IJLAI Transactions on Science and Engineering (ISSN: 3006-3418) 2025, vol. 3, no. 1, pp: 35–44

Chinese-English Machine Translation Based on Generative Adversarial Network

Tianxiao Wang[1]

College of International Business, Shenyang Normal University
Shenyang 11034 China
Received Feb. 28, 2025; Revised and Accepted Mar. 11, 2025

Abstract. Although Chinese-English machine translation is a resource-rich language pair, the problem of data sparsity still exists. For example, for the translation of some specific domains or low-frequency words, the amount of parallel corpus is limited, which makes it difficult to improve the translation quality of the model in these scenarios. Neural machine translation models usually need a large amount of alignment data to train, otherwise they are prone to over-fitting. Neural machine translation models are slow to train and decode, especially when dealing with long sentences or complex structures, which limits their efficiency in real-time application scenarios. Chinese-English machine translation can help people overcome the language barrier and promote the communication and cooperation between people with different language backgrounds, which is of great significance for international business, academic exchanges and cultural exchanges. This paper proposes a Chinese-English neural machine translation model based on generative adversarial network. This model applies the generative adversarial network to neural machine translation, and further optimizes the adversarial learning based neural machine translation model by improving the monotone decoding sequence from left to right or from right to left in the original machine translation model. At the same time, unlike previous generative adversarial networks, neural machine translation models are actually a sequence of discrete symbols that map source language sentences to target language sentences, both in discontinuous spaces. In this case, the generative adversarial network fails to transmit the gradient properly, causing the generator to lose its update direction. By introducing the strategy gradient algorithm in reinforcement learning, the generator optimization problem in adversarial learning is solved, and the translation performance of the model is improved. Finally, experiments on public data sets show that the proposed model can effectively improve translation quality compared with other advanced models.

Keywords: Chinese-English machine translation, Generative adversarial network, Strategy gradient algorithm, Generator optimization.

1. Introduction

From rule-based machine translation system to statistics-based machine translation system and neural machine translation system, with the continuous innovation of machine translation technology, we are getting closer and closer to the era of artificial intelligence. The research shows that although machine translation has achieved good results [1,2], the development of machine translation has not made a big breakthrough to some extent due to the problems such as sparse data, inadequate translation, monotonous translation sequence and unbalanced output. With the development of neural network, neural machine translation has gradually replaced the traditional machine translation. A large number of studies have shown that neural machine translation model has better processing effect than the traditional machine translation [3-5].

Jiang et al. [6] proposed a sequence-to-sequence translation model based on bidirectional recurrent neural networks to solve the problem that traditional translation models could not handle long text sequences. Although end-to-end neural machine translation could capture long-distance dependencies, it was difficult to actually handle them. To this end, by using long short-term memory (LSTM) and Gated Recurrent Neural Network (GRU), it can effectively overcome the long distance dependence problem based on network and the problem of gradient disappearance and gradient explosion in traditional convolutional networks [7,8]. In 2017, Google proposed a neural machine translation system based on the attention mechanism Transformer [9], and many subsequent machine translation fields had carried out research on this model, and achieved good results, greatly improving translation quality and translation speed. At present, the application of machine translation technology is becoming more and more extensive, such as Baidu Translate, Google Translate and Youdao Translate, etc. Some researchers have applied machine translation technology to libraries to provide multi-language translation services [10-12], which has brought great convenience to our lives.

Neural machine translation has achieved good results in the field of machine translation, but there are still some problems in neural machine translation, such as sparse data, inadequate translation and unbalanced output caused by single decoding sequence. For the problem of data sparsity, Taheri et al. [13] proposed to use Back-translation. Firstly, the existing parallel corpus was transformed into a source-ended monolingual corpus by using reverse translation technology, and mixed with parallel corpus for training. However, reverse translation required additional learning on the reverse translation pattern, resulting in an increase in computational effort. In addition, due to the large amount of noise in the collected monolingual corpus, the reverse translation of noisy sentences would degrade the quality of pseudo-parallel data, thus affecting the performance of the translation. Frontull et al. [14] proposed that machine translation could achieve automatic learning through dual tasks without the need for human labeling. Susini et al. [15] proposed a method of blending multiple words to perform word substitution by predicting the probability distribution of the next word.

