

Research Article

Multisensor Feature Fusion-Based Model for Business English Translation

Pingfei Zheng ^{1,2}

¹Universiti Utara Malaysia, Gubangbasu County, Kedah 06010, Malaysia

²Guangzhou Institute of Science and Technology, Guangzhou, Guangdong 510540, China

Correspondence should be addressed to Pingfei Zheng; gjyy@gzist.edu.cn

Received 13 December 2021; Revised 6 January 2022; Accepted 8 January 2022; Published 31 January 2022

Academic Editor: Baiyuan Ding

Copyright © 2022 Pingfei Zheng. 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.

With the wide use of computers, machine translation has been gradually applied in many fields from natural language processing, such as industry, education, and so on. Due to the increasing demand for multilanguage translation, it is an urgent problem to effectively improve the quality of text translation. Driven by the upsurge of artificial intelligence, neural network technology is increasingly integrated into the field of machine translation, which gradually expands the traditional machine translation method into neural machine translation method. With the continuous improvement of deep learning technology, machine translation has gradually integrated these methods and strategies and achieved good results in multiple tasks, but there are still some shortcomings. The most prominent problem is that, since word vector is the basis for the model to obtain semantic and grammatical information, the existing methods cannot obtain semantic and grammatical feature information, which greatly reduces the accuracy of English translation. Based on this, this paper proposed a method of splicing word vector with character-level and word-level encoding vector. The characterization of fusion of more word vector can effectively solve the word does not appear in the table, the word with some low frequency, can express meaning more complete information, performance directly affects the whole translation model, the results can be seen through the experiment, we put forward the characteristics of the fusion method and strategy, can effectively enhance the overall translation performance of the model.

1. Introduction

With the widespread use of computers, the speed of manual text translation is no longer enough to meet daily needs [1–3]. Therefore, how to use computers to realize the mutual conversion between multiple languages is an urgent problem to be solved. Machine translation technology has gradually attracted the attention of academic circles and developed rapidly. Driven by the upsurge of artificial intelligence, improving the quality of machine translation by using artificial intelligence technology is a hot research direction in the field of natural language processing. As the demand for multilingual text translation grows, many AI companies are developing translation software and providing online services. In the history of machine translation, British and French governments actively promote mutual translation between English and French, our country also supports the

study of the machine translation between Chinese and English from the perspective of the theory of value, and machine translation is a blend of mathematical linguistics automation technology such as computer science discipline, studies of machine translation at the same time [4, 5].

Since the 1890s, data-driven statistical machine translation has replaced rule-driven method and gradually become the mainstream machine translation method. Compared with rule-driven methods, statistical machine translation methods do not require any rules, extract language knowledge from large-scale parallel corpora, construct efficient translation models by establishing statistical and probabilistic models for the whole translation process, and adjust model parameters during training. Statistical machine translation has become the core technology of online machine translation systems at home and abroad because of its advantages such as low labor cost, short development time,

and good robustness, which overcomes the bottleneck of rule-making based translation methods. In the process of statistical machine translation, various translation knowledge are needed by human experts set of language features to say, but different language leads to structure is different also, when translating, the conversion structure will become very important, only rely on artificial design features, will lead to the structure of language cannot get comprehensive coverage. At the same time, the traditional statistical machine translation technology still faces severe challenges due to the serious problems of data sparsity, strong dependence on corpus, and too much time cost [6].

So far, machine translation technology has been developed for decades, although various algorithm models have been introduced, the accuracy of machine translation is still low, and it cannot replace professional translators. Among them, the most prominent problem is the poor translation effect of long clauses with many words and complex sentence structure. In English, the structural components of long sentences are more complex. In addition to the main sentence structure, there are various modifiers and conjunctions. In addition, long sentences may contain more than one clause, and the relationships between clauses may be nested and parallel, so syntactic analysis is a necessary prerequisite for long sentence translation. Therefore, syntactic pre-processing of long and difficult sentences is an effective way to improve the translation quality of long sentences. At present, syntactic analysis translation methods can be roughly divided into two categories: language template-based translation methods and statistics-based translation methods. The translation method based on language template, which refers to the surface features of a sentence, is the earliest technical method of grammar translation. The advantage of language template-based translation method is that the translation is most accurate when the sentence template features are strong, while the disadvantage is that the translation is inaccurate or even impossible when the sentence template features are weak [7–9].

2. Related Works

In order to improve the translation methods based on language templates, scholars around the country have carried out in-depth studies and gradually evolved into a translation method based on statistics. This method can use machine learning method to conduct a lot of data mining and feature learning for weak features of sentences, such as the features of conjunctions and sentence patterns of long sentences and the features of punctuation usage, which can make up for the defects of incomplete matching of sentence features in template matching translation. However, the method based on statistics has its own limitations, if the sentence itself has a small number of punctuation marks or connectives, the accuracy of statistical translation methods will also decline [10–13].

