

Research Article

Two-Way Neural Network Chinese-English Machine Translation Model Fused with Attention Mechanism

Jing Liang¹ and Minghui Du ²

¹*Yibin Vocational and Technical College, Yibin, Sichuan 644000, China*

²*Nanjing University of Chinese Medicine, School of Foreign Language Education, Nanjing, Jiangsu 210023, China*

Correspondence should be addressed to Minghui Du; 390052@njucm.edu.cn

Received 30 November 2021; Revised 24 December 2021; Accepted 27 December 2021; Published 25 January 2022

Academic Editor: Baiyuan Ding

Copyright © 2022 Jing Liang and Minghui Du. 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.

This study uses an end-to-end encoder-decoder structure to build a machine translation model, allowing the machine to automatically learn features and transform the corpus data into distributed representations. The word vector uses a neural network to achieve direct mapping. Research on constructing neural machine translation models for different neural network structures. Based on the translation model of the LSTM network, the gate valve mechanism reduces the gradient attenuation and improves the ability to process long-distance sequences. Based on the GRU network structure, the simplified processing is performed on the basis of it, which reduces the training complexity and achieves good performance. Aiming at the problem that some source language sequences of any length in the encoder are encoded into fixed-dimensional background vectors, the attention mechanism is introduced to dynamically adjust the degree of influence of the source language context on the target language sequence, and the translation model's ability to deal with long-distance dependencies is improved. In order to better reflect the context information, this study further proposes a machine translation model based on two-way GRU and compares and analyzes multiple translation models to verify the effectiveness of the model's performance improvement.

1. Introduction

Machine translation usually combines natural language processing and artificial intelligence. It retains the original meaning. Realize the mutual translation between two natural languages and complete the equivalent conversion. As an important method to realize the equivalent transmission, translation is rather important in this process [1–6].

In the previous research, linguists manually wrote the conversion rules between the two languages to be translated. Although the research on rule-based methods has reached the syntactic stage, this method is almost entirely dependent on the quality of language rules, which requires linguists to be very high, and is easily restricted in the field of application and cannot summarize. all the rules that the language will use. The emergence of a translation model based on word alignment proposed by IBM in 1993 marked the birth of statistical machine translation. Compared with the previous model, statistical machine translation learns the conversion

rules from the corpus, no longer needs to provide the language rules actively, and solves the bottleneck problem of knowledge acquisition. But in order to obtain perfect translation results, this method still has many problems. The translation results have poor utilization of global features and rely too much on data preprocessing steps. Each step is related to each other, and the accuracy of the processing steps affects the final translation result, greater impact. In fact, its related research has achieved good results. It is usually applied to provide new solutions for machine translation-related problems. The machine translation can be roughly divided into two types. One still uses the statistical machine translation system framework to improve key modules [7–11].

The research and exploration may be dated back to the birth of electronic computers in 1940s. Known as the pioneer of machine translation, Warren Weaver, as the head of the Natural Science Department of the Rockefeller Foundation, published “Machine Translation” in 1949, marking the

formal proposal of the idea of machine translation. He believes that the existence of multiple languages in the world is inherently consistent, and languages exist as a description tool of human society and objective things. For people who master a certain language, a foreign language is just a form of encoding of their mother tongue, and the foreign language can be translated into their mother tongue by cracking the password. Under the social background of significant achievements in the field of password cracking, people generally accepted Weaver's point of view and full of expectations for machine translation research. In the following ten years, the popularity of machine translation has continued to rise. The United States and the Soviet Union have caused a sudden upsurge in machine translation [12–15].

