) submodule and the target-to-source direction word prediction (denoted as tgt-src word predictor) submodule in Figure 3. Both submodules can be considered as a modified neural machine translation model, with the only difference being the language direction of the training data. The former is trained with a parallel corpus from the source language to the target language speech direction, while the latter is trained with a parallel corpus from the target language to the source language speech direction. This also allows both submodules to model the source-side sentences and the output translations of the system to be evaluated in the bilingual direction. In addition, the word prediction submodule can be based on a recurrent neural network or a self-attentive network.

4.2. Recurrent Neural Network-Based Feature Extraction Module. In the specific implementation, the start identifier <SOS> and the end identifier <EOS> are added to the input sentences at the source side and the target side, respectively. Unlike the word prediction module of the predictor-estimator model, the forward recurrent neural network at the decoding end does not share the context vector with the backward recurrent neural network at the decoding end. In other words, the forward recurrent neural network at the decoding end and the backward recurrent neural network at the decoding end share all parameters independently except for the word vector at the target end. Such a design ensures that when computing the context vector corresponding to the forward recurrent neural network, it only relies on the information from the start position of the sentence at the target end to the current position, while when computing the context vector corresponding to the backward recurrent neural network, it only relies on the information from the end position of the sentence at the target end to the current position. Therefore, the probability is calculated as shown in Equation (1):

$$
P(y_j | y_{<j}, X; \theta) = \text{softmax} \left( \left[ w'_{j+1}, w'_{j+1} \right], \left[ \tilde{s}_j, \tilde{s}_j \right], \left[ \tilde{c}_j, \tilde{c}_j \right] \right) \frac{\exp \left( \left( u_j^r \right)^T W_3 t_j'' \right)}{\sum_{k=1}^{K_j} \exp \left( \left( u_k^r \right)^T W_3 t_k'' \right)} \tag{1}$$

Among them, $[\tilde{c}_j; \tilde{c}_j]$ represents the spliced representation of the context vector corresponding to the forward recurrent neural network and the context vector corresponding to the backward recurrent neural network at the decoding end; $\tilde{s}_j$ is used to measure the correlation between $\tilde{s}_{j+1}$ and the hidden layer representation corresponding to each word at the source end, and $\tilde{s}_j$ is used to measure the correlation between $\tilde{s}_{j+1}$ and the hidden layer table corresponding to each word at the source end, which can be calculated by the attention module. The formula for the intermediate representation $t_j''$ is shown below:

$$t_j'' = U_3 [\tilde{s}_j; \tilde{s}_j] + V_3 [E w_j' + E w_j'' + C_3 [\tilde{c}_j; \tilde{c}_j]] \tag{2}$$

where $W_3 \in \mathbb{R}^{K \times l}$, $U_3 \in \mathbb{R}^{K \times 2m}$, $V_3 \in \mathbb{R}^{K \times 2m}$, $C_3 \in \mathbb{R}^{K \times 4m}$, $m$ denotes the dimensionality of the word vector, and $n$ denotes the number of hidden layer units of the forward recurrent neural network and the backward recurrent neural network at the decoding end. In the specific implementation, the original two parameter matrices $W_3$, $W_2$, and $W_1$ are combined into one parameter matrix. The variables that are not described remain consistent with the meaning of the variables in the previous section. Similarly, the feature vector $q_{ij}$ is designed using the information contained in the intermediate representation. i.e.,

$$q_{ij} = \left( \left( u_j^r \right)^T W_3 \right) \Theta t_j'' \tag{3}$$

Similarly, the only difference between the symmetric target-side language-to-source-speech direction word prediction module is that the input on the encoding side is the system output translation to be evaluated, and the input on the decoding side is the source-side sentence. It should be noted that, in the specific implementation, the two recurrent neural network-based word prediction modules need to be pretrained with a parallel corpus from the source language to the target language speech direction and a parallel corpus from the target language to the source language speech direction, respectively. The parallel corpora used here are the same, except that parallel corpora of different language directions are used in the training process of the different word prediction modules. In this way, for the source-side sentences to be evaluated and the output translations of the machine translation system, the corresponding feature vector representations at the source and target ends can be automatically learned by the two pretrained symmetric word prediction modules. Three specific types of feature vector representations are designed in this paper, taking the feature vector representation corresponding to the $j$th word at the decoding end in the source-to-target language word prediction module as an example:

1. Word Vector Representation at the Decoding End.

   The word vector representation corresponds to the
previous word and the next word of the currently predicted word at the decoding end \([Ew_{i-1};Ew_{i+1}]\).

