Journal of Multimedia Information System VOL. 10, NO. 1, March 2023 (pp. 59-68): ISSN 2383-7632 (Online)
https://doi.org/10.33851/JMIS.2023.10.1.59

Online Automatic English Translation Based on Mobile Edge Computing

Nengsheng Qiu[1], Xiaoqin Qiu[2*]

Abstract

The English translation is an important means of converting language as the basis of communication, is gradually becoming an indispensable part of people's daily life. With the rapid progress of science, cultural exchanges between countries in the world have become more and more common. And the advent of the Internet era has led to more frequent exchanges between different languages, and the demand for language translation has gradually increased. How to translate English quickly and accurately is the current challenge of machine translation tasks. In this paper, we investigate how to use the neural network model for the English translation in order to better assist the establishment of the machine translation system. We design an English translation method based on an encoder-decoder structure and an attention mechanism. First, we analyze the characteristics of the LSTM model. Second, we design an English translation framework using the seq2seq model. Third, we combine the attention mechanism to build a more robust translation model to improve the translation performance of the model. Finally, we validate the translation performance of our model on two public datasets and experimental results prove that the method proposed in this paper has good evaluation performance.

Key Words: English Translation, LSTM, Mobile Edge Computing, Attention Mechanism.

I. INTRODUCTION

Language and writing are important means of daily communication. Language is the ladder to promote the beneficial development of different countries or peoples. With the development of the Internet and other technologies, the connection between countries in the world has been strengthened, and the content that needs to be translated is increasing day by day. Simply relying on manual translation can no longer meet the needs of today's social development. As English is the universal language in the world [1], many experts and scholars have begun to explore the online assistance system that can effectively replace and assist manual English translation [2-3]. Before the advent of computers, humans had the idea of using machine translation instead of human translation. With the birth of computers and the continuous development of computer technology, the development and application of online assistance systems for English translation have been promoted [4-5]. Today, with the continuous integration of domestic and foreign cultures, more and more industries and scenarios require language translation. According to the research results of language usage rate, in daily life, there are often situations where online translation of English is required. Therefore, in recent years, the electronic dictionary translation industry has developed more and more rapidly. With the popularization of languages and the increasing demand for language translation, the continuous optimization of translation software has become a rigid demand in the translation market. Before the software is optimized, the online translation software is usually scored in the form of scoring, the translation score of the software is obtained, and optimization is carried out according to the weak part of the software score.

Machine translation is an indispensable part of natural language processing [6-7], and there are two core tasks to be solved, namely natural language understanding and generation. Such machine translation is complete only if the machine understands the natural language first and successfully generates another language. With the rapid progress of society and science, cultural exchanges between countries in the world have become more and more common. And the advent of the Internet era has led to more frequent exchanges between different languages, and the demand for language translation has gradually increased. Major compa-

Manuscript received March 03, 2023; Revised March 10, 2023; Accepted March 14, 2023. (ID No. JMIS-23M-03-008)
Corresponding Author (*): Xiaoqin Qiu, +86-18250877459, qiuxq_xmc@aliyun.com
[1]School of Foreign Languages, Xiamen, China. qiuns_xit@aliyun.com
[2]Department of General Education, Xiamen Medical College, Xiamen, China. qiuxq_xmc@aliyun.com

nies at home and abroad are constantly developing and improving their own machine translation systems, such as Youdao [8], Baidu [9], and Google [10]. Due to the high cost, low efficiency and some other reasons compared to human translation, machine translation has become an indispensable part of major industries due to its high efficiency and low cost. At present, machine translation generally consists of two directions: Statistical Machine Translation (SMT) and Neural Machine Translation (NMT) [11-13].