The development of machine translation has not been smooth sailing. For a period of time, the development of machine translation was put on hold. This was attributed to the “Language and Machine” report issued by the American Academy of Sciences Automated Language Processing Advisory Committee in 1966. In this report, the researchers of the committee denied the feasibility of machine translation with the shortcomings of slow speed, high accuracy, and high consumption and directly denied the possibility of the development of practical machine translation systems. At that time, the field of machine translation research was affected by the report, and the research investment was terminated one after another, making machine translation research into a period of underestimation. With the development of computer application technology and linguistics-related fields, coupled with the expansion of the demand for social information services at that time, machine translation has only begun to recover. The development of machine translation has mainly experienced the following aspects [16, 17].

The translation process is mainly composed of three parts: analysis, conversion, and generation. By parsing the source language sentence, the deep structure is obtained. The deep structure representation mentioned here can be divided into syntax-based, semantic-based, or intermediate language-based methods. Syntactic research focuses on analyzing the syntactic structure of the source language and generating the syntactic structure of the target language; semantic research uses semantic analysis on the source language to obtain the semantic representation and then transforms the semantic representation to generate the target language; the intermediate language method uses a mapping relationship with natural language. The universal intermediate language realizes through two mappings. Among them, machine translation based on syntactic rules has relatively mature syntactic theories such as generative transformation grammar and dependent grammar, which can obtain better translation results under certain conditions. The methods of semantic analysis and intermediate language are still in the preliminary research stage [18–22].

The rule-based machine translation method highly relies on the rules of language. Although it has a certain degree of versatility, the cost of obtaining the rules is relatively high. The quality of the language rules is too dependent on the experience and knowledge of the linguist, the maintenance

of the rules, and the compatibility of the new and the old rules. It is a bottleneck problem that is difficult to break through. After the rule-based method, researchers built a large-scale bilingual corpus. At the same time, machine learning research methods appeared. The combination of the two has led to improved machine translation methods. This leads to the emergence of instance-based and statistical-based machines. Translation methods are based on bilingual corpus for research.

After years of research, statistical machine translation has achieved good results, but there are still many problems to be solved. Because the linear model is used, linearity is inseparable when processing high-dimensional complex data, making it difficult for training and search algorithms to approach the theoretical upper limit of the translation space; the conversion is mainly at the vocabulary, phrase and syntax level realization, and lack of proper semantic representation; use context-independent features to design funny dynamic programming search algorithms, which makes it difficult to accommodate nonlocal context information in the model; in addition, it also includes the difficulty of feature design, data sparseness, and error propagation caused by pipeline architecture problem. The application of deep learning theory can better solve these problems of statistical machine translation. Existing research methods mainly include two kinds: one uses deep learning technology to improve key modules and built the model to achieve direct source language to target language [23–27].

The well-known professor Bengio in the field of deep learning proposed a language model based on the neural network in 2003, which effectively alleviated the problem of data sparseness through distributed representation. Compared with the traditional language model that only considers the first $n-1$ words of the target language, Jacob et al. believed that not only the historical information of the target language but also the relevant parts of the source language also play a certain role. They proposed a neural network joint model to increase the BLEU value. About 6%, Peng et al. used a recursive autoencoder to obtain the distributed representation of the word string and then built a neural network classifier to alleviate the problem of feature design. Lu et al. transformed the existing feature set into a new feature set, which significantly improved the expression ability of the translation model. Nal et al. first proposed end-to-end neural machine translation in 2013 and proposed a “encoding-decoding” model framework to directly model the translation probability. For the input source language sentence, the encoder was used to map it into a continuous, dense vector, the continuous representation method solves the previous sparse problem, and then the vector is converted into a target language sentence through the decoder. They use convolutional neural networks to construct encoders, and decoders use recurrent neural networks to obtain historical information and process variable-length strings. In view of the gradient attenuation and gradient explosion problems that are prone to occur in the training process, Sutskever et al. introduced long and short-term memory LSTM into end-to-end neural machine translation and used the method of setting threshold switches to

improve the recurrent neural network, aiming at the problem of the encoder generating fixed-length vectors [28–33].