(2) Intermediate Vector Representation at the Decoding End. The intermediate vector representation is used to predict the current word at the decoding end \(i^j\).

(3) Word Prediction Vector Representation at the Decoding End. The word prediction vector representation \(q_y\) used to predict the current word at the decoding end contains both the intermediate vector representation for predicting the current word and the information of the current predicted word, and it is not difficult to find that softmax\((\text{sum}(q_y))\) means the probability of predicting the current word.

The target-side language to source-side language inverse word prediction module accordingly extracts the feature vector representation corresponding to the decoding side, i.e., the source-side sentence to be evaluated. At this point, the feature vector representations corresponding to the source-side sentences to be evaluated and the system output translations can be obtained separately as the input to the next step of the quality evaluation module [11–14]; all words that are not in the vocabulary (out-of-vocabulary, OOV) are mapped to the special identifier \(<\text{UNK}>\) [15–17]. In the sentence-level machine translation quality assessment task, the quality label score prediction task (scoring) uses Pearson’s correlation coefficient, mean average error (MAE), and root mean squared error (RMSE) as evaluation metrics. Root mean squared error (RMSE) was used as evaluation indicators [18–21].

4.3. Quality Assessment Module. The feature extraction module automatically learns the feature vector representations, collectively called Quality Estimation Feature Vectors (QEFVs). The QEFVs can be considered as the link between the feature extraction module and the quality estimation model, which is used to transfer the bidirectional translation knowledge learned from the parallel corpus into the quality estimation module. The QEFV module consists of two bidirectional recurrent neural networks with the same structure, which compress the feature vector representation corresponding to each word in the source sentence and the feature vector representation corresponding to each word in the system output translation into a set of fixed dimensional hidden layer representations, specifically using the bidirectional LSTM model in this paper. That is,

\[
\hat{h}_{1:T_x}, \hat{h}_{1:T_x} = \text{Bi-LSTM} \left( \{ \text{QEFV}_{i=1}^{j=T_x} \} \right),
\]

\[
\hat{h}_{1:T_y}, \hat{h}_{1:T_y} = \text{Bi-LSTM} \left( \{ \text{QEFV}_{j=1}^{i=T_y} \} \right).
\]

Here, the two-way hidden layer representation at the target and the two-way hidden layer representation at the source are averaged separately, i.e., two fixed dimensional compressed vector representations can be obtained. The specific calculation method is shown as follows:

\[
h_{\text{src}} = \frac{1}{T_x} \sum_{j=1}^{j=T_x} \left( \hat{h}_j ; \hat{h}_j \right),
\]

\[
h_{\text{tgt}} = \frac{1}{T_y} \sum_{j=1}^{j=T_y} \left( \hat{h}_j ; \hat{h}_j \right).
\]

Finally, the corresponding compressed vector representations at the source and target ends are stitched together and used to calculate the quality label, i.e., the HTER value, of the output translation of the system to be evaluated. The specific calculation method is as follows:

\[
\text{QE}_{\text{sentence}}(y, x) = \sigma \left( w^T [h_{\text{src}}; h_{\text{tgt}}] \right),
\]

where \(w\) denotes the compression vector of the connected splice denotes the weight vector with the predicted quality label; \(\sigma\) denotes the nonlinear mapping function S-shaped function that can limit the final predicted HTER score of the model to between \([0, 1]\).

5. Experiment and Analysis

5.1. Dataset and Evaluation Indicators. The dataset used to train the word prediction module is derived from the WMT 2017 shared task: machine translation of news for the English-Chinese language oriented translation task, and the parallel corpus of this language oriented includes Europarlv7>Common Crawl corpus, News Commentary v12, Rapid corpus of EU press releases, totaling 5.2 million parallel sentence pairs, and using newstest2016 as the validation set (development data). All datasets of the word prediction module were used using the Moses tool. Preprocessing operations such as tokenization and truecase are performed; byte-pair encoding (BPE) is used for word separation. After the BPE operation, the vocabulary size of the source language (English) is 7734, and the vocabulary size of the target language (Chinese) is 90818; all words that are not in the vocabulary (out-of-vocabulary, OOV) are mapped to the special identifier \(<\text{UNK}>\).