Neural machine translation has been continuously improved with the rise of deep learning. Compared with the statistical machine translation model, one type of neural machine translation has a simpler structure but still retains the general framework of statistical machine translation. Correspondingly, the framework of deep learning is used to replace some middle modules. Another type of model has almost abandoned the basic structure of the previous statistical machine translation, and replaced it with a variant framework of the end-to-end (End to End) algorithm structure-Sequence to Sequence (Seq2Seq) [14-16]. The Seq2Seq model is used for neural machine translation, mainly by using the Encoder-Decoder mechanism, where the sequence is generally a sentence (an article fragment or an independent sentence), and the Seq2Seq framework is used to generate the target language or sentence. Compared with various translation strategies in the past, neural machine translation has more obvious advantages and generates target sentences with higher quality. In the early stage, machine translation uses the Seq2Seq network model composed of two RNN models [17], which can parse the original language into a fixed vector. At the same time, the intermediate vector can be decoded to generate the specified translation language. Many scholars have studied the mechanism of attention based on this. After calculating the attention matrix between relative words, the overall quality of the translation results was improved a lot. And the advantage is that it can visually align the words in the sentence through the calculation of attention, and has a significant advantage in the processing of long sentences.

In this paper, we investigate how to utilize the deep learning model to translate the English language in order to better assist the establishment of the machine translation system. In the previous translation process, the original scoring method was relatively general for the acquisition of English online translation scoring indicators, and did not divide the translation scoring indicators into reasonable levels, which often caused errors in the translation results. In view of the problems in the use of the original method, we use the attention mechanism to improve the original translation method, and design an automatic translation method for English online translation based on the deep learning model.

The rest of this paper is organized as follow. Section 2 describes the related work for English translation. Section 3 introduces the proposed method for English translation mechanism, including the basic LSTM model [18], the translation model based on seq2seq, and the translation with attention mechanism. Section 4 presents experimental results to compare and demonstrate the performance of proposed model for English translation. Section 5 concludes the whole paper.

II. RELATED WORK

In this chapter we briefly describe related work in English translation studies. Current neural network-based translation models mainly include recurrent neural networks, convolutional neural networks, and encoder-decoder structures.

Before using neural network as the model of machine translation, neural network was used to assist statistical machine translation, but because statistical machine translation has many problems, such as the problem of data sparseness, it is difficult to solve the problem of large model size, which leads to difficulty in training. Kalchbrenner et al. [19] proposed a class of continuous sentence-level neural network-based translation models. The representation in this method for language does not depend on alignment or phrase translation units. The model uses a convolutional neural network [20] (CNN) to encode the provided source language into a continuous vector, and then uses a recurrent neural network to decode the vector into the target language. It solves the long-distance reordering problem in statistical machine translation, and lays the foundation for the subsequent pure neural network for machine translation. The Back-Propagation algorithm [21] is the basis of recurrent network training. It back-propagates the error signal through the corresponding time series, and updates the weights of the parameters in the network by calculating the gradient at each moment. However, in the actual training process, there will still be gradient disappearance and gradient explosion. On this basis, Sutskever et al. invented the neural translation model of the Seq2Seq architecture based on the recurrent network of the LSTM structure. Because the long-short-term memory structure adopts a special gate mechanism, the problem of "gradient disappearance and explosion" in the entire network has been solved to a certain extent, and the long-distance dependence in the sentence is better solved. question. In the same year, Cho et al.[15] proposed a more compact network structure consisting of a gated recurrent unit (GRU) consisting of one state unit and two gated units [22]. And in many neural machine translation projects, its convergence speed and performance have achieved better performance.

Recurrent neural networks imitate the human translation process, and to a certain extent, have a good ability to model

time series. However, neither the original RNN nor the later LSTM, and GRU are out of the constraints of the overall network timing, and parallel training cannot be performed, resulting in very low training efficiency. Moreover, effective global information cannot be learned in long sentences, and the difficulty of long-distance dependence is not comprehensively solved.