2. Neural Networks

The use of activation functions adds nonlinear modeling capabilities to neural networks. It can only perform linear mapping expressions. The entire network is only equivalent to a single-layer neural network, losing the actual meaning of the hidden layer. The activation function can make the deep neural network have the ability of nonlinear mapping learning. Under normal circumstances, the activation function should have the properties of differentiability, monotonicity, and limited output value range. Because the neural network training process is optimized based on the gradient, the activation function must be differentiable to ensure the calculation of the gradient. The monotonic activation function can ensure that the single-layer neural network is convex. Since the optimization is carried out on the basis of gradient calculation, the finite output of the activation function can make the feature representation in the neural network more significantly affected by the finite weight, thereby making the optimization method more stable. At present, the common activation functions are mostly piecewise linear and nonlinear functions with exponential shape, mainly including sigmoid, tanh, ReLU, ELU, and PReLU. Here, we will introduce the two activation functions used in this article.

The sigmoid function is the most widely used activation function, and its functional formula is

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (1)$$

Tanh is a hyperbolic tangent function, and the function form is

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}. \quad (2)$$

As another commonly used activation function, the output range of the tanh function is between $[-1, 1]$, and the average output value is 0, which is the same as the sigmoid function. When the input is large or small, the output of the function is almost smooth. Although there will also be the problem of gradient disappearance, the tanh function converges faster and can reduce the number of iterations.

For the linear regression model, the hypothetical function is expressed as follows:

$$h_{\theta}(x_1, x_2, \dots, x_n) = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n. \quad (3)$$

where θ_i ($i = 0, 1, \dots, n$) is the model parameter and x_i ($i = 0, 1, \dots, n$) is the n eigenvalue of each sample. Adding $x_0 = 1$, the above formula can be transformed into

$$h_{\theta}(x_1, x_2, \dots, x_n) = \sum_{i=1}^n \theta_i x_i. \quad (4)$$

The corresponding cost function is

$$J_{\theta} = \frac{1}{2m} \sum_{j=0}^m [h_{\theta}(x_0^j, x_1^j, \dots, x_n^j) - y_j]^2. \quad (5)$$

The algorithm flow corresponding to the gradient descent method is shown in Figure 1:

In the system, the column search algorithm search tree is constructed. This method can greatly reduce the time and space resources occupied by the search process, but since each step is implemented by the greedy method, it cannot guarantee that the global optimal solution is obtained.

Then, it becomes

$$P(y_1, y_2, \dots, y_{T'}) = \prod_{t'=1}^{T'} P(y_{t'} | y_1, \dots, y_{t'-1}, c). \quad (6)$$

The dropout is shown as Figure 2.

Regarding the above systems, translation results cannot be directly reflected in the form of output text, and certain quantitative standards are needed for evaluation. In recent years, a variety of evaluation standards have appeared in the world, including BLEU, NIST, and METEOR. The most commonly used is the BLEU (Bilingual Evaluation Understudy). This method obtains the evaluation value by calculating the similarity between the translation result of the computer translation system and the manual translation result. Compare and count the number of n -grams that cooccur in the system translation and the reference translation, and then divide the number of matched n -grams. The mathematical formula is

$$\text{BLEU} = \text{BP} \cdot \exp\left(\sum_{n=1}^N \omega_n \log p_n\right). \quad (7)$$

Among them, BP is the penalty factor, and N is the longest n -gram, ω_n means that the weight of the cooccurring n -gram word usually takes the constant value $1/n$, and the value of N is usually 4, and $\omega_n = 0.25$ is the most commonly used BLEU4 indicator.

3. Neural Network Translation Model Incorporating Attention

The rule-based machine translation method highly relies on the rules of language. Although it has a certain degree of versatility, the cost of obtaining the rules is relatively high. The quality of the language rules is too dependent on the experience and knowledge of the linguist, the maintenance of the rules, and the compatibility of the new and old rules. It is a bottleneck problem that is difficult to break through. After the rule-based method, researchers built a large-scale bilingual corpus. At the same time, machine learning research methods appeared. The combination of the two has led to improved machine translation methods. This leads to the emergence of instance-based and statistical-based machines. Translation methods are based on bilingual corpus for research.

