

Research Article

Generative Adversarial Network-Based Short Sequence Machine Translation from Chinese to English

Wenting Ma,¹ Bing Yan,¹ and Lianyue Sun ²

¹English Major, Art and Law Department, Hua Xin College of Geo University, Shijiazhuang, Hebei 050700, China

²Shijiazhuang Posts and Telecommunications Technical College, Shijiazhuang, Hebei 050020, China

Correspondence should be addressed to Lianyue Sun; sunly@sjzpc.edu.cn

Received 19 November 2021; Accepted 18 December 2021; Published 28 January 2022

Academic Editor: Baiyuan Ding

Copyright © 2022 Wenting Ma et al. 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 acceleration of economic globalization, the economic contact, information exchange, and financial integration between countries become more and more frequent. In this context, the communication between different languages is also closer, so accurate translation between languages is of great significance. However, existing methods give little thought to short sequence machine translation from Chinese to English. This paper designs a generative adversarial network to solve the above problem. First, a conditional sequence generating adversarial net is constructed, which includes two adversarial submodels: a generator and a discriminator. The generator is designed to generate sentences that are difficult to distinguish from human-translated sentences, and the discriminator is designed to distinguish the sentences generated by the generator from human-translated sentences. In addition, static sentence-level BLEU values will be used as reinforcement targets for the generator. During training, both dynamic discriminators and static BLEU targets are used to evaluate the generated sentences, and the evaluation results are fed back to the generator to guide the generator's learning. Finally, experimental results on English-Chinese translation dataset show that the translation effect is improved by more than 8% compared with the traditional neural machine translation model based on recurrent neural network (RNN) after the introduction of generative adversative network.

1. Introduction

As one of the important tools for human communication, language is the fundamental ability for human beings to distinguish themselves from other creatures. There are more than 5,600 languages in the world. The diversity of languages promotes the development of cultural diversity, but it also sets up barriers for human communication in different regions [1–3]. With the acceleration of globalization, the contradiction between cross-language communication and language gap becomes more and more obvious. The existing human resources and translation capacity are insufficient to cope with the translation needs of large-scale texts and multiple language pairs in the future. Therefore, machine translation (machine translation) has the advantages of low cost, fast speed, and no language restriction. Translation technology is regarded as the most effective way to break the barrier of language translation. Machine translation (MT) is

a process of automatic translation from one language (source Language) to another language (target language) using computers. In recent years, machine translation technology has developed vigorously and become one of the research hotspots of natural language processing (NLP) task [4, 5].

After more than 70 years of development, machine translation has developed from rule-based machine translation to instance-based machine translation and then to statistics-based machine translation (SMT) [6] and to the present neural network-based neural machine translation (NMT). With the rapid development of deep learning, the performance of neural machine translation has significantly surpassed. It has not only become the hotpot research method of language translation, but also the core technology of commercial online machine translation systems such as Google and Baidu. Neural machine translation models generally adopt encoder-decoder framework. The encoder

model translates the source sentences and encodes them into a set of hidden layer vector representations with fixed dimensions. The decoder generates the word sequence of the target end word for word according to the hidden layer vector output by the encoder. Encoders and decoders of neural machine translation models can be realized by different network structures, which are divided into self-attention network (SAN), convolutional neural networks (CNN), and recurrent neural networks (RNN) according to the characteristics of different network topologies. Compared with statistics-based and rule-based machine translation methods, the neural machine translation model does not require such steps as word extraction from translation rules, word alignment, and ordering but completely relies on the neural network to reflect the relationships between source language and target language automatically, which greatly simplifies the complexity of the model [7].

Although the results of neural machine translation have been enhanced largely under the condition of large-scale training corpus and powerful computing power, there are still many problems to be solved in neural machine translation, such as out of vocabulary (OOV), overtranslation, undertranslation [8], and exposure bias [9]. These problems become the development bottleneck of neural machine translation and restrict the further improvement of translation performance.

2. Related Works

Based on the coverage idea in statistical machine translation, coverage mechanism is introduced into neural machine translation model in [10], and coverage vector (CV) is used to store the historical translation in the decoding process and integrate it into the calculation process of attention weight, so as to guide the attention mechanism to give more resource to the untranslated words and reduce the weight of translated words. As for the phrase-based translation model, literature [11] marks the corresponding source language phrase as translated when adding the translation results of candidate phrases into the output sequence, so as to ensure that each source language phrase will not be repeated or omitted in the translation process. However, because there is no similar mechanism for explicitly storing translation history information in the neural machine translation model, and all source words are involved in the prediction of target words, overtranslation and missing-translation are inevitable. Literature [12] uses full coverage embedding vector to represent the translation degree of source language words in the decoding process and reduces the role of translated words in future decoding by reducing the encoding vector of translated words. In literature [12], it combines a recurrent attention mechanism and conditioned decoder to provide more ordering information in the translation process, thus reducing the phenomenon of repeated translation. In [13], the authors use two circulating neural networks to store the past information and future information in translation process, respectively, and use this information to guide the attention mechanism and decoding state. These methods can alleviate the phenomenon of

overtranslation and missed translation to a certain extent but cannot completely eliminate the problem [14].