End-to-end learning makes machine learning no longer need to go through tedious data preprocessing, feature selection, dimensionality reduction and other processes like traditional feature engineering methods, but directly uses artificial neural networks to automatically extract and combine more complex features from simple features, which greatly improves model capability and engineering efficiency. In traditional methods, image classification requires many stages of processing. First, some hand-designed image features need to be extracted, and after dimensionality reduction, classification algorithms such as SVM need to be used to classify them. Compared with this multi-stage pipeline-like processing flow, end-to-end deep learning trains only one neural network, the input is the pixel representation of the picture, and the output is directly the classification category. Traditional machine learning requires a large number of manually defined features, and the construction of these features often leads to implicit assumptions about the problem. In the feature engineering method of traditional machine learning, the feature extraction process often relies on a large number of prior assumptions, which greatly increases the development cycle of related systems. The final system performance is very dependent on the selection of features. Data and features determine the upper limit of machine learning, but human intelligence and cognition are limited. Therefore, the accuracy and coverage of manually designed features are limited. For different tasks, traditional machine learning feature engineering methods need to select different features. Features that perform well on one task may not perform well on other tasks. End-to-end learning liberates people from a large number of feature extraction tasks, and does not require much prior knowledge of people. In a sense, the feature extraction of the problem is all automatic, which means that the development of the relevant system can be completed even if we are not experts on the task. Furthermore, the end-to-end learning actually implies a new form of representation for the problem, namely the distributed representation. Under this framework, the input of the model can be described as a distributed real number vector, so that the model can have more dimensions to describe a thing, and at the same time avoid the discretization of the traditional symbol system for the characterization of objective things. For example, in natural language processing tasks, representation learning redefines the distinction between words and sentences.

In Refs. [23-24], the author proposes a translation method, which uses the CNN model to build an encoder. And the authors prove that their proposed method can better capture the dependencies in sentences through a large number of experiments. Subsequently, Refs. [24-25] improved on the model proposed by the former. They made full use of the attention mechanism and the convolutional gating model to design an English translation model, which performed well on multiple data sets. The ablation experiment also verified that the proposed translation model can better find the structural relationships in sentences. Ref. [26] combines the advantages of attention mechanism and neural network to propose a lightweight translation model, which has better parallelism. The experimental results also show that the model is lighter and more efficient.

III. METHOD

As mentioned above, we investigate how to utilize the deep learning model for the English translation in order to better assist the establishment of the machine translation system [27]. We design an English translation method based on an encoder-decoder structure and an attention mechanism. First, we introduce the characteristics of the LSTM model. Then, we design an English translation framework using the seq2seq model. Finally, we combine the attention mechanism to build a more robust translation model to improve the translation performance of the model.

3.1. The Basic LSTM Model

The structure of recurrent neural network is designed as a structure specifically for processing languages [28]. Subsequently, recurrent neural network models have been greatly developed in the study of image and language processing. For example, let's assume that the language sequence is $\{x_1, x_2, x_3, \dots, x_n\}$. Then the neural network is trained in the following ways. First, the recurrent neural network will process each data x_i, and then calculate and get a hidden state data h_j. The hidden state is often referred to as the memory of the RNN, which is used to store the information of the previous i-1 data. The update method of RNN's hidden state is the key of RNN. For the simplest form of RNN, the state update equation is as follows:

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

where $f(x)$ is an activation function. The value of the hidden state is associated with the previous hidden state and x_i. In other words, each state is the accumulation of the previous state.

$$h_t = \sigma(W_{hh}h_{t-1} + W_{xh}x_t), \tag{2}$$

where W_{hh} and W_{xh} are the parameters of the network. RNNs have an output in addition to a hidden state. In a simple form of RNN as shown above. At the same time, we can also use the hidden vector generated by the calculation to perform the classification task. In other words, the hidden vector classification is calculated as follows.

$$P_t = \text{softmax}(W_{hy} y_t). \tag{3}$$

In the above equation, P_t is a probability distribution over all possible categories. In language models or other models used to predict the next word, the dimension of P_t is the size of the vocabulary V, d is the dimension of the output of the RNN, usually between 100 and 1000. However, V is usually very large, so in the whole model, the place with the largest amount of calculation is on W_{hy}. For the optimization of the W_{hy} layer, researchers have also proposed many effective methods, such as multi-level softmax or negative sampling [29].

The problem of gradient vanishing and gradient explosion can be alleviated by the entrance element to some extent [30]. And the LSTM and GRU are the most widely used gating units. It is noted that the cause of the vanishing gradient comes from the update of the hidden state in the RNN. The value of the hidden state grows cumulatively repeatedly, but the gradient problem occurs when the network updates the parameters. To solve this problem, the following items are added to the RNN model:

$$h_t = h_{t-1} + \sigma(W_{hh} h_{t-1} + W_{xh} x_t). \tag{4}$$

However, the structure shown in the formula still has certain downside. That is, the hidden state will be spread directly, which is not in line with the characteristics of language. For example, when encountering punctuation marks, stop words or new subjects, it may be necessary to partially reset the previous state. Based on this concern, LSTM adds three gates, including the forget gate, the output gate, and the input gate. There is an important difference between LSTM and traditional RNN. The nerve cell state and cell output of traditional RNN are the same, which are represented by h_t [31]. But in LSTM, the state of nerve cells and cell output are not the same. The cell state and cell output of the LSTM are calculated as follows.