After years of research, statistical machine translation has achieved good results, but there are still many problems

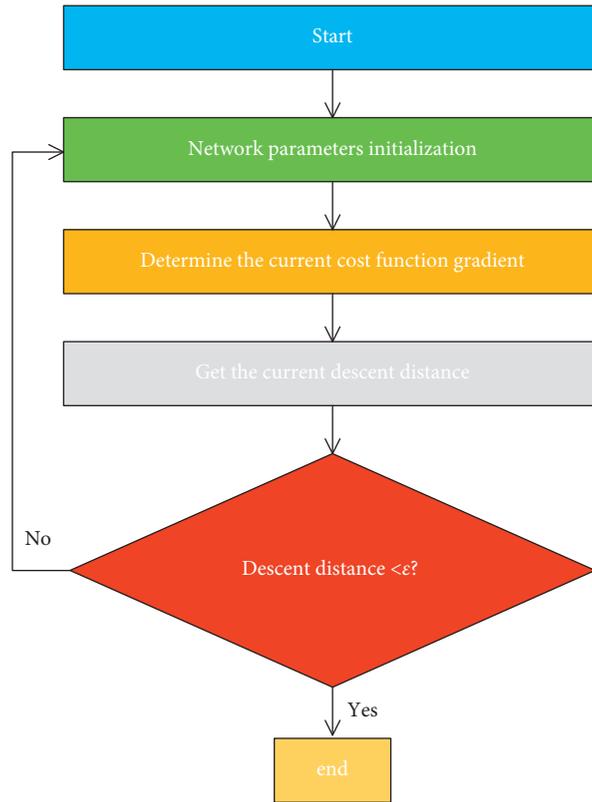


FIGURE 1: Gradient descent algorithm flow.

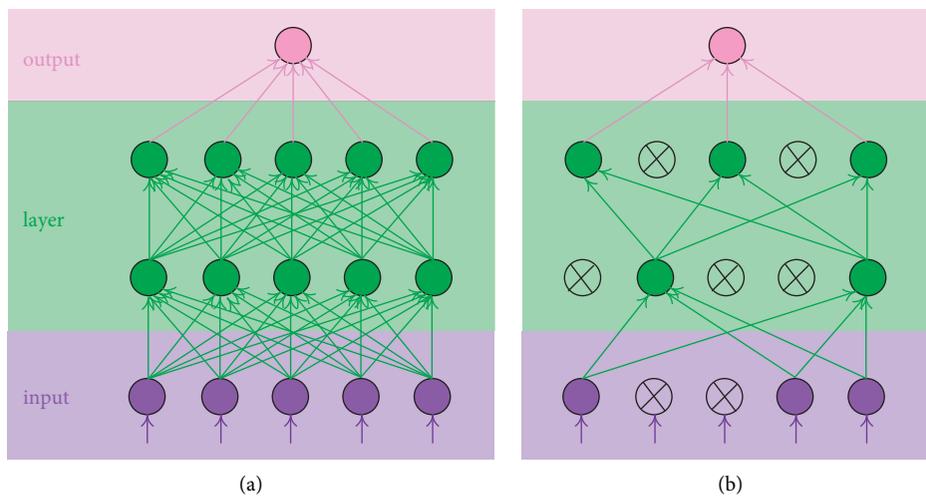


FIGURE 2: Dropout figure. (a) Traditional neural network. (b) Dropout neural network.

to be solved. Because the linear model is used, linearity is inseparable when processing high-dimensional complex data, making it difficult for training and search algorithms to approach the theoretical upper limit of the translation space; the conversion is mainly at the vocabulary, phrase and syntax level realization, and lack of proper semantic representation; use context-independent features to design funny dynamic programming search algorithms, which makes it difficult to accommodate nonlocal context information in the model; in addition, it also includes the

difficulty of feature design, data sparseness, and error propagation caused by pipeline architecture problem. The application of deep learning theory can better solve these problems of statistical machine translation. Existing research methods mainly include two kinds: one uses deep learning technology to improve key modules and built the model to achieve direct source language to target language.