Neural machine translation uses an encoder-decoder architecture and generates target translations from left to right. Despite this remarkable success, there are still some weaknesses [16], the most prominent of which is the problem of unbalanced output. To alleviate these problems, Liu et al. [17] proposed to use an independent bidirectional decoder for neural machine translation, training two left-to-right and right-to-left neural machine translation models separately. The candidate translations were then translated and reordered using the two decoding scores. Zhang et al. [18] proposed an asynchronous bidirectional decoding algorithm for neural machine translation that extended the traditional encoder-decoder framework by using a reverse decoder. However, these methods are more complex than traditional neural machine translation frameworks, requiring two neural machine translation models or decoders [19]. In addition, left-to-right and right-to-left decoders are independent of each other, or only the forward decoder can utilize information from the backward decoder. Therefore, Dong et al. [20] proposed a new algorithm model (SB-NMT), which could realize synchronous bidirectional decoding in a single model, and the experiment proved that the model significantly accelerated the decoding speed while improving the generation quality. Lo et al. [21] improved the monotone sequential generation method and proposed a framework for training text generation models. This method learned strategies for text generation without specifying the generation order. To some extent, it alleviated the problems of unbalanced output and inadequate translation of machine translation, and improved the speed of translation while ensuring the quality of translation.

The application of generative adversarial network in machine translation brings a new development opportunity for machine translation. Its network framework was first proposed by Goodfellow et al in 2014 and had achieved remarkable results in image data [22]. At present, some studies have applied the generative adversarial network architecture and its variants to different scenarios [23], such as sentiment analysis [24], natural language processing and other tasks. Ye et al. [25] proposed a CycleGAN algorithm for cross-domain image-to-image translation to enhance the data model, and the experimental results showed that the use of CycleGAN algorithm could improve the generalization of data in image segmentation. However, when applying generative adversarial networks to machine translation, there is a problem that discrete data cannot be generated. With the maturity of machine translation technology, Nyamathulla et al. [26] applied generative adversarial network to machine translation for the first time, and established a machine translation model of adversarial training by using deep reinforcement technology, making machine translation results similar to manual translation results. Experimental results proved that such adversarial training narrowed the gap between manual translation and machine translation results. Chen et al. [27] carried out further research on conditional generation adversarial network, using sequence generation adversarial network to jointly train generator and discriminator, and using BLEU enhanced generative adversarial network to improve the performance of machine translation. However, the existing generator model is still unable to capture contextual semantic information effectively, which leads to poor translation quality. However, in the process of machine translation, the influence of sentence information and semantic structure on the translation result is equally important. Therefore, how to use model design to solve the problem of semantic information and grammatical structure, so that the model can generate higher quality translation is a problem that needs to be solved at present. Therefore, the generative adversarial network combined with the improved decoder sequence proposed in this study has certain research value and significance to solve the above problems.

2. Proposed Chinese-English Neural Machine Translation Model

This section proposes a neural machine translation model based on generative adversarial network optimization. This model applies the generative adversarial network to neural machine translation, and further optimizes the adversarial learning based neural machine translation model by improving the monotone decoding sequence from left to right or from right to left in the original machine translation model. At the same time, unlike previous generative adversarial networks, neural machine translation models are actually a sequence of discrete symbols that map source language sentences to target language sentences, both in discontinuous Spaces. In this case, the generative adversarial network fails to transmit the gradient properly, causing the generator to lose its update direction. By introducing the strategy gradient algorithm in reinforcement learning, the generator optimization problem in adversarial learning is solved, and the translation performance of the model is improved.

2.1. Problem Analysis

Compared with traditional statistical machine translation (SMT) [28], neural machine translation models achieve similar or even better translation results in an end-to-end framework. At the same time, many different machine translation models have emerged, especially Transformer, which relies on a self-attention mechanism to achieve state-of-the-art performance on different benchmarks. Liu et al. [29] proposed the concept of hierarchical coordination in neural machine translation model, designed hierarchical attention and mixed attention mechanisms, and further shared parameters of each layer between encoder and decoder. Experiments showed that this model was better than Transformer baseline models.

In general, neural machine translation models use left-to-right sequential decoding to generate target translations [30,31]. Despite some success, neural machine translation models still have some weaknesses. One of the most prominent problems is the output imbalance problem, the sequential one-way decoding property, which limits each output word to the previously generated output. However, in the process of translation, the semantic information of the target side cannot be fully considered, which leads to the limitations of the decoder and leads to the problems of inaccurate translation and poor translation.

In this section, adversarial learning based on generative adversarial network is adopted to solve the above problems. Generative machine translation is to improve the decoding sequence of neural machine translation model and take the improved translation model as a generation network to improve the ability of capturing contextual semantic information. In addition, in the framework of generative adversarial network, through the adversarial training of generator and discriminator, the generative ability of generative network is constantly strengthened to promote the generator to generate higher quality translations. At the same time, the strategy gradient method in reinforcement learning is used to improve the discreteness of Chinese text data in natural language processing.