$$c_t = c_{t-1} + \sigma(W_{hh} h_{t-1} + W_{xh} x_t). \tag{5}$$

$$h_t = c_t, \tag{6}$$

where c_t represents the cell state of the neural unit, and h_t represents the output of the neural unit. On the basis of the above two formulas, LSTM adds forget gate, input gate and output gate, which are defined as follows.

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma(W_{hh} h_{t-1} + W_{xh} x_t). \tag{7}$$

$$h_t = o_t \circ \sigma(c_t), \tag{8}$$

where $\circ$ represents the element-wise multiplication of the elements of the vector (or matrix), f_t, i_t and o_t represent the forget gate, the input gate, and the output gate correspondingly, and $\sigma(x)$ is a nonlinear function. In LSTM, f_t, i_t, and o_t are adjustable and are decided by the current signal x_t and h_{t-1} of the previous state. And we want its value to be between [0, 1]. 0 means completely discarded, 1 means completely retained. Therefore, LSTM uses the sigmoid function to calculate f_t, i_t and o_t. Using tanh as the nonlinear function $\sigma(x)$, the calculation formula of LSTM is obtained as follows.

$$f_t\& = \text{sigm}(W_{hf} h_{t-1} + W_{xf} x_t). \tag{9}$$

$$i_t\& = \text{sigm}(W_{hi} h_{t-1} + W_{xi} x_t). \tag{10}$$

$$o_t\& = \text{sigm}(W_{ho} h_{t-1} + W_{xo} x_t). \tag{11}$$

$$\widehat{h_t} = \tanh(W_{hh} h_{t-1} + W_{xh} x_t). \tag{12}$$

$$c_t\& = f_t \circ c_{t-1} + i_t \circ \widehat{h_t}. \tag{13}$$

$$h_t\& = o_t \circ \tanh(c_t). \tag{14}$$

3.2. The Translation Model Based on Seq2seq

Sequence to sequence (seq2seq) model is a special recurrent neural network architecture, which usually solves the complex language problems, such as machine translation, question answering, creating chat robots, text summarization. And significant breakthroughs have been made in all of these areas. The main architecture of the seq2seq model is an encoder-decoder model, as shown in Fig. 1. During translation, the encoder encodes the source language sequence of indefinite length into a fixed-length representation vector, and then the decoder generates the target language sequence of indefinite length according to the input representation vector.

The encoder and decoder in the Seq2Seq model are both LSTM models or GRU models. Where the encoder reads the input sequence and computes the collected information on a hidden state vector or context vector. Seq2seq model starts translation after the input sequence encoding is completed and the full meaning of the source language is obtained. The translation mechanism of Seq2Seq model is more consistent with human habits, that is, the model will focus on the true meaning of language. Therefore, the neural translation model can capture the longer dependencies and solve the different word order characteristics of differ-

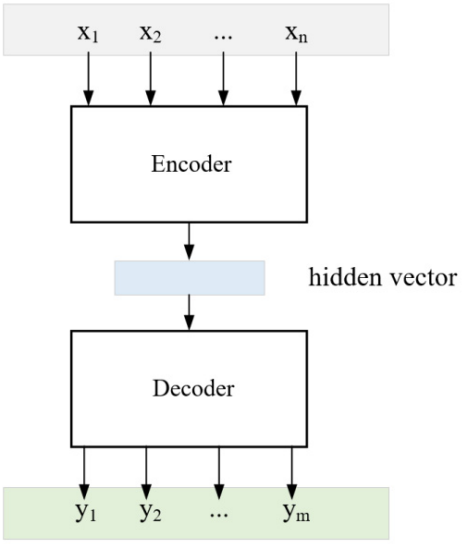

Fig. 1. The encoder-decoder framework.

ent languages to produce smoother translation. Most of the encoders and decoders in the seq2seq model use recurrent neural networks. The difference lies in the following points: (1) the type of recurrent neural network, including ordinary RNN, LSTM, and GRU; (2) the depth of the recurrent neural network, single-layer or multi-layer; (3) the direction of the recurrent neural network, unidirectional or bidirectional.