Since the neural machine translation model often selects sentences with high probability but short length as translation results, coverage can also be used as an evaluation index to screen translation results [15]. In [16], coverage penalty is introduced into the beam search algorithm to make the model consider the probability of sentence generation and fidelity to the original text at the same time when selecting the translation, so as to avoid the bias towards short sentences. On this basis, literatures [17–19] introduced the detection of coverage into each cluster search step and improved the calculation method of coverage score (CS), making it suitable for a variety of mapping relations between source language and target language. These methods select better translation results by improving search evaluation methods without changing the structure of the neural machine translation model [20, 21].

In order to give consideration to the fluency and fidelity of the target text, the context gate structure is introduced in [22, 23] to dynamically control the proportion of the influence of source language and target language context on the generation of words in the target language during decoding. Combined with the overlay mechanism, this method can improve the coverage of the translation results to the source language and the fluency of the sentence, but the target context does not fully utilize the generated translation information. Literature [24] has modeled the structural relations of all words, so that the neural machine translation model can make better use of the context features of source language and target language. In addition, literature [25] studies the features of omitted words in source languages, finds that words with higher translation entropy are more likely to be omitted, and proposes to use coarse-to-fine framework to improve the translation quality of sentences and words and reduce the number of missed translation of words with high entropy.

From the previously mentioned analysis, we know that the previously mentioned methods have alleviated the problem of overtranslation and overtranslation in NMT to some extent, but these problems are still unavoidable due to the soft alignment of attention mechanism and the imperfection of coverage mechanism in NMT model word-for-word prediction. On the other hand, no scholar has applied English-Chinese short sequence machine translation till now, so the research here is still a blank, which has a great theoretical research and practical application value [26–28].

The contribution of this paper is the GAN model, which is first used to solve the problem of short sequence machine translation from Chinese to English. In this paper, a survival adversarial network-based machine translation method is proposed. Specifically, BLEU value is used to generate D for network, and BLEU value is established as feedback to enhance the training of generator G. So, generator G can generate the translation result which is closer to the real sentence, thus improving the accuracy of translation. In future work, we may try to combine more neural machine translation models with generative adversarial networks or use multiadversarial network frameworks to experiment

with different parameters and construct different reinforcement feedback to improve translation effects.

This paper consists of five parts. The first and second parts give the research status and background. The third part is the short sequence machine translation by generative adversarial 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 sixth part summarizes the full paper. The study of this paper gives the user portrait system of operators and gives the realization scene of precision marketing. The paper can complete the analysis function according to the data provided by operators to form user portraits, which has the application value of precision marketing of operators.

3. Short Sequence Machine Translation by Generative Adversarial Network

3.1. GAN Model and Transformer Model. The process of short sequence machine translation by generative adversarial network is divided into two parts, and the overall architecture of the model is shown in Figure 1 [29]. The left half is made up of generator G and discriminator D, where G is the neural machine translation model, which generates target sentences. D discriminates between the sentences generated by G and the artificial translation sentences and generates the feedback results. The right part carries out strategy gradient training for G, and the final feedback is provided by D and Q, where Q is BLEU value.

Another type of encoding and decoding similar to the GAN model is the transformer model, which is shown in Figure 2 [30]. And the best translation effect of current neural machine translation is achieved. Similarly, a set of source language sequences X with input length m and target language sequences Y with corresponding length n are given. Transformer encoders are stacked by N identical network layers. Every network layer contains two subnetwork layers: the first layer is self-attention layer, and the second layer is a fully connected neural network. Residual network and layer normalization (LN) connections are used after each sublayer. Multihead attention network maps input values into various subspaces and then uses scaled dot product attention network to calculate context vectors of different spaces and splices these context vectors to the final output result.

Similar to the encoder, the transformer decoder also contains N isomorphic network layers, each containing three subnetwork layers: the first layer is self-attentional network because the decoder can only see the generated word information when decoding, so mask technology is used to shield the ungenerated word information. The second layer is multiattentional network, which models the hidden state of source language sentences and the hidden state of target language to generate the context vector of source language sentences. The third layer is a fully connected feedforward neural network, which also uses residual networks and hierarchically normalized connections after each sublayer. Since both encoders and decoders based on full attention networks do not consider location information,

which is very important for language understanding and generation, the transformer model includes location coding in the input vector of the lowest level encoder and decoder. Position encoding can be fixed position encoding, relative position encoding, or learned position encoding.

3.2. Generator G. Generator G uses a recurrent neural network (RNN) based neural machine translation model RNN search, which consisted of two parts: the encoder and coder [31]. The encoder adopts bidirectional cyclic gate to control the unit, codes the input sequence formula $x = (x_1, \dots, x_m)$, and calculates the hidden state of forward and backward propagation as follows:

$$\vec{h} = (\vec{h}_1, \dots, \vec{h}_m), \quad (1)$$

$$\overleftarrow{h} = (\overleftarrow{h}_1, \dots, \overleftarrow{h}_m), \quad (2)$$

where the final annotation vector h_p is calculated by \vec{h}_p and \overleftarrow{h}_p together.