Despite the promise of statistical machine translation, it is still difficult to translate statements in one language completely and correctly into statements in another language. Since 2013, deep learning composed of neural networks has become once powerful. The advantages of deep learning can well solve the disadvantages of statistical machine translation. Therefore, introducing deep learning methods into the field of machine translation has become a hot research direction. La et al. [14] put forward both short-term and long-term memory network, and its application in the frame of the end-to-end nerve machine translation, both short-term and long-term memory in circular neural network on the basis of introducing the door mechanism can control self-circulation, in order to achieve sustainable flow gradient can keep for a long period of time, avoids the disappear because of the gradient and the gradient semantic loss phenomenon caused by the explosion. Shuang et al. [15], for the first time, treated machine translation as an end-to-end learning task and effectively solved the problem of variable length of input and output by using recurrent neural network. Huang et al. [16] proposed a gated recurrent unit (GRU) to replace LSTM in handling machine translation tasks, GRU is actually an optimization of LSTM, which simplifies the internal structure, reduces the training parameters, and improves the training efficiency. Choi et al. [17] invented a new and simple network framework. Based on this, Ahmed et al. [18] proposed the attention mechanism, which effectively solved this problem and brought machine translation to a new level. The attentional mechanism is essentially a small neural network trained simultaneously with the RNN-RNN network. Awad et al. [19] proposed a two-way simultaneous decoding method, in which the decoding direction of each word is dynamically determined by the model. Domestic scholars started relatively late in the field of machine translation, but they have also made a lot of achievements. The minimum risk training criterion was proposed by [20] to deal with the mismatch between training and testing in the codec framework of attention mechanism, and the performance of machine translation was significantly improved. Chen et al. [21] proposed an unsupervised adaptive method to solve the problem of domain migration, which fine-tuned the pretrained out-of-domain neural machine translation model by using pseudo-in-domain corpora. Specifically, the model first performed lexical induction to extract the dictionaries in the domain, and then adopted the pseudo-parallel intradomain corpus is constructed by reverse translation of the target sentences word by word.

From the above analysis, we know that the above methods have studied business English translation problem to some extent, some problem still exists [22–24]. For example, no scholar has applied multisensor feature fusion to this field till now, so the research here is still a blank, which has great theoretical research and practical application value for business English translation. In addition, almost all translation models are based on the encoder decoder framework. Although its structure achieves good results, all translation ends from left to right.

The contributions of this paper is: 1. multirepresentation fusion is proposed for the word vector of translation, encoding the input data with various granularity. 2. Meanwhile, a new neural machine translation model is proposed combining with the idea of refining neural networks and transformer network. Next, I will introduce this chapter according to the sequence of character-level coding refining neural networks.

This paper consists of five parts. The first and second parts give the research status and background. The third part is the translation model of fusion character encoding and negotiation network. The fourth part shows the experimental results and analysis. The experimental results of this paper are introduced and compared and analyzed with relevant comparison algorithms followed. Finally, the fifth part concludes the full paper.

3. Translation Model of Fusion Character Encoding and Negotiation Network

3.1. Multifeature Fusion Based on Transformer. Although the word-level vector in neural machine translation has achieved good results, there are still many inevitable defects, for example, it cannot solve the problem of accurate expression of rare words and words that do not appear in the training vocabulary, so it is generally only used in the translation of some languages with rich corpus, such as English, German, and French. Sometimes researchers solve this problem by increasing the vocabulary, but the algorithm complexity during training and decoding increases linearly with the vocabulary used, leading to a vicious circle [25]. Faced with these existing problems, some scholars proposed character-based neural machine translation model in 2017, as shown in Figure 1. One of the advantages of letter-level coding is that it is more suitable for multilanguage translation than word-level model, while word-level coding requires each language to use a separate vocabulary [26].

First, the input sequence is mapped to the corresponding character embedding vector and then the convolution operation is carried out using the convolution kernel with different window sizes, and then the output is connected. For example, in Figure 1, there are three convolution kernels whose window sizes are 3, 4, and 5, which are equivalent to learning character-based triples. Then the output of the convolution layer is input to the maximum pooling layer, which is equivalent to selecting the most important features to generate fragments and embed them. So far, from the initial input character embedding, we can get the fragment embedding which the system considers as language synthesis. Then, all fragments are embedded through the Highway Network layer (its function is similar to residual Network, and information flow is controlled by adding some gating systems) and two-way GRU, so that the encoder output can be obtained by the model at last. Finally, the decoder uses the attention mechanism and character-level encoding GRU Network decoding [27, 28].

3.2. Character Level Encoding Based on CNN. Since a single character contains little information and no rich semantic

information, the input sequence is supplemented by convolution and GLU network, and finally the splicing word vector is input into transformer for training, as shown in Figure 2. As for GLU mentioned above, this section will introduce its specific structure and principle in detail. Its schematic diagram is shown in Figure 2, and the formula is as follows [29]:

$$h_l(X) = (X * W + b) \otimes \sigma(X * V + c), \quad (1)$$

where the input sequence is represented by X , L represents the number of layers, W and V represent different convolution kernels, respectively, and b and c represent the corresponding bias, respectively, and σ is the sigmoid activation function. Half of the convolution is normalized by sigmoid activation function, which controls what values of the other half of the convolution can be transferred to the next GLU layer. To put it simply, a gating mechanism is added to the original convolutional network, which is equivalent to an output gate in LSTM, enabling the model to learn and control the output of the convolutional layer and improving the interpretability of the model. The gating mechanism adjusts the flow direction of data information in the network, and some scholars have proved its reliability in the circular network.