3.3. Translation with Attention Mechanism

Although neural machine translation based on seq2seq model has reached the highest level in large-scale translation at that time, the translation of long sentences is still a challenge in neural machine translation. The attention mechanism in deep learning is a method that draws on the selective visual attention mechanism of humans [32-34], and its goal is to select the information that is more important to the current task goal from the information to be calculated. In the field of machine translation, attention mechanisms are often attached to encoder-decoder models. Intuitively, this paper uses the attention mechanism to process English sentences gracefully. First of all, place names are very rare in the distribution of daily texts, and are often not included in the translation lexicon. We use the unknown symbol "< U >" to express. Thus, the attention mechanism can perform the substitution of "unknown" words. The idea of this method is very simple. After obtaining the preliminary translation results and the corresponding attention weight matrix, the two are analyzed to find the characteristics of the attention weight vector of place names and proper names and their corresponding translated words, and replace the corresponding translated words with transliterated words. When looking at a picture, humans will first quickly scan the whole picture to get the key areas that need attention, and then focus their attention on these key areas to reduce the interference of other information. This mechanism of human brain greatly improves the efficiency and accuracy of human processing visual signals.

For encoder, it is mainly divided into embedded layer and hidden layer. The embedding layer mainly converts input words into real-value vectors through mapping tables. The hidden layer is presented as the bidirectional cyclic neural network hidden layer, that is, a positive coding cyclic neural network layer and a reverse coding cyclic neural network layer, whose output at a certain moment is the series of the forward and reverse output vectors. The decoder is mainly divided into embedded layer, hidden layer and softmax layer. The hidden layer is different from the encoder in that it is a one-way cyclic neural network layer. The softmax layer accepts the output from the hidden layer, which is used to calculate the probability and find the word with the highest probability in the word list for output. By providing additional information to the encoder and decoder, the attention mechanism solves the problem of low accuracy of the encoder-decoder model in related tasks due to the information loss and noise caused by the encoding of different information into a certain length vector. The seq2seq model can align each word and calculate the context vector during the decoding process. The main steps are described below:

(1) Calculate the attention force weight at each moment by combining the state of encoder and decoder;
(2) The context vector is obtained by weighting the average state values at each moment by means of the attention weight;
(3) Calculate the attention vector by combining the context vector values and the decoding stage values;
(4) Calculate the results through the softmax function; The process can be calculated as

$$\alpha_{ts} = \frac{exp\left(score\left(h_t, \bar{h}_s\right)\right)}{\sum_{s'} exp\left(score\left(h_t, \bar{h}_{s'}\right)\right)}. \tag{15}$$

$$c_t = \sum_s \alpha_{ts} \bar{h}_s. \tag{16}$$

$$a_t = \tanh(W_c[c_t; h_t]). \tag{17}$$

$$p(y_t \mid y_{<t}, x) = \text{softmax}(W_s a_t), \tag{18}$$

where a_{ts} is the attention weight corresponding to the encoding stage at time s in the decoding stage at time t; a_t, c_t, h_t indicates the attention vector, the background vector, the hidden state of t moment. The overall process is shown in Fig. 2.

IV. EXPERIMENTS AND RESULTS

In this section, we experimentally verify the translation

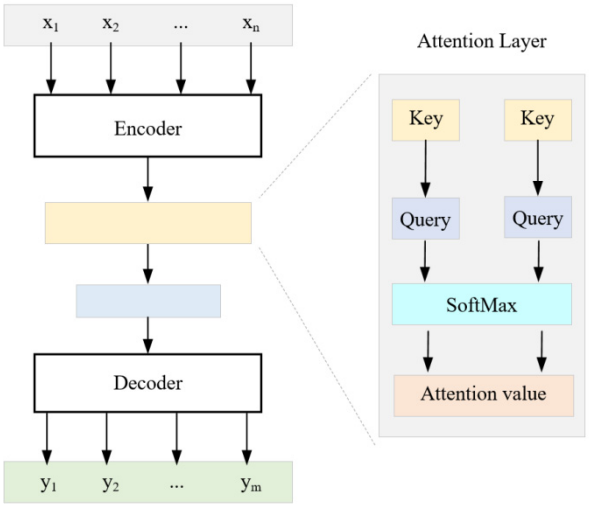

Fig. 2. The translation model with attention mechanism.

model based on the seq2seq model and attention mechanism proposed in this paper. This section includes two parts, data processing and performance evaluation. First, we introduce the data selection and define the evaluation indicators. Second, we compare and analyze the model translation results based on public datasets through experiments, and verify the performance of our method.