The encoder uses recurrent neural network to predict target sequence $y = (y_1, \dots, y_m)$. The current prediction of each word y_i is calculated based on the current hidden state s_i and the prediction y_{i-1} of the previous moment, as well as the context vector c_i , where c_i is derived from the weighted sum of the annotated vectors h_p .

3.3. Discriminator D. The discriminator uses convolutional neural network structure to discriminate the results generated by the generator. Since the length of the sequence generated by the generator is variable, the sequence is filled with a fixed length of T . Source matrix $X_{1:T}$ and target matrix $Y_{1:T}$ are established, respectively, for source sequence and target sequence as follows:

$$X_{1:T} = x_1; x_2; \dots; x_T, \quad (3)$$

$$Y_{1:T} = y_1; y_2; \dots; y_T. \quad (4)$$

When x_t and $y_t \in R^k$ are both the k dimensional word vectors, the convolution kernel $W_1 \in R^{l \times k}$, and the convolution calculation formula is shown as follows:

$$c_{ji} = f(W_1^* X_{i:i+l-1} + b), \quad (5)$$

where W_1^* is the weight of the convolution kernel, $X_{i:i+l-1}$ is the vector matrix in the window from i to $i+l-1$, b is the bias term, f is the activation function, and ReLU function is adopted in this paper.

After using different convolution kernels for convolution operation, we conducted the maximum pooling (Max) operation on feature vector c_{ji} of each convolution kernel, that is, proposing the maximum value of each feature vector, splicing the pooling result, and obtaining the feature vector of the source sequence, as follows:

$$c_x = [c_1, c_2, \dots, c_{T-l+1}]. \quad (6)$$

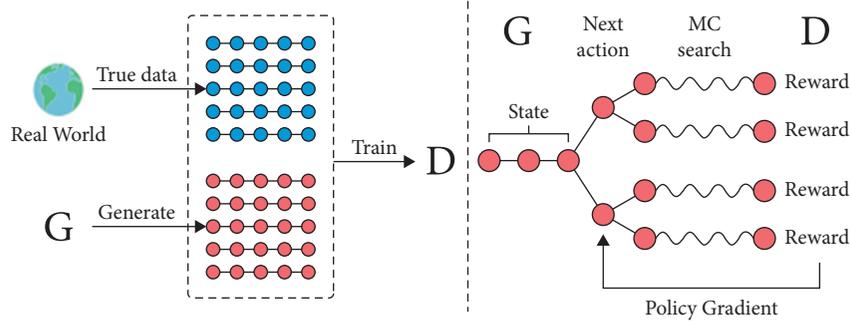


FIGURE 1: The process of short sequence machine translation by GAN.

Similarly, the feature vectors c_y of the target sequence are given by the matrix $Y_{1:T}$. The calculation of the probability of target sequence depends on c_x and c_y , which can be expressed as follows:

$$p = \varphi(V[c_x; c_y]), \quad (7)$$

where V is the parameter matrix, $[c_x; c_y]$ are converted into two-dimensional vectors, and φ is the softmax function.

3.4. Strengthening Target Feedback. BLEU value Q of bilingual translation quality assessment is used as the reinforcement target, and the guidance generator can generate higher BLEU value. Q is a static function and is not updated during training. We apply BLEU value as the specific target of generator and assign the sentence y_g generated by generator and the artificially translated sentence y_d . By calculating the n-gram accuracy of y_g , the feedback $Q(y_g, y_d)$ of

target Q is obtained. The same as the output of discriminator D , $Q(y_g, y_d)$ also ranges from 0 to 1, which makes it easier for Q and D to merge.

3.5. Strategy Gradient Training. Since the goal of generator G is defined as maximizing the expected feedbacks from the initial state of the generated sequence. Formally, the objective function is as follows:

$$J(\theta) = \sum_{Y_{1:T}} G_\theta(Y_{1:T} | X) \cdot R_{D,Q}^{G_\theta}(Y_{1:T-1}, X, y_T, Y^*), \quad (8)$$

where θ is a parameter of generator G .

The discriminator provides N feedbacks for N sentences sampled, and the resulting feedbacks are the summation average of these feedbacks. For the target sentence with length T , the feedback calculation of y_T is shown as follows:

$$R_{D,Q}^{G_\theta}(Y_{1:t-1}, X, y_T, Y^*) = \begin{cases} \frac{1}{N} \sum_{n=1}^N \lambda(D(X, Y_{1:T_n}^n) - b(X, Y_{1:T_n}^n)) + (1 - \lambda)Q(Y_{1:T_n}, Y^*), & t < T, \\ \lambda(D(X, Y_{1:t}) - b(X, Y_{1:t})) + (1 - \lambda)Q(Y_{1:t}, Y^*), & t = T. \end{cases} \quad (9)$$