2.2. Overall Model Framework

In view of the above problems faced by machine translation, this study designs a new machine translation model based on generative adversitary network optimization, which is modified on the original Transformer model architecture. The encoder remains consistent with the Transformer encoder. For the decoder side, the original model is decoded from left to right, and the decoding is performed from the middle prediction, regeneration into the word on the right, and finally generate the word on the left, and then generate the entire target sequence.

In this paper, the improved encoder-decoder structure replaces the generator of the original generation adversal network, which improves the unbalanced model output, greatly reduces the decoding delay, and improves the decoder's ability to capture context information more effectively. In addition, the reward mechanism method in reinforcement learning is used to improve the sentence dispersion problem in machine translation, and the fluency of the translation is greatly improved [32,33].

The Chinese-English neural machine translation model based on generative adversarial network optimization is shown in Figure 1. Externally, the overall framework is mainly composed of a generative network and a discriminant network, both of which conduct adversarial training. The generative network is based on the real sample set and aims to improve the similarity between the generated sample and the real sample, while the discriminant network aims to distinguish whether the translation results come from the generated sample or the real sample, and feed the discriminant results back to the generative network to guide the generative network to learn and optimize.

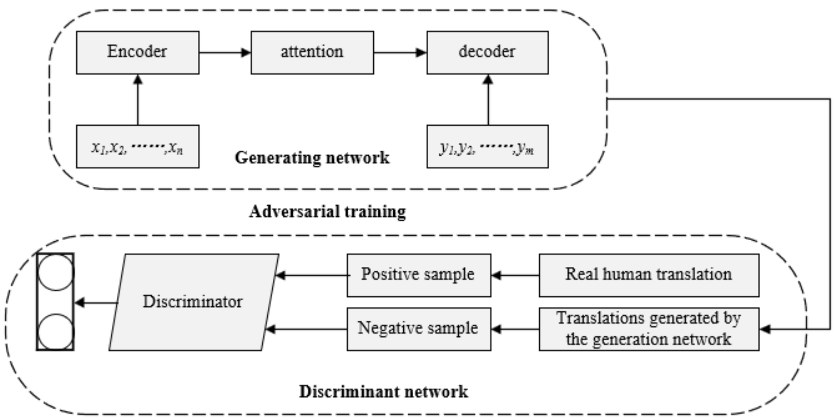

Fig. 1. Proposed model

2.3. Objective function

The model design is based on the Transformer structure, which is divided into two stages. Firstly, given the source statement $X = (x_1, x_2, \cdots, x_n)$, the neural machine translation model $P_\theta(Z_k|X)$ is used to predict the intermediate translation Z_k from the middle of the sentence. Where $Z_k = (y_{m-k+1}, \cdots, y_m, [M], y_1, \cdots, y_{m-k})$. $[M]$ is the $k-th$ word of Z_k. Secondly, the final translation Y is constructed from the intermediate translation Z_k. Given the source statement, it first predicts a word y_{m-k+1}, and then the model predicts the right part $(y_{m-k+1}, \cdots, y_m)$ of the sentence each time. Furthermore, when it predicts the symbol $[M]$, the model begins to predict the left part of the sentence $(y_1, \cdots, y_{m-k})$. Finally, the final translation Y is obtained from the intermediate translation Z_k. In this neural machine translation model, the probability of the decoder generating the target word from the middle prediction is shown in equation (1).

$$P_\theta(Z_k|X) = \prod_{1 \leq j \leq m-k} P_\theta(y_i|X; y_1, \cdots, y_{j-1}, \cdots, y_m). \tag{1}$$

Where, i and j represent the $i-th$ and $j-th$ words of the target end sentence, and $[M]$ is the $k-th$ word of the middle translation Z_k.

2.4. Generator Construction Procedure

Adversarial networks are based on two models, a generative model and a discriminant model. The goal of generator G is to generate target language sentences indistinguishable from human translation on the basis of source sentences. In order to improve translation efficiency, the generation model adopts Transformer structure, which improves the translation speed and improves the performance of machine translation. The Transformer structure is shown in Figure 2.