The gradient calculation formula of GLU is shown in (2), this can generally be thought of as a kind of multiplicative hop connection that allows gradients to pass smoothly through each layer. This allows gradient information to flow smoothly in multiple time steps. Without these gating mechanisms, gradient disappearance may occur in the transition of the information. In addition, the network structure allows the model to perform nonlinear transformation. With GLU's multilayered stack structure, long-term dependencies can be captured without the need for gradient extinction.

$$\nabla[X \otimes \sigma(X)] = \nabla X \otimes \sigma(X) + X \otimes \sigma'(X) \nabla X. \quad (2)$$

3.3. Fusion of Multifeature Codes. In this chapter, the direct vector splicing strategy is adopted to connect word vector TW and character-level encoding vector (character-level encoding model output in the previous section) to obtain the final representation vector of word. In this chapter, this seemingly simple method is adopted, but the experimental results show that it is very efficient. This strategy is frequently used not only in the research of neural machine translation, but also in its other natural language processing tasks, through vector splicing:

$$x_t = [w_t; chr_t], \quad (3)$$

x_t is the final representation vector, chr_t is character-level encoding vector. In this section, all corpus data collected will be fully utilized to encode the input at word level and character-level, respectively, and then be used as the feature vector of the final text by means of splicing. In the face of rare words and unknown words that do not appear in the word list of the training set, more information can be obtained by

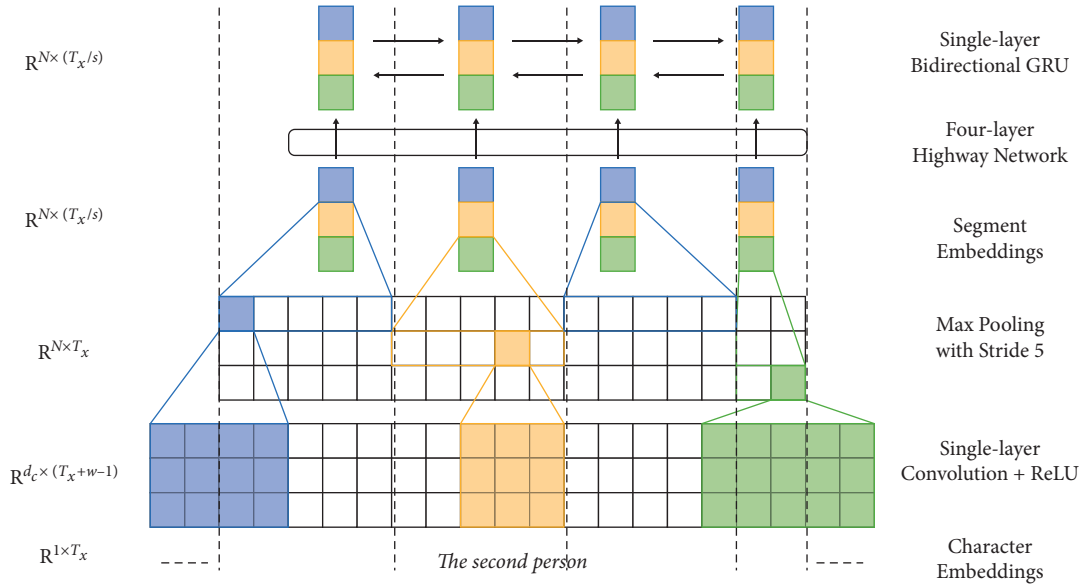


FIGURE 1: Word vector coding based on convolutional neural network.

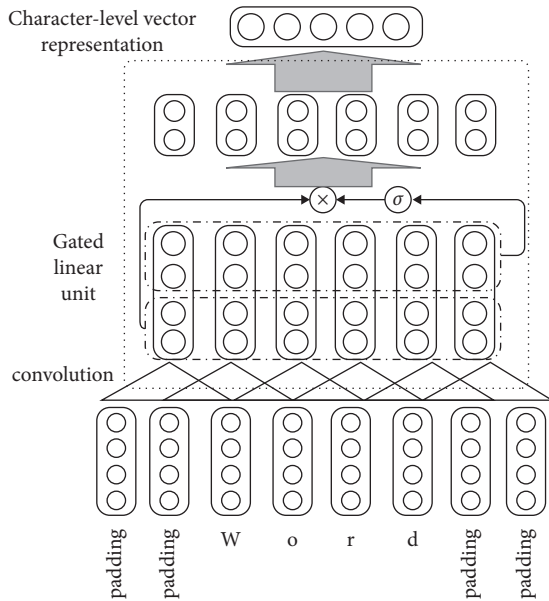


FIGURE 2: Schematic diagram of character-level coding model.

character-level coding vector, so as to alleviate the impact of this kind of problem. Retraining of word-level coding can obtain more semantic and sentential information, which is conducive to the integrity of the whole sentence information. When training the front part of neural network. The encoding of word vector is first trained. After the model converges, character-level encoding is added to ensure the high efficiency of the model. In the pretraining stage of the model, Transformer model structure is used, and the loss function of cross entropy type is adopted in this chapter for modal optimization. Just like the conventional neural machine translation model, attention mechanism is also added to the network. The formula for attention calculation is as follows:

$$ctx_e = \sum_{i=1}^{T_x} \alpha_i h_i,$$

$$\alpha_i \propto \exp(v_\alpha^T \tan h(W_{at,h}^c h_i + W_{att,s}^c \hat{s}_{j-1})) \forall i \in [T_x],$$

$$\sum_{i=1}^{T_x} \alpha_i = 1. \quad (4)$$

\hat{s}_{j-1} is the corresponding hidden layer state quantity, T_x present the time point, α_i is the parameter, and W is the weight. Different from the conventional network based on encoding and decoding mechanism, there is more D2 part in the deliberative network. The specific model calculation details are as follows:

$$ctx_c = \sum_{j=1}^{T_j} \beta_j [\hat{s}_j; \hat{y}_j],$$

$$\beta_j \propto \exp(v_\beta^T \tan h(W_{dt,j}^d [\hat{s}_j; \hat{y}_j] + W_{dt,s}^d s_{t-1})) \forall i \in [T_y],$$

$$\sum_{j=1}^{T_j} \beta_j = 1. \quad (5)$$

From the above calculation process, it can be known that at moment T , the second decoding calculates the attention of each sequence generated by the first, which is equivalent to referring to the words after moment T in the second decoding process. Due to the RNN structure based on the neural network, there are two decoder parts and the target prediction space is very large, it is very difficult to calculate, gradient calculation is also very difficult, so the network uses Monte Carlo method to optimize the maximum likelihood loss.

However, in the syntactic feature model proposed in this paper, it is stored according to units. The syntactic feature

unit model will be established below, which directly stores the syntactic relations between words into corresponding syntactic units. The unit construction rules are as follows:

$$M_i = Mx_i, MPx_i, MCx_i, MBx_i. \quad (6)$$

M_i represents the storage syntactic unit of the i th word in a sentence. MPx_i, MCx_i, MBx_i denote the position of the i th parent node word, child node word, and adjacent node word in the sentence. The mathematical definition of conditional random fields is: suppose an undirected graph

$$G = T(V, E), \quad (7)$$

where V is the set of vertices and E is the set of edges. Let X be the condition that satisfies Y_v , then the variable Y_v satisfies the following equation:

$$p(Y_v|X, Y_u, u \neq v) = p(Y_v|X, Y_u, u: v), \quad (8)$$

where u, v represents two vertices contained in graph T , then (X, Y) is a random field of strips.

4. Experimental Results and Analysis

4.1. Introduction to Experimental Environment and Data Set.

In order to compare the quality of multiattentional neural machine translation models, we conducted several experiments on the (a) WMT 14 English-German translation task. The training data set of the experiment is provided by WMT 14 English-German translation task. After preprocessing the corpus (deleting too long or too short sentences, removing blank lines, etc.), it contains a total of 4.5 million English-German sentence pairs, involving about 116M English words and 110M German words. Meanwhile, (b) newstest 2014 English-German Parallel Corpus database was used as the test set, which contained 2737 parallel statement pairs, and (c) Newstest2013 English-German parallel Corpus database was used as the verification set, which contained 3000 parallel statement pairs. Then byte pair encoding is used to process the parallel corpus training set test set and verification set, and a fixed dictionary with a size of 32000 is extracted from the training data set, which is mixed with the source language lists with the highest frequency words and target language words are shared between input text and output text, and the dictionary does not change during training [28]. All the models in this paper are coded by Python language, and all the experiments in this paper are carried out on a hardware device of NVIDIA 1080Ti GPU.

In the training process of the model, the model uses the stochastic gradient descent algorithm to optimize the model parameters, and the initial value of the learning rate is 1.0, and the maximum number of training steps is set to 340K. After 170K steps, the learning rate is halved every 17K steps. At the same time, the ownership weight parameters in the model are initialized to 0.1, and the deviation is initialized to 1.0. In order to prevent the generation of gradient explosion, gradient clipping algorithm and dropout algorithm (dropout probability set to 0.2) were used in the training so that the norm of the gradient would not exceed 5.0.

4.2. *Experimental Results Analysis.* In order to verify the performance of the proposed LSTM-based attention embedding translation model, the standard LSTM model and the LSTM model of attention embedding were used to train the experimental data set. The results are shown in Figure 3. As can be seen from the figure, compared with the standard LSTM model, the BLEU value of the attention-embedded LSTM model is higher, indicating that the attention-embedded LSTM model has better translation effect for long sentences and the model performance is effectively improved.

As shown in Table 1, the results are mainly shown and compared with some previously well-known models that are relevant to this study. The algorithm is briefly explained as follows:

RNNsearch: This model is the first to bring the attention mechanism to the field of neural translation, which has special significance for the development of the following attention mechanism and machine translation. Their encoder and decoder are based on the RNN of the gated mechanism. The encoder generates the required middle hidden layer vector, and the decoding process passes the note. The semantic mechanism solves the context representation of the output current moment, and finally combines the decoding output of the upper moment with the context representation to complete word prediction.

CovS2S: This model is a neural translation model proposed by Facebook based on convolutional neural network, which preserves the long-distance dependence of words in sentences through gating mechanism combined with multihop convolution and solves the problem of gradient descent disappearing in the training process. This model relies on the pattern of recurrent neural network for neural translation break, improve the model training parallel ability, and model training efficiency in the translation effect is also very good.

Deep-Att + posUnk: This model is a neural translation model proposed by Facebook based on convolutional neural network, which preserves the long-distance dependence of words in sentences through gating mechanism combined with multihop convolution and solves the problem of gradient descent disappearing in the training process. This model relies on the pattern of recurrent neural network for neural translation break and improve the model training parallel ability, and model training efficiency in the translation effect is also very good.

GNMT + RL: This network proposes corresponding solutions to the situation that the calculation cost is too high in the process of deep learning training and translation, and there are rare words in the translation sentences by adopting low-precision algorithms and cutting words into common wordpieces.