4.1. Data and Metrics

We selected TED2013 database [35] and UM-Corpus database [36] as experimental data.

The TED2013 database can be downloaded from OPUS. The database contains 155K sentence pairs in both Chinese and English. The English sentence contains 2.99M English words, and the Chinese sentence contains 2.90M words or 5.42M Chinese characters. On average, each English sentence contains 19.3 English words and each Chinese sentence contains 35.0 Chinese characters. In the corpus, there are 31,280 English words with frequency greater than 2 (case sensitive), and 33,607 Chinese words with frequency greater than 2, or 5647 Chinese characters.

The UM-Corpus database contains article sentence patterns in eight fields, such as education, law, and journalism. The database contains 2.22M sentences in both Chinese and English. The English sentence contains 28.1M English words. The Chinese sentence contains 25.1M words or 46.9M Chinese characters. On average, each English sentence contains 12.7 English words and each Chinese sentence contains 21.2 Chinese characters. In the corpus, there are 98,903 English words with frequency greater than 5, and 83,416 Chinese words with frequency greater than 5, or 15,845 Chinese characters.

In our experiment, we choose the general *BLEU* value as an indicator for evaluating the translation quality of the translation model [37]. The larger the value, the higher the translation quality. The calculation method of the *BLEU* value is shown below.

$$BLEU = BP \times \exp(\sum_{n=1}^{N} w_n \log_{10} p_n), \tag{19}$$

$$BP = \begin{cases} 1 & lc > lr \\ \exp(1 - lr/lc) & lc \le lr' \end{cases} \tag{20}$$

where BP is the penalty factor, N indicates the longest tuple length, usually N is 4, n is the number of tuples; w_n is the weight of tuples n, p_n is the proportion of tuples n, lc indicates the length of the machine translation, and lr indicates the length of the shortest reference translation sentence.

4.2. Performance Evaluation

In order to compare the performance of the models proposed in this paper, we compare the translation performance of different translation models on two datasets, including RNN, LSTM, Transformer, and the proposed Seq2seq with attention model (S1). Table 1 compares the translation performance of the four models on the two datasets, namely the BLEU value.

As can be seen from the table, our proposed model achieves the highest score, which also shows that the model has a good translation effect in the English translation task. At the same time, we also found that the translation effect of the LSTM model is better than that of the RNN model, which also reflects that the forgetting gate added in the LSTM model plays a role in retaining the information features of long sentences.

In this paper, one of the innovations of our method is the addition of an attention mechanism to the seq2seq model. Therefore, we compare and verify the role of the attention mechanism in the translation model. Fig. 3 shows the translation performance of models with and without attention on two datasets. We abbreviate Seq2seq with attention as S1 and Seq2seq without attention as S2. As can be seen from Fig. 3, on both datasets, the translation performance of the model with attention mechanism is better than that of the translation model without attention mechanism, which fully shows that the attention mechanism plays a positive role in model translation. This also demonstrates the effectiveness of our proposed method in English translation.

Table 1. Translation performance comparison with other models.