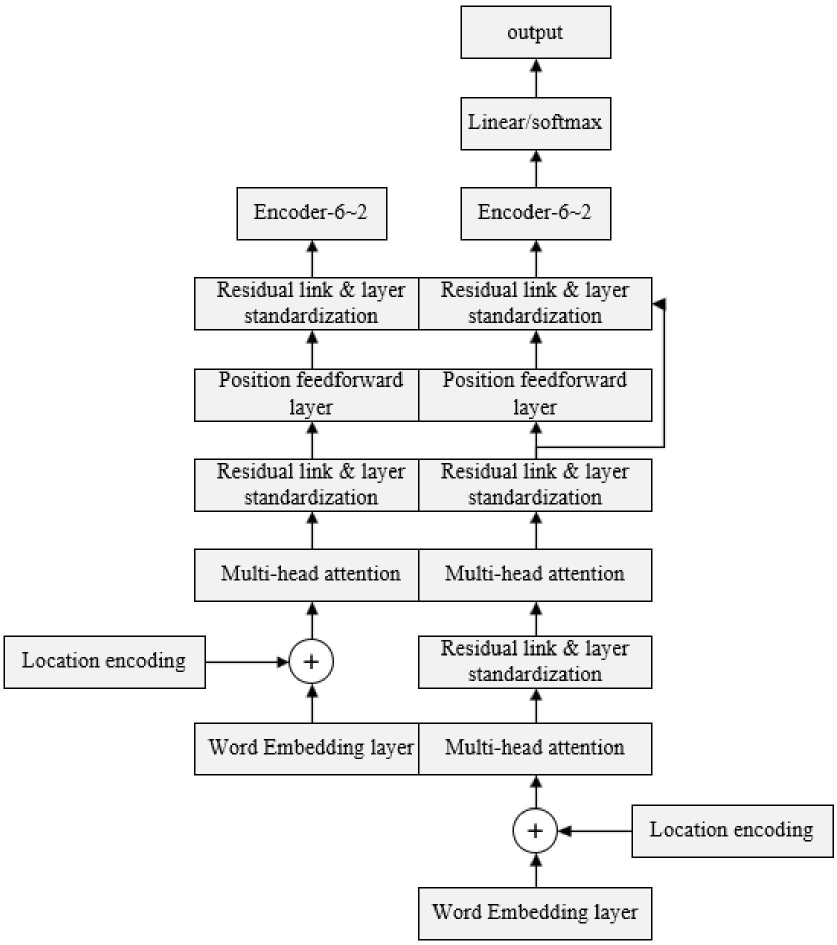

Fig. 2. Transformer model structure

For the encoder, different sentences, the first thing to do is to input the word vector into the model, which is the embedding layer to obtain the word vector of the source language and target language. Since the network structure with sequential input is not adopted, the word position should be marked before the word vector enters the model. The position marking formula is as follows.

$$PE_{(pos,2i)} = \sin(pos/1000^{2i/d_{model}}). \tag{2}$$

$$PE_{(pos,2i+1)} = \cos(pos/1000^{2i/d_{model}}). \tag{3}$$

Where d_{model} is the length of the word vector.

The word vector with position information is obtained by adding these two formulas to the word vector. Next, a word vector with location information is entered into self-attention to get a self-attention score for each word. Multi-head attention is at the heart of the Transformer model and its operation is shown in Figure 3.

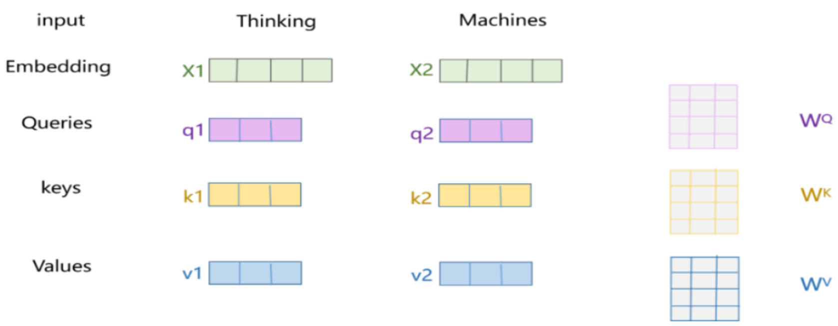

Fig. 3. self-attention computation

Where x_1 and x_2 represent the input word vector. W^Q, W^K and W^V represent the three weight matrices respectively, and $W \in R^{d_{model} \times v}$ (v represents the length of the vocabulary).

It multiplies x_i by the three weight matrices respectively to get the query vector q_i, the key vector k_i, and the value vector v_i. The calculation formula is as follows.

$$q_i = x_i \cdot W^Q. \tag{4}$$

$$k_i = x_i \cdot W^K. \tag{5}$$

$$v_i = x_i \cdot W^V. \tag{6}$$

When Q, K, and V represent the same sequence, the attention mechanism is calculated as equation (7).

$$attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}}) \cdot V. \tag{7}$$

Where d_k represents the dimension of the key vector K, that is, the dimension of the model. $\sqrt{d}$ is a scale factor. This is used to standardize the value of the input softmax function.