Transformer: This model is only based on the attention mechanism and the feedforward neural network

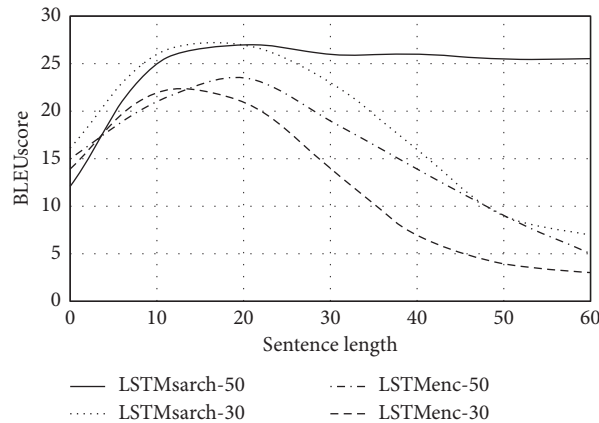


FIGURE 3: Translation performance of attention mechanism.

TABLE 1: BLEU value of different models.

Model	German-English translation	English-French translation
RNNsearch-50 + UNK	—	36.2
ConvS2S	25.2	40.8
Deep-Att + posUnk	—	39.5
GNMT + RL	24.9	39.8
Transformer	27.8	38.2
Transformer (big)	29.9	41.2
Ours	26.5	38.9

method, which can train the model in parallel to shorten the training time. The self-attention mechanism is adopted to shorten the distance between sentences. This model not only improves the translation effect, but also brings researchers a new model design idea.

In order to verify the translation effect of the proposed English machine translation model based on LSTM attention embedding, the research adopts the traditional LSTM English machine translation model RNN English machine translation model Gru-attention English machine translation model, respectively. As can be seen from the figure, the BLEU value of the proposed English machine translation model based on LSTM attentional force embedding on the development machine and test set. Both are higher than the BLEU values of the traditional LSTM and RNN Gru-attention machine translation models, indicating that the proposed translation model improves the performance and translation quality by embedding attention machine in the LSTM network compared with the comparative translation models. The validity of LSTM-based attention embedding model for English machine translation is verified (Figure 4).

Comparing the three backbone network RNNSearch, ConvS2S, and transformer models in the test set, the transformer model based on self-attention mechanism performs better than RNNSearch and ConvS2S model on translation tasks in both rich and low-resource scenarios, as shown in Figure 5. In Figure 5, from the overall translation performance, the transformer model, and GT; ConvS2S

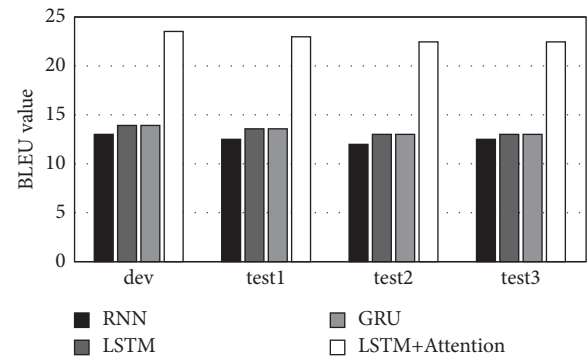


FIGURE 4: BLEU values of different translation models on experimental data sets.

model and gt; RNNSearch model is more obvious in the rich resource scenario, while the difference is not obvious in the low-resource scenario.

As can be seen from Figure 6, the degree of data dependence of NEURAL machine translation can be further verified. No matter in low-resource scenarios or resource-rich scenarios, BLEU value in newstest2016 test set also increases when the scale of training data increases.

However, the trend of BLEU value increase gradually slows down, which means that the performance improvement of pseudo-bilingual corpus constructed by artificial data enhancement is limited and cannot be increased indefinitely. The core of translation performance is still the size

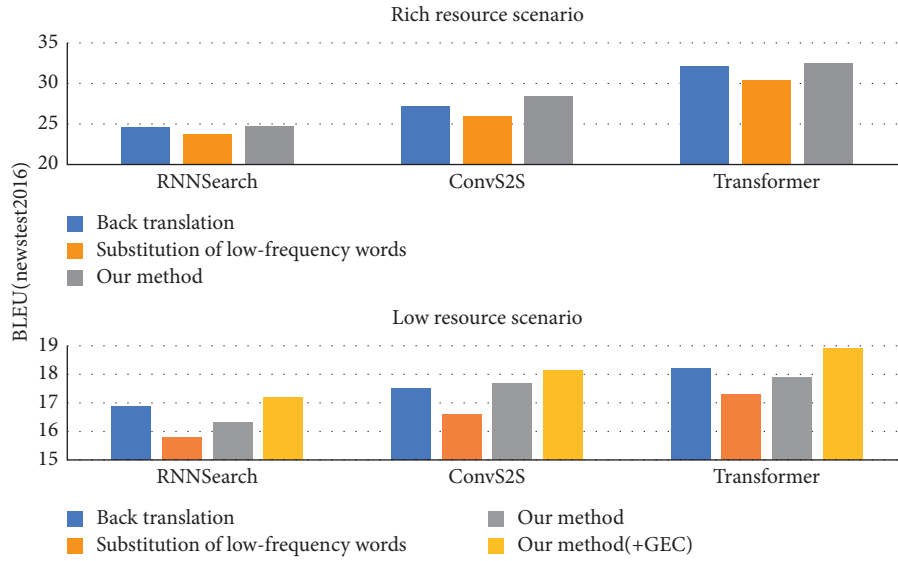


FIGURE 5: Different translation models apply different data enhancement methods to enhance translation performance in different scenarios.