The gradient calculation of the objective function $J(\theta)$ to the parameter θ of generator G is shown as follows:

$$\nabla J(\theta) = \frac{1}{T} \sum_{t=1}^T \sum_{y_t} R_{D,Q}^{G_\theta}(Y_{1:t-1}, X, y_T, Y^*), \quad (10)$$

$$\nabla_\theta (G_\theta(y_t | Y_{1:t-1}, X)) = \frac{1}{T} \sum_{t=1}^T E_{y_t \in G_\theta} [R_{D,Q}^{G_\theta}(Y_{1:t-1}, X, y_T, Y^*) \cdot \nabla_\theta \log(G_\theta(y_t | Y_{1:t-1}, X))]. \quad (11)$$

4. Experimental Results and Analysis

4.1. Introduction to Dataset. In this part, experiments on Chinese-English translation tasks are carried out to verify the effectiveness of the proposed neural machine translation method. The training data set used in the Chinese-English

translation task is 1.25 million Chinese-English parallel sentence pairs that are extracted from Linguistic Data Consortium (LDC). The test sets are NIST02 NIST03 NIST04 NIST05 and NIST08 from the National Institute of Standards and Technology in 2002. NIST06 was used as the test set. Byte pair encoder (BPE) encoding is performed for

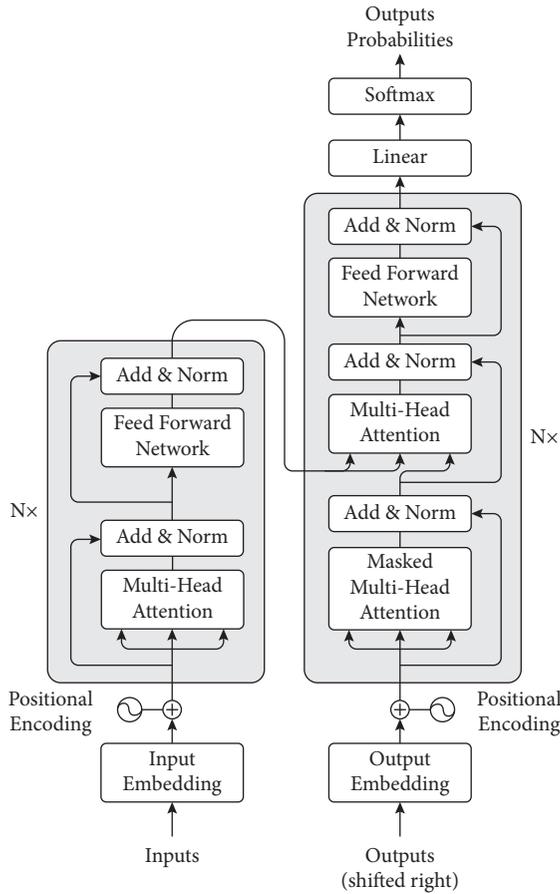


FIGURE 2: The structure of transformer model.

the Chinese and English materials, respectively. The BPE table is used in Chinese and English, respectively, and the size of both tables is 32000. The size of the Chinese is set to 40,000, and the size of the English is set to 30,000. The UNK takes the place of the low frequency of the words that are not in the word table.

4.2. Introduction to the Experimental Platform. In order to verify the effectiveness of the proposed method in machine translation, the experimental hardware environment is Intel (R) Core (TM) I7-8550U CPU @ 1.80 GHz 1.99 GHz, 16.0 GB memory, SSD 512 GB, Windows 10 Professional operating system model training and testing using Google's open source deep learning framework TensorFlow, experimental data analysis software environment, and test environment PyCharm 2019 Professional edition.

4.3. Experimental Results Analysis

4.3.1. Experimental Results of the Data Generalization Method. This section introduces the experimental results given by the data generalization based neural machine translation model. During the test, the 10 models with the highest BLEU value in the training stage are averaged to get the final experimental results of the model, as shown in Table 1 (the best result is shown in bold).

In Table 1, RNNSearch and Transformer are the experimental results of the baseline system on the original corpus, while RNNSearch_G and Transformer_G are the experimental results of processing the unknown words in the corpus using the data generalization strategy. The last row is the proposed method, and the experimental results (in bold) show that the performance of the machine translation model is significantly improved by the generalization of unknown words. The average BLEU values of the RNNSearch, RNNSearch_G, and Transformer_G models are 21 and 22.9, 22.8, 23.5, and 28.2, respectively. That is to say, the proposed methods is 4.7 percentage points higher than the suboptimal model, which has proved the superiority and effectiveness of the proposed method.

In order to examine the translation effect of unknown words in the NIST04test set, 200 sentences were randomly selected for manual evaluation. According to the experimental results, the proposed model has a better performance than Transformer_G models, so the experiment is only carried on the proposed model and Transformer_G model. The 200 sentences include 79 time expressions, 124 number expressions, 135 number expressions, and 46 website and special expressions, which are used to evaluate the translation effect of unknown words. The calculation formula is as follows:

$$ACC = \frac{\text{number of untranslated words correctly}}{\text{total number of unlogged words}}. \quad (12)$$

The experimental performances are shown in Table 2 (the best result is shown in bold).

According to the experimental results in Table 2, the generative adversarial network can greatly improve the translation effect of unknown words in corpus. In terms of names, Chinese names account for 72% of the total number of names in the corpus, so the Chinese Pinyin translation method has the highest improvement, reaching 40.1%. Regarding the time, number, website, and special expressions, due to the rules-based method, the accuracy of all translations has been significantly improved, reaching 10.37%, 24.02%, and 34.44%, respectively.