The Transformer model completely abandons the architecture based on circular network and convolutional network, and the self-attention mechanism can better capture global information. Meanwhile, the parallel processing greatly improves the model training speed. However, as the output of Transformer adopts the weighted average method, attention will be distracted and the relationship between adjacent words will be ignored. In this paper, the decoder sequence of Transformer is improved to enhance the ability of model capture dependency. The improved Transformer model is used as a new generative model in generative adversarial network, and is trained together with discriminant network to improve the robustness of neural machine translation model.

2.5. Discriminator Construction Process

The setting of discriminator is consistent with that of reference [34], which is realized based on convolutional neural network CNN. Discriminator D, conditioned on source language sentences, attempts to distinguish between machine-generated sentences and human-translated sentences. Because D is updated synchronously with G, D can be considered a dynamic target. Since the length of the sequence generated by the generation model is different, the output sequence is first extended to a fixed length T, and the source matrix X and the target matrix X are constructed according to the source language sequence and the target sequence respectively. The formula is as follows.

$$X_{1:T} = x_1; x_2; \cdots; x_T. \tag{8}$$

$$Y_{1:T} = y_1; y_2; \cdots; y_T. \tag{9}$$

Where $x_t, y_t \in R^k$ are word embedding vectors of dimension k. The source matrix $X_{1:T}, W_j \in R^{l \times k}$ is a convolution kernel of length L. The convolution operation on the convolution kernel of length L produces a series of feature mappings, whose convolution formula is shown in (10).

$$C_{ji} = f(W_j \otimes X_{i:i+l-1} + b). \tag{10}$$

Where W_j is the weight of the convolution kernel. $X_{i:i+l-1}$ is the word vector matrix from i to $i+l-1$ in the window. $\otimes$ is the convolution between elements. b is the offset term. f is a nonlinear activation function.

Features are extracted by using different number of kernel functions and different window sizes, and then the maximum value of each feature vector is spliced together to form the feature vector C_x of the source sequence. The feature map of each convolution kernel is max pooled to get the final feature result.

$$C_x = max c_{j1}, c_{j2}, \cdots, c_{j(T-l+1)}. \tag{11}$$

By using the above method, the target sequence feature vector is extracted from the source matrix $Y_{1:T}$, and the probability that the target statement is the true sample is calculated by C_x and C_y, and the calculation formula is shown in (12).

$$P = \Phi(V[C_x; C_y]). \tag{12}$$

Where V is the transformation matrix that connects C_x and C_y into a two-dimensional word vector. Φ is the softmax function.

3. Experiment and Result Analysis

3.1. Experimental Environment and Data

The experimental environment of this research is as follows: The CPU used in the hardware is Intel Core i7-9700 and the main memory is 16G. The operating system used for the software is Windows 10 Professional Edition, using Python 3.5 programming, using the Python language under the Pytorch framework.

For English-German translation data, the WMT14 translation corpus is selected in this experiment, and the data set includes training set, test set and verification set. 400000 English-German sentence pairs are randomly selected from WMT14En-De as a training set. newstest2013 and newstest2014 are selected as the test set, and newstest2015 is selected as the verification set. For monolingual data, 50000 English monolingual data and 50000 German monolingual data are extracted from WMT [35,36].

For Chinese-English translation data, this experiment obtains them from the LDC corpus. The training data consistes of 10000 sentence pairs. The NIST 2002, 2003, 2006 data set are selected as the test set, and the NIST 2005 data set is selected as the validation set. For monolingual data, 50000 Chinese monolingual data and 50000 English monolingual data are extracted from Gigaword corpus.

3.2. Comparison Experiments

The data enhancement method proposed in this chapter is dynamic data augmentation combined with sampling decoding strategy by adding noise to the target language. The experiment is conducted from two aspects to verify the effect:

First, the comparison experiments between different network structures are mainly from the following two network structures.

(1) RNNSearch: The neural machine translation model architecture proposed by Bahdanau et al. [37] applied the attention mechanism in the field of natural language processing for the first time. By combining the attention mechanism with Bi-LSTM, the performance of machine translation was improved, and the translation effect was comparable to that of the phrase-based translation system.

(2) Transformer: The model architecture proposed by Vaswani et al. [38], which was completely based on the self-attention mechanism to carry out machine translation tasks, had achieved good results and become the mainstream network structure of our recent research.

Second, in order to solve the problem of sparse data in neural machine translation, this section proposes a data enhancement method based on dynamic data expansion and sampling decoding strategy. Therefore, this study is compared with the work related to data enhancement and different decoding strategies as follows:

(1) base: Only the original parallel corpus is used, without any data enhancement methods.