Method	TED2013	UM-Corpus
RNN	15.7	16.8
LSTM	18.2	19.9
Transformer	20.4	21.7
S1 (Ours)	**21.6**	**23.1**

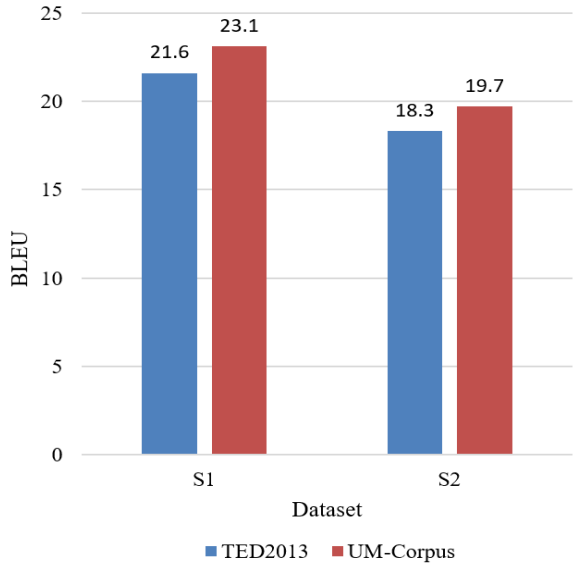

Fig. 3. The performance comparison of translation models with and without attention mechanism on two datasets.

V. CONCLUSION

In this paper, we present an online English translation mechanism model based on the Seq2seq model to provide fast and accurate English machine translation. We analyze the characteristics of current translation models and design an English translation method based on an encoder-decoder structure and an attention mechanism. First, we analyze the characteristics of the LSTM model. Second, we design an English translation framework using the seq2seq model. Third, we combine the attention mechanism to build a more robust translation model to improve the translation performance of the model. The attention mechanism enables the model to better notice and extract features in long sentences. Finally, we validate the translation performance of our model on two public datasets.

ACKNOWLEDGMENT

This work was supported by Fujian Province Education Science "13th Five-Year Plan" 2020 Project (Granted No. FJJKCG20-202), 2020 Annual University-Level Scientific Research and Innovation Team Project at Xiamen Institute of Technology (Granted No. KYTD202008), and Xiamen Education Science The 13th Five-Year Plan Project: A Study on the Ideological and Political Development Path of College English Curriculum from the Perspective of Blended Teaching (Granted No.: 20031).

REFERENCES

[1] D. G. Drubin and D. R. Kellogg, "English as the universal language of science: Opportunities and challenges," *Molecular Biology of the Cell*, vol. 23, no. 8, pp. 1399-1399, 2012.

[2] B. Songbin and M. Fanqi, "The design of massive open online course platform for English translation learning based on Moodle," in *2015 Fifth International Conference on Communication Systems and Network Technologies*, *IEEE*, 2015, pp. 1365-1368.

[3] C. K. Chang and C. K. Hsu, "A mobile-assisted synchronously collaborative translation–annotation system for English as a foreign language (EFL) reading comprehension," *Computer Assisted Language Learning*, vol. 24, no. 2, pp. 155-180, 2011.

[4] T. Yang and H. Fan, "Application of computer technology in english translation," *Journal of Physics: Conference Series*, *IOP Publishing*, vol. 1575, no. 1, p. 012029, 2020.

[5] Z. Zhao, "Research on English translation skills and problems by using computer technology," *Journal of Physics: Conference Series, IOP Publishing*, vol. 1744, no. 4, pp. 042111, 2021.

[6] A. Lopez, "Statistical machine translation," *ACM Computing Surveys (CSUR)*, vol. 40, no. 3, pp. 1-49, 2008.

[7] H. Somers, *Machine Translation: History*, *Development*, *and Limitations*, 2011.

[8] K. Fu, J. Huang, and Y. Duan, "Youdao's winning solution to the nlpcc-2018 task 2 challenge: A neural machine translation approach to chinese grammatical error correction," in *CCF International Conference on Natural Language Processing and Chinese Computing*, Cham, 2018, pp. 341-350.

[9] M. Sun, B. Jiang, and H. Xiong, et al., "Baidu neural machine translation systems for WMT19," in *Proceedings of the Fourth Conference on Machine Translation*, 2019, pp. 374-381.

[10] Y. Wu, M. Schuster, and Z. Chen, et al., "Google's neural machine translation system: Bridging the gap between human and machine translation," arXiv preprint arXiv:1609.08144, 2016.

[11] Y. Al-Onaizan, J. Curin, and M. Jahr, et al., "Statistical machine translation," Final Report, JHU Summer Workshop, 1999.

[12] P. Koehn, "Neural machine translation," arXiv preprint arXiv:1709.07809, 2017.

[13] F. Stahlberg, "Neural machine translation: A review," *Journal of Artificial Intelligence Research*, vol. 69, pp. 343-418, 2020.