4.3.2. Source Language Sentence Length Influences. The translation effect of long sentences is one of the important indexes to evaluate the performance of neural machine translation models. To study the translation performance of the multicoverage fusion model regarding the sentence length range of different source languages, the source languages in the test set were grouped according to the sentence length in the experiment. The BLEU values of different models in the range of source language length (0, 10], 10, 20], 20, 30], 30, 40, 50 and (50, +) were compared, as shown in Figure 3.

The following conclusions can be drawn from the experimental results shown in Figure 3. Overall, the performance of proposed model is better than that of RNNSearch and Transformer model. In the length interval of (0, 10), the BLEU value of proposed model is lower than that of the Transformer model, but higher than that of RNN search system. With the further increase of sentence length, BLEU values of RNNSearch system and proposed model decreased significantly.

TABLE 1: BLEU scores of different methods incorporating data generalization.

Method	NIST02	NIST03	NIST04	NIST05	NIST06	AVG
RNNSearch	26.2	21.7	19.0	18.4	19.7	21
RNNSearch_G	29.7	22.73	21.9	19.7	22.1	22.9
Transformer	23.4	24.2	23.5	21.6	20.6	22.8
Transformer_G	24.6	24.33	25.7	20.5	21.2	23.5
Proposed	31.2	28.6	26.2	26.8	25.3	28.2

TABLE 2: The translation results of different methods.

Method	Name (%)	Time expression (%)	Numeral expression (%)	URL and special expressions (%)
Transformer_G	31.82	68.29	52.24	52.37
Proposed	71.92	78.66	76.26	86.81

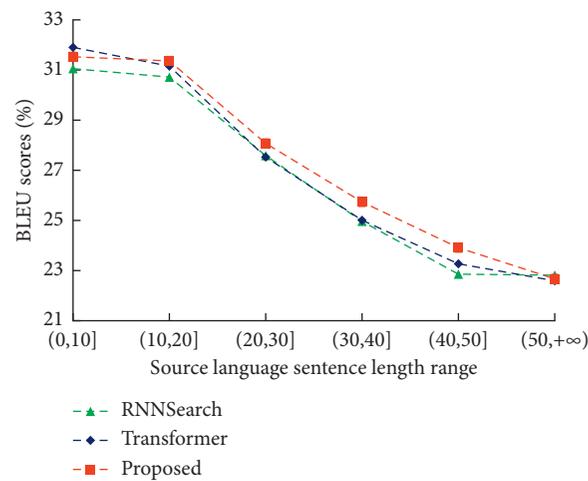


FIGURE 3: The comparison of BLEU scores with respect to the source length.

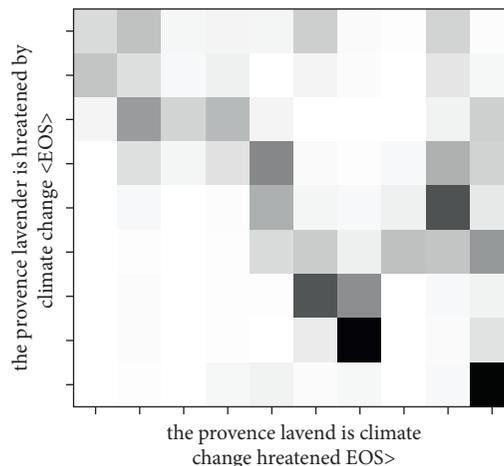


FIGURE 4: The proposed model alleviates the overtranslation phenomenon.

In the experiment, the overtranslation of different models was evaluated by counting the number of source language words in the test set. In Figure 4, the proposed model correctly translates Provence lavender as “Provence lavender.” It not only eliminates duplication of translation but also correctly aligns attention between

Provence and lavender, thus producing higher quality translation results.

As shown in Figure 5(a), in the baseline system, the translation of the word payment is omitted, leading to the loss of context meaning, so mobile is mistranslated as movement. The translation deviates completely from the