(2) Data Enhancement (DA)+beam: data enhancement is used, and beam search algorithm is used to generate translation in the decoding stage.

(3) Proposed: Data enhancement (DA) is used to generate the final translation by sampling algorithm in the decoding stage.

This section conducts experiments on WMT14 English-German and LDC corpus Chinese-English data sets respectively. The baseline model is based on RNNSearch and Transformer, and the comparison methods are base (without data enhancement) and Edunov (reverse translation method +beam), proposed method (The target end dynamically adds noise +sampling decoding strategy). The experimental results of each model method on different data sets are shown in Tables 1 and 2.

Table 1. BLEU value of English-German translation experiment/%

Network	Model	Tst2013	Tst2014	Average
RNNSearch	base	20.83	23.01	21.92
RNNSearch	DA+beam	22.70	23.24	22.97
RNNSearch	Proposed	23.25	24.43	23.84
Transformer	base	26.63	27.33	26.98
Transformer	DA+beam	26.99	27.95	27.52
Transformer	Proposed	27.55	28.47	27.91

On the English-German translation dataset, the BLEU index of 23.84 for RNNSearch model and 27.91 for Transformer model is obtained by using the proposed method, which is better than any other data enhancement method (base). Based on the use of data enhancement, it can be found that the sampling decoding strategy algorithm has an average increase of 0.87% and 0.49% BLEU respectively on different backbone networks than the data enhancement method using the beam search algorithm.

The proposed method in this study compares the two groups of experiments on the Chinese-English translation dataset. The experimental results given in Table 2 show that: compared with the baseline system, the bilingual translation evaluation BLEU value brought by the data enhancement method combined with the beam search algorithm has improved by 0.36%. Based on the data enhancement method by adding noise to the target end, the BLEU value using sampling decoding strategy is improved by 0.71% on average compared with the BLEU value searched by beam. This verifies the validity of the noisy data expansion combined with sampling decoding strategy. In addition, it also shows that the generation model with sampling decoding strategy can increase the diversity of target sentences and generate higher quality translations.

According to the experimental results in the above table, the following figure is further analyzed. As shown in figure 4 and figure 5, comparing the effects of the two network architectures RNNSearch and Transformer model on the test set, it can be seen that the Transformer model based on the self-attention mechanism has better performance in machine translation tasks than the RNNSearch model.

As shown in Figure 5, Transformer model is superior to RNNSearch model in terms of overall translation performance. In addition, sampling decoding strategy is better than beam search when data expansion is combined with different decoding strategies to generate target sentences.

Table 2. BLEU value of English-Chinese translation experiment/%

Network	Model	NIST02	NIST03	NIST06	Average
RNNSearch	base	39.79	38.26	37.96	38.67
RNNSearch	DA+beam	39.97	38.55	38.12	38.88
RNNSearch	Proposed	40.66	39.44	39.30	39.80
Transformer	base	46.85	45.13	44.76	45.58
Transformer	DA+beam	46.98	45.64	45.20	45.94
Transformer	Proposed	47.70	46.36	45.89	46.65