[14] T. Glasmachers, "Limits of end-to-end learning," in *Asian Conference on Machine Learning*, *PMLR*, 2017, pp. 17-32.

[15] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[16] J. Gehring, M. Auli, D. Auli, et al., "Convolutional sequence to sequence learning," in *International Conference on Machine Learning*, *PMLR*, 2017, pp. 1243-1252.

[17] T. Mikolov, M. Karafiát, and L. Burget, et al., "Recurrent neural network based language model," *Interspeech*, vol. 2, no. 3, pp. 1045-1048, 2010.

[18] S. Hochreiter, J. Schmidhuber, "Long short-term memory," *Neural Computation*, 1997, vol. 9, no. 8, pp. 1735-1780.

[19] N. Kalchbrenner, P. Blunsom, "Recurrent convolutional neural networks for discourse compositionality," arXiv preprint arXiv:1306.3584, 2013.

[20] J. Gehring, M. Auli, and D. Grangier, et al., "A convolutional encoder model for neural machine translation," arXiv preprint arXiv:1611.02344, 2016.

[21] S. M. Bohte, J. N. Kok, and H. La Poutre, "Error-backpropagation in temporally encoded networks of spiking neurons," *Neurocomputing*, vol. 48, no. 1-4, pp. 17-37, 2002.

[22] J. Chung, C. Gulcehre, and K. H. Cho, et al., "Empirical evaluation of gated recurrent neural networks on sequence modeling," arXiv preprint arXiv:1412.3555, 2014.

[23] M. T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," arXiv preprint arXiv:1508.04025, 2015.

[24] J. Gehring, M. Auli, and D. Grangier, et al., "Convolutional sequence to sequence learning," in *International Conference on Machine Learning*, *PMLR*, pp. 1243-1252, 2017.

[25] A. Graves, N. Jaitly, and A. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM," in *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, *IEEE*, 2013, pp. 273-278.

[26] A. Vaswani, N. Shazeer, and N. Parmar, et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, p. 30, 2017.

[27] K. Han, A. Xiao, and E. Wu, et al., "Transformer in transformer," *Advances in Neural Information Processing Systems*, vol. 34, pp. 15908-15919, 2021.

[28] R. D. Beer, "Dynamical approaches to cognitive science," *Trends in Cognitive Sciences*, vol. 4, no. 3, pp. 91-99, 2000.

[29] T. Mikolov, I. Sutskever, and K. Chen, et al., "Distributed representations of words and phrases and their compositionality," *Advances in Neural Information Processing Systems*, p. 26, 2013.

[30] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157-166, 1994.

[31] K. Cho, B. Van Merriënboer, and C. Gulcehre, et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014.

[32] V. Mnih, N. Heess, and A. Graves, "Recurrent models of visual attention," *Advances in Neural Information Processing systems*, vol. 27, 2014.

[33] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014.

[34] M. T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," arXiv preprint arXiv:1508.04025, 2015.

[35] J. Tiedemann, "Parallel data, tools and interfaces in OPUS," *Lrec*, vol. 2012, pp. 2214-2218, 2012.

[36] L. Tian, D. F. Wong, and L. S. Chao, et al., "Um-corpus: A large english-chinese parallel corpus for statistical machine translation," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 2014, pp. 1837-1842.

[37] K. Papineni, S. Roukos, and T. Ward, et al., "Bleu: A method for automatic evaluation of machine translation," in *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 2002, pp. 311-318.

Journal of Multimedia Information System VOL. 10, NO. 1, March 2023 (pp. 59-68): ISSN 2383-7632 (Online)
https://doi.org/10.33851/JMIS.2023.10.1.59

AUTHORS

Nengsheng Qiu received his Bachelor Degree at Fujian Normal University in 2002 and received his Master Degree at Suzhou University in 2010. He is currently a professorsenior lecture at School of Foreign Languages, Xiamen Institute of Technology. His research interests include English translation and computer technology.

Xiaoqin Qiu received her Bachelor Degree at Jiangsu University of Technology in 2007 and received her Master Degree at Xiamen University in 2011. She is currently a senior lecturer at Department of General Education, Xiamen Medical College. Her research interests include English translation and computer technology.