The dataset used for the training quality evaluation module is derived from the English-Chinese language oriented QE dataset of WMT 2017 shared task: quality estimation task I: sentence-level QE, which is \(T\)-domain related and includes training data, validation set, and test set (test data). Each dataset includes the source sentences, the machine translation results, the postediting results obtained by human postediting of machine translation, and the quality label HTER corresponding to the machine translation results, where the machine translation results are obtained by translating the source sentences using phrase-based SMT (PBSMT). The statistical information of the English-Chinese oriented sentence-level quality assessment task dataset is shown in Table 1, where the test set includes both the test set of 2017 and the test set of 2016. In order to
ensure the consistency of data processing for the whole QE model, all datasets of QE need to be uniformly tokenized, truecase preprocessed, and BPE word separation [22].

In the sentence-level machine translation quality assessment task, the quality label score prediction task (scoring) uses Pearson’s correlation coefficient, mean average error (MAE), and root mean squared error (RMSE) as evaluation metrics. Root mean squared error (RMSE) was used as evaluation indicators. Among them, Pearson’s correlation coefficient is the main evaluation index, which is denoted by the Greek letter $\rho$. The Pearson’s correlation coefficient between variables $X$ and $Y$ can be calculated by Equation (9)

$$
\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}, \tag{9}
$$

where $\sigma_X$ denotes the variance of variable $X$, $\sigma_Y$ denotes the variance of variable $Y$, and $\sigma_{XY}$ denotes the covariance of variable $X$ and variable $Y$. Because the variance of variables $X$ and $7$ is greater than $0$ ($\sigma_X > 0$ and $\sigma_Y > 0$), therefore the positivity of $\rho$ depends on the positivity or negativity of the covariance of $X$ and $T$. Variables $X$ and $Y$ are data point sets $(x_i, y_i)$, and therefore, the Pearson correlation coefficient can be refined.

Table 2 shows the detailed interpretation of Pearson’s correlation coefficient, and if the correlation coefficient is less than $0$, the two are negatively correlated, and if the correlation coefficient is greater than $0$, the two are positively correlated. By comparing the positive correlation between the predicted HTER value and the real HTER value, if the positive correlation is higher, it indicates that the machine translation quality assessment model is more effective.

The average absolute error is used to calculate the average of the absolute error between the predicted HTER and the real HTER, and the smaller the value, the better. It is calculated as shown in Equation (10).

$$
\text{MAE} = \frac{\sum_{i=1}^{n}|x_i - y_i|}{n}, \tag{10}
$$

$$
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}}. \tag{11}
$$

5.2. Experimental Results. The experimental analysis in this section is mainly based on the experimental results of the machine translation quality assessment model on the test set. In this section, we first build on the monolingual to machine translation quality assessment model, i.e., only the source-to-target direction word predictor (denoted as src-tgt word predictor) submodule and its quality assessment module are retained, and the target-to-source direction word predictor (denoted as tgt-src word predictor) submodule and its quality assessment module are removed. In other words, the model is degraded to an architecture similar to that of the predictor-estimator model and the bilingual expert model to investigate the impact of different quality evaluation feature vectors on the translation quality evaluation task of the tgt-src word predictor and its quality evaluation module and the impact of different training methods on the quality evaluation task of machine translation.

From the experimental results in Table 3, the best experimental results can be achieved by using the word prediction vector representation corresponding to each word at the decoding end, and the experimental results of the intermediate vector representation are the second best; this is because the quality assessment feature vector is the link between the feature extraction module and the quality assessment module, and the good or bad quality assessment feature vector will directly affect the final machine translation quality assessment effect. The word vector representation only contains the information of the previous word and the next word of the current predicted word; the intermediate vector representation can take into account all the contextual information related to the current predicted word and the source word information through the bidirectional recurrent neural network and attention mechanism at the decoding end; the word prediction vector representation also contains the information of the current predicted word on the basis of the intermediate vector representation. Therefore, by designing more informative quality assessment feature vectors, the effectiveness of machine translation quality assessment can be effectively improved.

Table 4 shows the experimental results of the recurrent neural network-based monolingual to machine translation quality assessment model using different quality assessment feature vectors on the machine translation quality assessment task dataset under the joint training condition. Fine-tuning indicates that the feature extraction module and the quality assessment module are trained jointly, i.e., the training data of the machine translation quality assessment task is used to update both the quality assessment module and parameters of the feature extraction module, i.e., the parameters of the pretrained source-to-target language word prediction module are fine-tuned. It should be noted that the dropout operation of the feature extraction module is turned off during the whole process of joint training of the model.