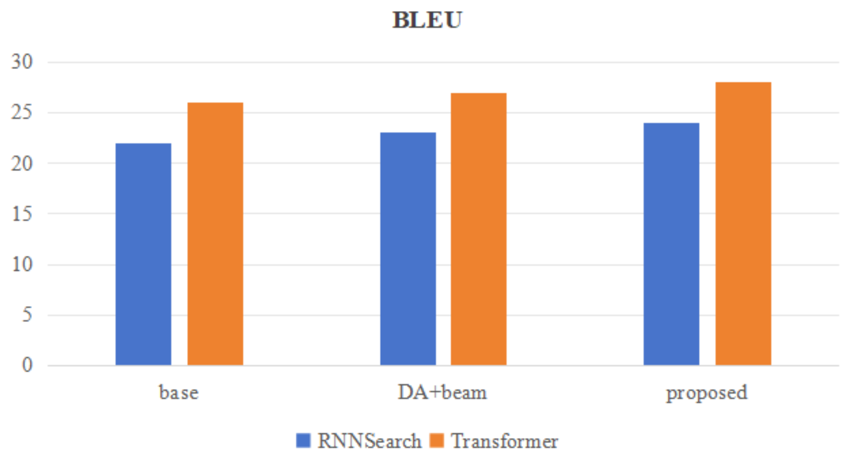

Fig. 4. Results of English-German translation

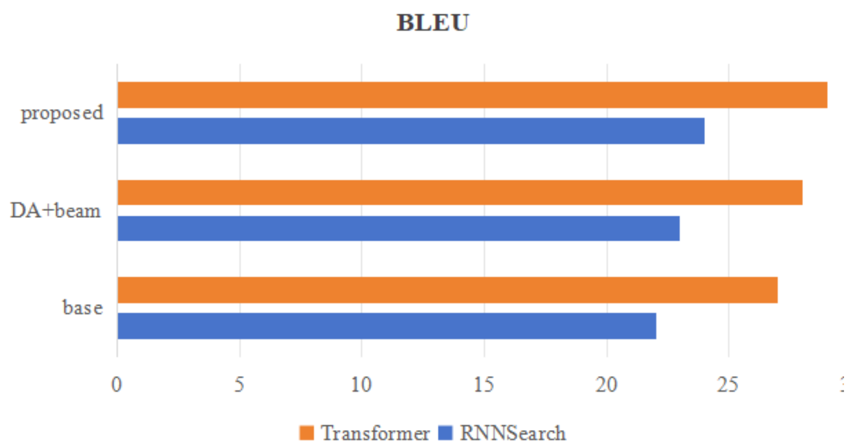

Fig. 5. Results of English-Chinese translation

4. Conclusion

The main content of this paper is to solve the problem of sparse data in neural machine translation, using the method of data expansion, adding noise to the target sentence according to a specific strategy, and randomly replacing and overwriting the target sentence. In this way, the noisy target statement is obtained, and the pseudo-parallel corpus is formed with the source statement. The noisy target statement is restored to the original statement by the constraint encoder, and the generalization ability and anti-interference ability of the model are enhanced. At the same time, in order to improve the diversity of the generated final translation, sampling decoding strategy is adopted during decoding, so that more and more authentic translations can be generated at the sentence level. The experimental results show that the method of data expansion combined with sampling decoding can effectively improve the robustness of neural machine translation model and improve the performance of machine translation on the two standard data sets of WMT14 English and German and LDC Chinese and English.

5. Conflict of Interest

The authors declare that there are no conflict of interests, we do not have any possible conflicts of interest.

Acknowledgments. None.

References

1. Rajan, Remya, Remya Sivan, Remya Ravindran, and K. P. Soman. Rule based machine translation from English to Malayalam. In 2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies, pp. 439-441. IEEE, 2009.
2. Lagarda, Antonio-L., Vicent Alabau, Francisco Casacuberta, Roberto Silva, and Enrique Diaz-de-Liano. ”Statistical post-editing of a rule-based machine translation system.” In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers, pp. 217-220. 2009.
3. Yin S, Li H, Laghari A A, et al. FLSN-MVO: Edge Computing and Privacy Protection Based on Federated Learning Siamese Network With Multi-Verse Optimization Algorithm for Industry 5.0[J]. IEEE Open Journal of the Communications Society, 2024.
4. Wang H, Wu H, He Z, et al. Progress in machine translation[J]. Engineering, 2022, 18: 143-153.
5. Stahlberg F. Neural machine translation: A review[J]. Journal of Artificial Intelligence Research, 2020, 69: 343-418.
6. Jiang L, Yu R, Zhao Y. Theoretical perspectives and factors influencing machine translation use in L2 writing: A scoping review[J]. Journal of Second Language Writing, 2024, 64: 101099.
7. Yu J, Lu Z, Yin S, et al. News recommendation model based on encoder graph neural network and bat optimization in online social multimedia art education[J]. Computer Science and Information Systems, vol. 21, no. 3, pp. 989-1012, 2024. doi: 10.2298/CSIS231225025Y.
8. Teng L, Wang J. A Novel Gesture Recognition Network based on LSTM[J]. Journal of Science and Engineering, 2024, 1(1): 1-6.
9. Ibrar M, Sun Y. SEIR Model Based Epidemic Transmission Risk Deep Prediction[J]. Journal of Science and Engineering, 2024, 1(1): 25-31.
10. Zhang W, Wang J. English Text Sentiment Analysis Network based on CNN and U-Net[J]. Journal of Science and Engineering, 2024, 1(1): 13-18
11. Yin S, Li H, Laghari A A, et al. An anomaly detection model based on deep auto-encoder and capsule graph convolution via sparrow search algorithm in 6G internet-of-everything[J]. IEEE Internet of Things Journal, vol. 11, no. 18, pp. 29402-29411, 2024. DOI: 10.1109/JIOT.2024.3353337.
12. Wang Z, Wang Y. Digital Library Book Recommendation System Based on Tag Mining[J]. Journal of Artificial Intelligence Research, 2024, 1(1): 10-16.
13. Taheri A, Zamanifar A, Farhadi A. Enhancing aspect-based sentiment analysis using data augmentation based on back-translation[J]. International Journal of Data Science and Analytics, 2024: 1-26.
14. Frontull S, Moser G. Rule-based, neural and LLM back-translation: Comparative insights from a variant of Ladin[J]. arXiv preprint arXiv:2407.08819, 2024.
15. Susini M, Ana I W. Back-Translation as a Tool for Identifying Language Change[J]. Journal of Language Teaching and Research, 2024, 15(4): 1178-1188.
16. Shah T Z, Imran M, Ismail S M. A diachronic study determining syntactic and semantic features of Urdu-English neural machine translation[J]. Heliyon, 2024, 10(1).
17. Liu X. QDLTrans: Enhancing English Neural Machine Translation with Quantized Attention Block and Tunable Dual Learning[J]. IEEE Access, 2025.
18. Zhang X, Zhou Y, Yang G, et al. Context-aware code generation with synchronous bidirectional decoder[J]. Journal of Systems and Software, 2024, 214: 112066.