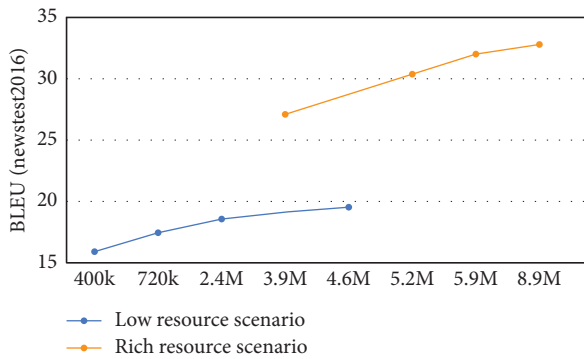


FIGURE 6: Variation relationship between corpus size and BLEU value.

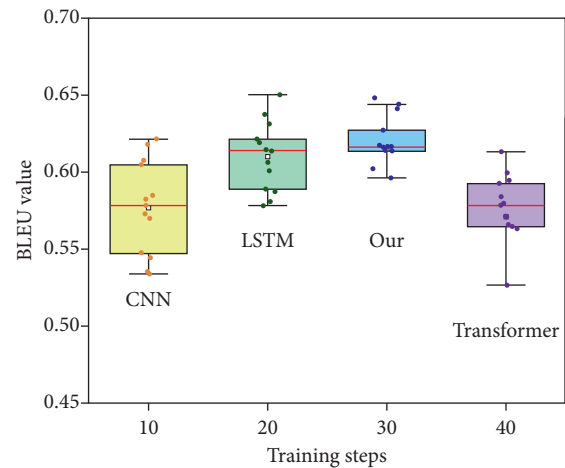


FIGURE 7: BLEU value line chart.

of the real bilingual corpus. Compared with the real parallel corpus scale of 400K in the low-resource scenario, the overall BLEU value of the translation task in the resource-rich scenario is more than 10 times higher than that in the low-resource scenario.

In order to further understand how the performance of the hyperbolic tangent neural machine translation model exceeds that of the comparative model, this paper uses a broken line graph to more intuitively reflect the training process.

As shown in Figures 7 and 8, the horizontal axis of the line graph is the training time step, the vertical axis is the BLEU score, the orange curve is the hyperbolic tangent neural machine translation model of this paper, the blue curve is the RNNSearch model, and the gray one is the Open NMT model. As can be seen from Figures 7 and 8, the curve of the model proposed in this paper is very steep at the beginning of training, indicating that BLEU score value of the model increases greatly during this period, and it is easy to see that the curve of BLEU value of the model proposed in this paper tends to be gentle after steep, indicating that it

soon reaches convergence. However, the curve of the comparison model rises slowly and then reaches a gentle level. From the comparison of the line graph, it can be seen that the model proposed in this paper can greatly improve the training speed, achieve convergence quickly, and save the training time.

In this article, the attention thermal diagram after Sparsemax is adopted in the cross attention layer of decoder part is shown in Figures 9 and 10. As can be seen from the experimental results, after sparse normalization is used to calculate the probability, the corresponding attention score of some unrelated words in Figure 8 is zero (the color is white), which reduces the error size of model data distribution in the problem of induction bias. The accuracy of direct word alignment is improved, which not only improves the effect of the model but also improves the interpretability of the model. Some researchers also adopt the mechanism of local attention to deal with such bias. But in practice,

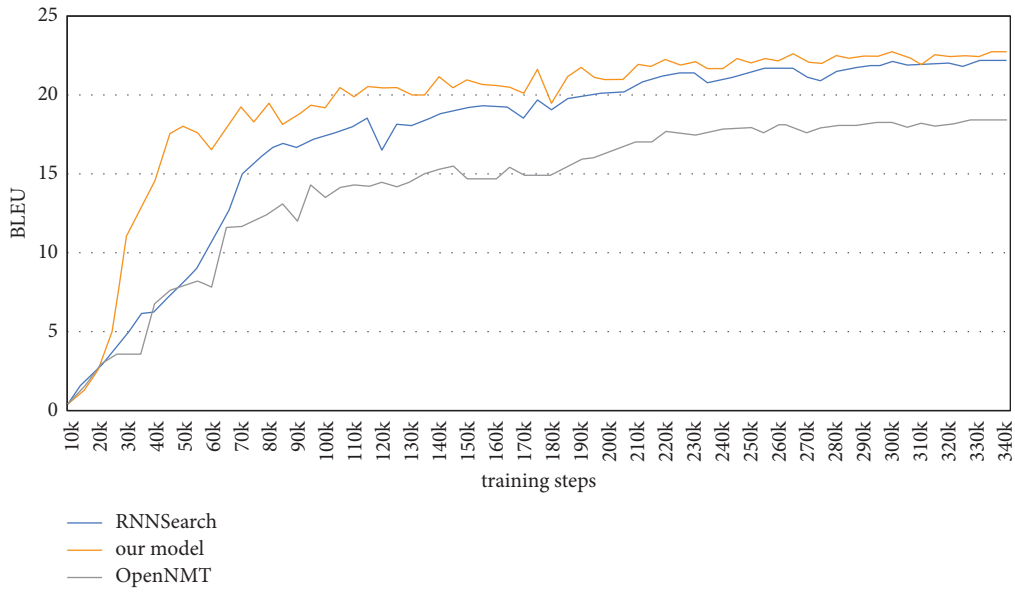


FIGURE 8: BLEU value line chart.