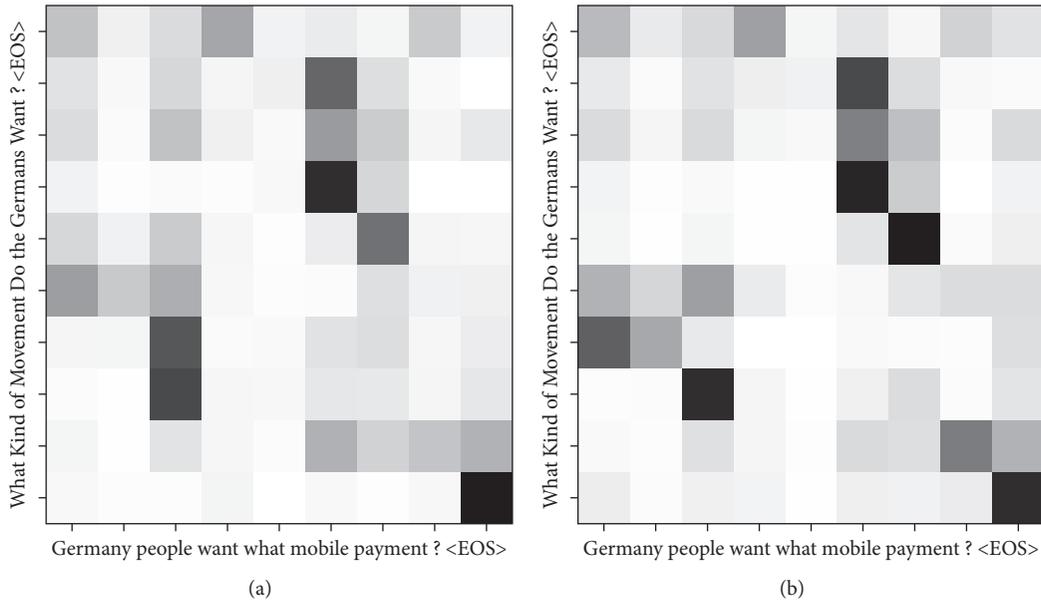


FIGURE 5: The proposed model alleviates the overtranslation phenomenon. (a) Baseline. (b) GAN.

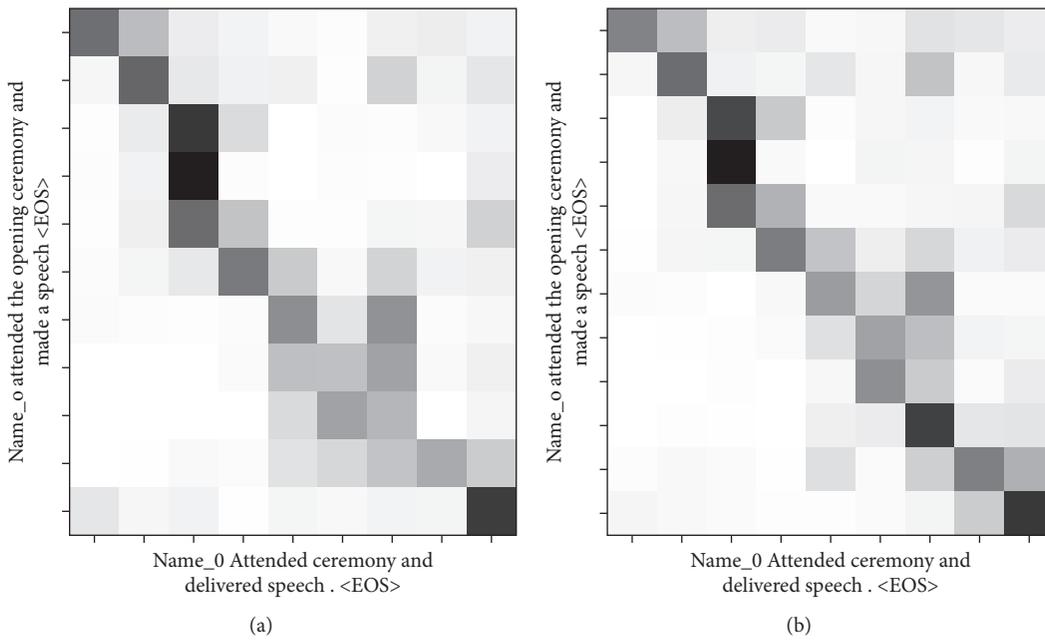


FIGURE 6: Alignment example.

meaning of the original sentence. In contrast, in Figure 5(b), the proposed model eliminates the phenomenon of translation leakage and correctly translates mobile payment into “mobile payment,” thus ensuring the consistency of translation results with the original sentence meaning.

In addition to further alleviating the problems of overtranslation and overtranslation, the multicoverage fusion model can also improve the alignment quality of attentional mechanisms by studying the alignment relationship. As shown in Figure 6, in the baseline system, the second half of the text gives the keynote speech. There was

an alignment error in translation. The keynote and speech were misaligned to speech and make, respectively, and the word keynote was omitted. The proposed model not only correctly establishes the alignment between keynote and speech but also eliminates the phenomenon of missing translation, which improves the quality of translation results.

4.3.3. *Convergence Analysis of the Model.* In order to better demonstrate the performance of the proposed method, its convergence is studied. Figures 7 and 8 show the variation

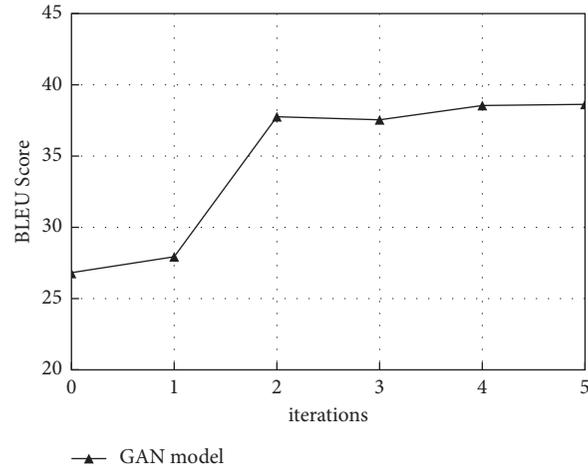


FIGURE 7: Performance curves of Chinese to English models.