19. Meng X, Wang X, Yin S, et al. Few-shot image classification algorithm based on attention mechanism and weight fusion[J]. Journal of Engineering and Applied Science, 2023, 70(1): 14.
20. Dong J. Syntax-aware bidirectional decoding Neural Machine Translation model[J]. Applied and Computational Engineering, 2024, 37: 8-15.
21. Lo P C, Lim E P. Non-monotonic generation of knowledge paths for context understanding[J]. ACM Transactions on Management Information Systems, 2024, 15(1): 1-28.
22. Ghosheh G O, Li J, Zhu T. A survey of generative adversarial networks for synthesizing structured electronic health records[J]. ACM Computing Surveys, 2024, 56(6): 1-34.
23. Li P, Laghari A A, Rashid M, et al. A deep multimodal adversarial cycle-consistent network for smart enterprise system[J]. IEEE Transactions on Industrial Informatics, 2022, 19(1): 693-702.
24. Mao Y, Liu Q, Zhang Y. Sentiment analysis methods, applications, and challenges: A systematic literature review[J]. Journal of King Saud University-Computer and Information Sciences, 2024: 102048.
25. Ye M, Meng Z, Qian Y. Building Cross-Domain Mapping Chains from Multi-CycleGAN for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024.
26. Nyamathulla S, Veeranjaneyulu N. Analysis of Pix2Pix and CycleGAN for Image-to-Image Translation: A Comparative Study[C]//2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE). IEEE, 2024: 1-6.
27. Chen Z, Bian Y, Shen E, et al. Moment-Consistent Contrastive CycleGAN for Cross-Domain Pancreatic Image Segmentation[J]. IEEE Transactions on Medical Imaging, 2024.
28. Das S B, Panda D, Mishra T K, et al. Statistical machine translation for indic languages[J]. Natural Language Processing, 2025, 31(2): 328-345.
29. Liu X, Zhang W, Shao C, et al. Autonomous collaborative optimization control of earth pressure balance shield machine based on hierarchical control architecture[J]. Engineering Applications of Artificial Intelligence, 2024, 137: 109200.
30. Yin, S., Ivanović, M. Real-Time Image processing with deep neural networks and optimization algorithms. Computer Science and Information Systems, vol. 19, no. 3, iiiCvi. (2022), https://doi.org/10.2298/CSIS220300iiiY.
31. Zhao Y, Li H, Yin S. A Multi-channel Character Relationship Classification Model Based on Attention Mechanism[J]. Int. J. Math. Sci. Comput.(IJMSC), 2022, 8: 28-36.
32. Gao Z, Wang Q, Mei T, et al. An enhanced encoder-decoder network architecture for reducing information loss in image semantic segmentation[C]//2024 5th International Conference on Information Science, Parallel and Distributed Systems (ISPDS). IEEE, 2024: 116-120.
33. Chaoyang Z, Shibao S, Wenmao H, et al. FDR-TransUNet: a novel encoder-decoder architecture with vision transformer for improved medical image segmentation[J]. Computers in Biology and Medicine, 2024, 169: 107858.
34. Yu J, Zhao L. A novel deep CNN method based on aesthetic rule for user preferential images recommendation[J]. Journal of Applied Science and Engineering, 2021, 24(1): 49-55.
35. Wu Y, Xing Y. Efficient Machine Translation with a BiLSTM-Attention Approach[J]. arXiv preprint arXiv:2410.22335, 2024.
36. Li F, Liu B, Yan H, et al. Incorporating bilingual translation templates into neural machine translation[J]. Scientific Reports, 2025, 15(1): 5547.

Biography

Tianxiao Wang is with the College of International Business, Shenyang Normal University. Research direction is English education, computer application and AI.