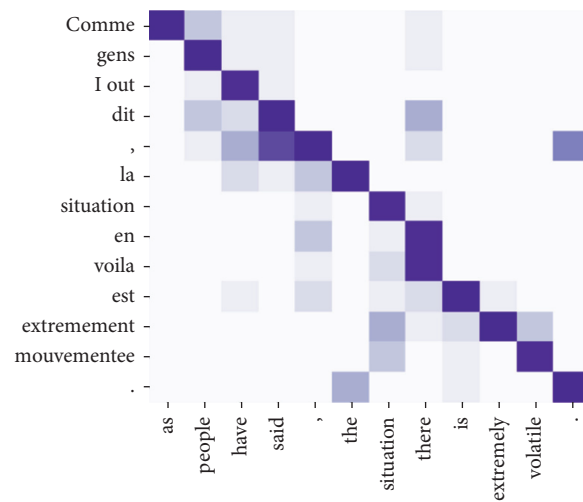


FIGURE 9: Matrix thermogram of attention score in French-English translation.

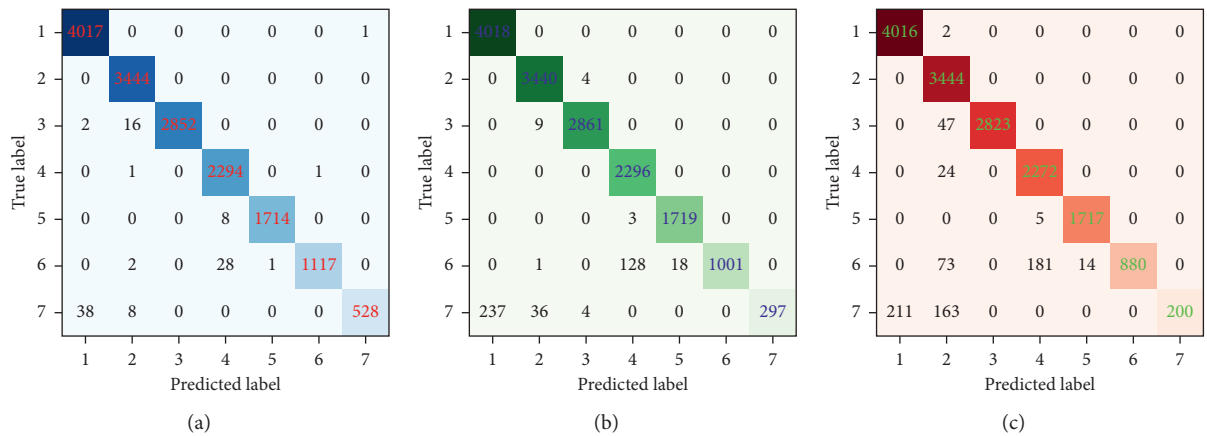


FIGURE 10: Confusion matrix of 7 words translation results: (a) Data set a; (b) Data set b; (c) Data set c.

discreteness and nondifferentiability require Monte-Carlo method to perform gradient approximation, which greatly increases the training complexity of the model. Sparsemax is differentiable, easy to calculate, and easy to use.

5. Conclusions

To study in depth and the rapid development of society now, people more and more demand for artificial intelligence brings the convenience of nerve has important role in machine translation, can help us faster to communicate in daily life, with the development of deep learning neural machine translation also got the attention of researchers and development. Especially in the application of encoder and decoder in natural language processing, the translation performance has been significantly improved. However, there are still many problems in deep learning, which is the focus and hotspot of current deep learning research, but also has many difficulties.

In this paper, transformer models, which are popular in various fields, and some existing problems of machine translation, are proposed to match the research methods and improvement strategies. Whether the text input sequence is represented well or not and whether the information contained is sufficient will directly affect the performance of the whole translation model. In most deep networks, word-level coding is carried out directly, although the effect is not wrong. However, if the embedded word vector is only at the word level, the quality of the corpus is required to be high while the learning effect is guaranteed. In the process of word vector training, the size of the model has a great influence on the number of iterations and translation effect. Therefore, on the basis of the current research work, this chapter uses the character-level vector embedding to ensure that the word vector can integrate more semantic information. Future research may study the specific translation performance of single decoding and integrates the global information of the first translation by integrating the idea of elaboration network, so that the model translation is more smooth and neat, and the generalization error of the whole model is smaller.

Data Availability

The data set can be accessed upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This study was supported by 2021 Educational Science Programming Projects (Special Program for Higher Education), Research on the Role Evolution and Core Competences Construction of Teachers of Application-oriented Undergraduate in Guangdong-Hong Kong-Macao Greater Bay Area (No. 2021GXJK316).