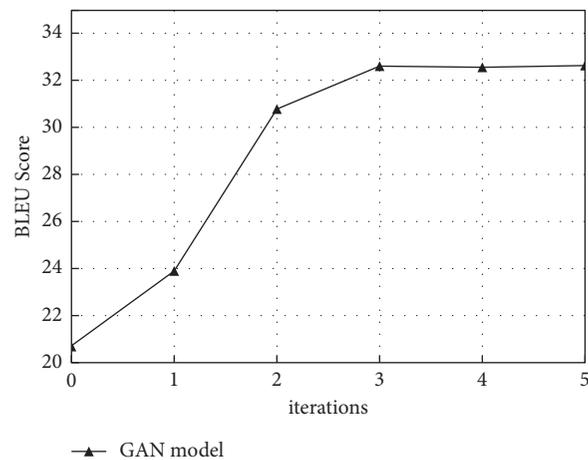


FIGURE 8: Performance curves of English and Chinese models.

curves of the model's performance with each iteration in the experimental process of different translation tasks. Figure 6 shows the model performance change curve of the proposed method in the Chinese-English translation task on the NIST2006 dataset.

Figure 8 shows the performance change curve of Transformer model with the number of iterations of EM algorithm in the translation task of English and Chinese on NIST2006 dataset. It can be seen from Figures 7 and 8 that, after five iterations, the model tends to converge, and it is difficult to improve the performance of the model, which also proves that the method in this paper tends to be stable and has a fast convergence speed after certain iterations. It shows that the proposed method has good performance.

5. Conclusions

Neural machine translation (NMT) model based on encoder and decoder structure is a mainstream approach in the field of machine translation. Although the performance of neural machine translation model has far exceeded that of traditional statistical machine translation

methods, due to the lack of word list size and coverage mechanism, neural machine translation still has the problem of overtranslation and omitted translation of unknown words.

In this context, the communication between different languages is also closer, so accurate translation between languages is of great significance. However, existing methods give little thought to short sequence machine translation from Chinese to English. This paper designs a generative adversarial network to solve the previously mentioned problem. To avoid the loss of coverage information affecting the quality of attention alignment, both the coverage vector and the coverage score are obtained when calculating the attention score. According to the fusion mode of coverage vector and coverage fraction, the GAN model is constructed. Experimental results show that this model can further alleviate the problems of overtranslation and overtranslation, and GAN model has a more obvious improvement effect. In addition, the proposed GAN model has a fast convergence rate.

Although the previously mentioned methods can effectively improve the problem of unregistered words and

overtranslation, there are still some shortcomings. First, a large number of manual rules are needed to achieve the maximum alignment of unknown words, and the translation quality of people's names depends too much on dictionaries. Second, there are still many overtranslation and overtranslation phenomena in the multicoverage fusion model. In the following work, on the one hand, we will expand the scope of unknown words, such as the names of places, organizations, and proper nouns in specific fields, and explore other methods to translate unknown words, so as to reduce the dependence of thesaurus. On the other hand, phrase alignment in statistical machine translation is given to the neural machine translation model as an overlay reference to further improve overtranslation and overtranslation.