From the experimental results in Table 4, it can be seen that the experimental results on the word vector representation decrease by the joint training method, which
demonstrates the importance of a valid (or information-rich) quality assessment feature vector for the final machine translation quality assessment. Because the information contained in the word vector representation is not sufficient to reflect the “quality” of the output translation to be evaluated, coupled with the joint training method, the misguided information learned in the word vector representation “spreads” to the whole machine translation quality evaluation model, resulting in poor quality evaluation. The final quality assessment results are poor. On the contrary, the experimental results were improved for both the intermediate vector representation and the word prediction vector representation, and the best results were obtained for the word prediction vector representation. Therefore, based on the use of information-rich quality assessment feature vectors, the effect of machine translation quality assessment can be effectively improved by the joint learning training method.

From the experimental results in Table 5, the monolingual-oriented machine translation quality assessment model based on the self-attentive network, compared
with the monolingual-oriented machine translation quality assessment model based on the recurrent neural network, has improved the final quality assessment effect under the experimental conditions of using three different quality assessment feature vectors. Among them, under the same experimental conditions of using the hidden layer (middle) vector representation, the self-attentive network-based monolingual-oriented machine translation quality assessment model improves the Pearson correlation coefficient by +0.0689 (0.6749 vs. 0.606) on the test2016 dataset, and the Pearson correlation coefficient improves by +0.0756 (0.6805 vs. 0.6049); under the same experimental conditions using word prediction vector representation, the monolingual vectorizer translation quality assessment model based on self-attentive network improved the Pearson correlation coefficient by +0.0581 (0.704 vs. 0.6459) on the test2016 dataset and by +0.0562 on the test2017 dataset (0.698l vs. 0.6419).

This subsection first visualizes and compares the experimental methods explored in this paper in the form of bar graphs. Figure 4 shows the comparison results of the recurrent neural network-based machine translation quality assessment model on the test2017 test set in different experimental conditions in terms of Pearson’s correlation coefficient. Only two quality assessment feature vectors with relatively good results, namely, the intermediate vector representation at the decoding end (RNNsearch+latent) and the word prediction vector representation at the decoding end (RNNsearch+wordpredicting), are given in the figure for comparison of the experimental results. Figure 4 can visually reflect the experimental conclusions of this section. First, under the experimental condition of fusing only one-way translation knowledge from source to target, the training method using joint learning (fine-tuning) is better than the training method using nonjoint learning (no-fine-tuning).

Second, under the experimental conditions where joint training was used, the method of incorporating bidirectional translation knowledge (backtranslation+fine-tuning) was superior to the method of incorporating only unidirectional translation knowledge (fine-tuning). Third, under the same experimental conditions, using valid (or more informative) quality assessment feature vectors can effectively improve the quality assessment of machine translation, i.e., the information contained in the word prediction vector representation is better than the intermediate vector representation.

Figure 5 shows the comparison results of the Pearson correlation coefficient of the self-attentive network-based machine translation quality assessment model on the test2017 test set under different experimental conditions. Only two quality assessment feature vectors with relatively better results, i.e., the comparison of experimental results on the hidden layer vector representation at the decoding side (transformer+latent) and the word prediction vector representation at the decoding side (transformer+wordpredicting), are presented in the figure.

Similarly, Figure 5 can visually reflect the experimental findings in this section. First, under the experimental conditions of joint training, the method of fusing bidirectional translation knowledge (backtranslation+fine-tuning) outperforms the method of fusing only unidirectional translation knowledge (fine-tuning). Second, under the same experimental conditions, the use of effective (or more informative) quality assessment feature vectors can effectively improve the effectiveness of machine translation quality assessment, i.e., the information contained in the word prediction vector representation is better than the hidden layer vector representation. Combining with Figure 4, another experimental conclusion can be obtained that, under the same experimental conditions, using a word prediction module with better performance, i.e., the self-

6. Conclusions

In this paper, we propose a new machine translation model for bidirectional Chinese-English translation, including a recurrent neural network-based translation quality assessment model and a self-focused network-based model. The designed model has a better translation performance metric module in both monolingual-oriented machine translation quality assessment model and bilingual-oriented machine translation quality assessment model and obtains better results on different datasets. The experimental results in this paper show that the scheme designed in this paper has the best results for Chinese-English translation under comparison with baseline.

Data Availability

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

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

Acknowledgments

This study was supported by the Key Scientific Research Project of Zhanjiang University of Science and Technology: The translation and research on Fitzgerald’s works in China from the perspective of Bourdieu’s Sociology (CJKY201908).

References