References

- [1] R. Cattoni, M. A. Di Gangi, L. Bentivogli, M. Negri, and M. Turchi, "MuST-C: a multilingual corpus for end-to-end speech translation," *Computer Speech & Language*, vol. 66, Article ID 101155, 2021.
- [2] C. Palladino, M. Foradi, and T. Yousef, "Translation alignment for historical language learning: a case study," *Digital Humanities Quarterly*, vol. 15, no. 3, pp. 11–24, 2021.
- [3] W. Ling, "Research on the application of artificial intelligence in machine translation," *Journal of Frontiers in Engineering Technology*, vol. 1, no. 1, pp. 1–3, 2021.
- [4] L. Deng, "The process-oriented assessment model of business English translation course in a flipped learning context," *Higher Education Studies*, vol. 10, no. 4, pp. 1–11, 2020.
- [5] L. Bowker, "Machine translation literacy instruction for international business students and business English instructors," *Journal of Business & Finance Librarianship*, vol. 25, no. 1-2, pp. 25–43, 2020.
- [6] L. Gao, "A study on the application of functional equivalence to business English E-C translation," *Theory and Practice in Language Studies*, vol. 8, no. 7, pp. 759–765, 2018.
- [7] C. A. I. Ruizhen, "Status quo, problems and countermeasures of research on teaching of business English translation in China," *Journal of Ningbo University of Technology*, vol. 03, pp. 25–37, 2017.
- [8] W. Hu, "On legal English translation from the perspective of legal linguistics," *Review of Educational Theory*, vol. 2, no. 3, pp. 6–10, 2019.
- [9] J. Li, "Application of nord's functionalist translation theory to the translation of English business news discourses," *International Journal of Social Science and Education Research*, vol. 2, no. 5, pp. 46–56, 2019.
- [10] L. Shi and Y. Liu, "A study on equivalence between English translation and cross-cultural communication based on translation coordination theory," *Revista Argentina de Clinica Psicológica*, vol. 30, no. 1, p. 283, 2021.
- [11] Q. Xu and L. Deng, "Investigating the use of translation continuation tasks in commercial translation teaching: a study on translating user manuals," *Chinese Journal of Applied Linguistics*, vol. 44, no. 3, pp. 366–381, 2021.
- [12] D. Viorela-Valentina, "Translation practice—A means for enhancing student employability," *Dialogos*, vol. 22, no. 38, p. 215, 2021.
- [13] A. V. Novikova and L. A. Mylnikov, "Problems of machine translation of business texts from Russian into English," *Automatic Documentation and Mathematical Linguistics*, vol. 51, no. 3, pp. 159–169, 2017.
- [14] S. Y. La and L. W. Wan, "Research on the LSTM Mongolian and Chinese machine translation based on morpheme encoding," *Neural Computing & Applications*, vol. 32, no. 1, pp. 41–49, 2020.
- [15] K. Shuang, Z. Zhang, J. Loo, and S. Su, "Convolution-deconvolution word embedding: an end-to-end multi-prototype fusion embedding method for natural language processing," *Information Fusion*, vol. 53, pp. 112–122, 2020.
- [16] Z. Huang, F. Yang, F. Xu, X. Song, and K.-L. Tsui, "Convolutional gated recurrent unit-recurrent neural network for state-of-charge estimation of lithium-ion batteries," *IEEE Access*, vol. 7, pp. 93139–93149, 2019.
- [17] H. Choi, K. Cho, and Y. Bengio, "Fine-grained attention mechanism for neural machine translation," *Neurocomputing*, vol. 284, pp. 171–176, 2018.

- [18] K. Ahmed, N. S. Keskar, and R. Socher, "Weighted transformer network for machine translation," 2017, <https://arxiv.org/abs/1711.02132>.
- [19] M. Awad, S. M. Ibraheem, S. A. Napoleon, W. Saad, M. Shokair, and M. E. Nasr, "Secrecy enhancement of cooperative NOMA networks with two-way untrusted relaying," *IEEE Access*, vol. 8, pp. 216349–216364, 2020.
- [20] E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui, and J. Long, "A survey of clustering with deep learning: from the perspective of network architecture," *IEEE Access*, vol. 6, pp. 39501–39514, 2018.
- [21] K. Chen and M. D. Sacchi, "Time-domain elastic Gauss-Newton full-waveform inversion: a matrix-free approach," *Geophysical Journal International*, vol. 223, no. 2, pp. 1007–1039, 2020.
- [22] L. Deng, "The project-based flipped learning model in business English translation course: learning, teaching and assessment," *English Language Teaching*, vol. 11, no. 9, pp. 118–128, 2018.
- [23] L. Wu and L. Wu, "Research on business English translation framework based on speech recognition and wireless communication," *Mobile Information Systems*, vol. 2, pp. 14–19, 2021.
- [24] H. Wang, "Marine environment salinity measurement based on data classification system and features of business English translation," *Arabian Journal of Geosciences*, vol. 14, no. 15, pp. 1–14, 2021.
- [25] Z. Cao, "Retracted article: evaluation of air pollution and business English vocabulary translation in coastal cities based on distributed storage," *Arabian Journal of Geosciences*, vol. 14, no. 17, pp. 1–7, 2021.
- [26] Y. Jiang, "Critical thinking oriented teaching reform of business English translation," *Studies in Literature and Language*, vol. 19, no. 1, pp. 96–99, 2019.
- [27] H. Feng, I. Crezee, and L. Grant, "Form and meaning in collocations: a corpus-driven study on translation universals in Chinese-to-English business translation," *Perspectives*, vol. 26, no. 5, pp. 677–690, 2018.
- [28] J. Lyu, "The modularized construction on translation competence for business English majors in China," *English Language Teaching*, vol. 13, no. 7, pp. 124–129, 2020.
- [29] Z. Li, "The construction of the turning classroom of business English translation teaching in higher vocational education under the Internet+ environment," *Frontiers in Educational Research*, vol. 2, no. 5, 2019.