Data Availability

The dataset can be accessed upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] D. Shilton, M. Breski, D. Dor, and E. Jablonka, "Human social evolution: self-domestication or self-control?" *Frontiers in Psychology*, vol. 11, Article ID 134, 2020.
- [2] N. M. Makhmudovna, N. G. Davronovna, E. D. Shonazarovna, and M. D. Shavkatovna, "language is a means of cognition and communication for teaching in education," *International Journal of Psychosocial Rehabilitation*, vol. 24, no. 1, pp. 247–256, 2020.
- [3] V. M. Vysotki, E. V. Patalashko, V. E. Stepanenko, T. V. Makarova, and I. A. Balabanova, "Language as a communication resource and its place in the representation of world practices," *Linguistics and Culture Review*, vol. 5, no. S3, pp. 574–584, 2021.
- [4] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang, "Pre-trained Models for Natural Language Processing: A survey," *Science China Technological Sciences*, vol. 63, no. 10, pp. 1–26, 2020.
- [5] C. Park, Y. Yang, K. Park, and H. Lim, "Decoding strategies for improving low-resource machine translation," *Electronics*, vol. 9, no. 10, Article ID 1562, 2020.
- [6] Y. Xia, "Research on statistical machine translation model based on deep neural network," *Computing*, vol. 102, no. 3, pp. 643–661, 2020.
- [7] Y. Belinkov, N. Durrani, F. Dalvi, H. Sajjad, and J. Glass, "On the linguistic representational power of neural machine translation models," *Computational Linguistics*, vol. 46, no. 1, pp. 1–52, 2020.
- [8] K. Bringmann, M. Künnemann, and A. Nusser, "Discrete fr chet distance under translation," *ACM Transactions on Algorithms*, vol. 17, no. 3, pp. 1–42, 2021.
- [9] S. Knobloch-Westerwick, C. Mothes, and N. Polavin, "Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information," *Communication Research*, vol. 47, no. 1, pp. 104–124, 2020.
- [10] B. Zhang, D. Xiong, J. Xie, and J. Su, "Neural machine translation with GRU-gated attention model," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4688–4698, 2020.
- [11] Y. Liu, K. Wang, C. Zong, and K. Y. Su, "A unified framework and models for integrating translation memory into phrase-based statistical machine translation," *Computer Speech & Language*, vol. 54, pp. 176–206, 2019.
- [12] H. Choi, K. Cho, and Y. Bengio, "Fine-grained attention mechanism for neural machine translation," *Neurocomputing*, vol. 284, pp. 171–176, 2018.
- [13] M. Chen, Y. Li, and R. Li, "Research on neural machine translation model," *Journal of Physics: Conference Series*, vol. 1237, no. 5, Article ID 052020, 2019.
- [14] S. Willis, "Mistranslation, missed translation: h l ne cixous' vivre L'orange," *Rethinking Translation*, pp. 106–119, Routledge, England, UK, 1st edition, 2018.
- [15] W. M. Bramer, D. Giustini, and B. M. R. Kramer, "Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study," *Systematic Reviews*, vol. 5, no. 1, pp. 1–7, 2016.
- [16] Z. Ren, D. Dong, H. Li, and C. Chen, "Self-paced prioritized curriculum learning with coverage penalty in deep reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2216–2226, 2018.
- [17] C. Leacock, M. Chodorow, M. Gamon, and J. Tetreault, "Automated grammatical error detection for language learners," *Synthesis lectures on human language technologies*, vol. 3, no. 1, pp. 1–134, 2010.
- [18] S. Sun, C. Luo, and J. Chen, "A review of natural language processing techniques for opinion mining systems," *Information Fusion*, vol. 36, pp. 10–25, 2017.
- [19] C. Leacock, M. Chodorow, M. Gamon, and J. Tetreault, "Automated grammatical error detection for language learners, second edition," *Synthesis lectures on human language technologies*, vol. 7, no. 1, pp. 1–170, 2014.
- [20] F. Cairo, R. Rotundo, P. D. Miller, and G. P. Pini Prato, "Root coverage esthetic score: a system to evaluate the esthetic outcome of the treatment of gingival recession through evaluation of clinical cases," *Journal of Periodontology*, vol. 80, no. 4, pp. 705–710, 2009.
- [21] F. Isaia, R. Gyurko, T. C. Roomian, and C. E. Hawley, "The root coverage esthetic score: intra-examiner reliability among dental students and dental faculty," *Journal of Periodontology*, vol. 89, no. 7, pp. 833–839, 2018.
- [22] M. A. Marrs, B. D. Vogt, and G. B. Raupp, "Comparison of wet and dry etching of zinc indium oxide for thin film transistors with an inverted gate structure," *Journal of Vacuum Science and Technology: Vacuum, Surfaces, and Films*, vol. 30, no. 1, Article ID 011505, 2012.
- [23] F. A. Gers and E. Schmidhuber, "LSTM recurrent networks learn simple context-free and context-sensitive languages," *IEEE Transactions on Neural Networks*, vol. 12, no. 6, pp. 1333–1340, 2001.
- [24] Y. S. G. Kim, "Structural relations of language and cognitive skills, and topic knowledge to written composition: a test of the direct and indirect effects model of writing," *British Journal of Educational Psychology*, vol. 90, no. 4, pp. 910–932, 2020.
- [25] P. Satasivam, B. Y. Poon, B. Ehd ie, A. J. Vickers, and J. A. Eastham, "Can confirmatory biopsy be omitted in patients with prostate cancer favorable diagnostic features on

- active surveillance? The,” *Journal of Urology*, vol. 195, no. 1, pp. 74–79, 2016.
- [26] J. Su, J. Zeng, D. Xiong, Y. Liu, M. Wang, and J. Xie, “A hierarchy-to-sequence attentional neural machine translation model,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 623–632, 2018.
- [27] M. Ghazvininejad, V. Karpukhin, and L. Zettlemoyer, “Aligned cross entropy for non-autoregressive machine translation,” in *Proceedings of the International Conference on Machine Learning*, pp. 3515–3523, Vienna, Austria, April 2020.
- [28] S. Wu, D. Zhang, Z. Zhang, N. Yang, M. Li, and M. Zhou, “Dependency-to-Dependency neural machine translation,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 11, pp. 2132–2141, 2018.
- [29] H. Zhang, Y. Sun, L. Liu, X. Wang, L. Li, and W. Liu, “ClothingOut: a category-supervised GAN model for clothing segmentation and retrieval,” *Neural Computing and Applications*, vol. 32, no. 9, pp. 4519–4530, 2020.
- [30] N. Chiesa, B. A. Mork, and H. K. Høidalen, “Transformer model for inrush current calculations: simulations, measurements and sensitivity analysis,” *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2599–2608, 2010.
- [31] J. C. W. Lin, Y. Shao, Y. Djenouri, and U. Yun, “ASRNN: a recurrent neural network with an attention model for sequence labeling,” *Knowledge-Based Systems*, vol. 212, Article ID 106548, 2021.