At present, many researchers have proposed many improved methods for NMT, among which the most effective is the codec system based on the attention

architecture. Figure 3(a) shows the structural principle according to the attention NMT model, which is divided into three parts: the coding and decoding layers and the structure of the middle cascade that introduces the attention mechanism.

The NMT system first converts all the divided symbols into a sequence, namely, word embedding. In this process, each character must be processed separately, and finally, the original sequence after the word embedding is generated. In the figure, the word embedding layer, NMT uses a bidirectional recurrent neural network (biRNN) to obtain a representation of the entire original sequence after training. Between the encoding layer and the decoding layer, an attention mechanism is added to fuse all time steps of the input sequence and focus on the current time step of the decoding layer. In the process of generating the target word, the controller will integrate three items: the last generated word, the current hidden layer state, and the context information calculated by the attention mechanism to determine the final target word.

The RNN coding layer is very important for NMT based on the attention model. However, it is difficult for traditional RNNs to achieve multilayer information integration, and machine translation increasingly requires this network structure. Therefore, this article proposes a multichannel attention mechanism encoder, the network of which is shown in Figure 3(b). This structure adds an external storage to assist the RNN to complete more complex integrated learning. In addition, the hidden layer state of the RNN and the word embedding sequence together generate gated annotations for the attention mechanism between the codec layers. From another perspective, integrating the word embedding sequence into the attention mechanism model can also be seen as establishing a short-circuit connection, which can alleviate the degradation problem. This short-circuit connection does not introduce any additional parameters while enhancing the network function and does not cause computationally complex upgrades.

Figure 4 shows the detailed rules for reading and writing the memory of the coding layer of the neural translation system. In each time step, the state node in the RNN queries the context information in the memory, and the memory is stored according to the attention-based mechanism. In this design, the previous state node is used to query and obtain context information as the input state of the gated recurrent unit (GRU), instead of directly feeding back the previous state to the GRU. This operation ensures that the controller can obtain more context information before generating the current state, which can potentially help the GRU make judgments. While designing read memory operation, write operation is also designed in the system. The purpose of this design, according to the research work of the Baidu team, is to hope that the RNN and NTM can learn different types of associations through different update strategies.

The function of the encoder is to convert an input sequence of variable length into a background vector C of fixed length.

The encoder part is mainly divided into three steps:

- (1) First, the input source language sequence $X = \{x_1, x_2, \dots, x_n\}$ is represented by one-hot encoding, and each word x_i in the source language sentence is represented as a column vector w_i . The value of the dimension represented by the index value is 1, and the rest are 0.
- (2) Map the vector obtained in the previous step to the low-dimensional semantic space to form a word vector. Since the vector obtained by only one-hot encoding is too sparse and cannot describe the semantic similarity, it is necessary to use a word vector model of distributed representation to convert it and map it to the low-dimensional semantic space to generate a fixed-dimensional word vector.
- (3) Using neural network coding to generate source language, as shown in the following equation:

$$h = \sigma(h_{i-1}, S_i). \quad (8)$$

Among them, h_0 is an all-zero vector, σ represents a nonlinear activation function, and the equation $h = \{h_1, h_2, \dots, h_n\}$ obtained is the input state coding sequence of n source language words, which is the background vector C mentioned at the beginning of this section.

The task of the decoder is to obtain the maximum probability. The specific process is as follows:

- (1) For a certain time i in the sequence, calculate the next hidden layer state Z_{i+1} according to the background vector C generated by the source language sequence. Calculate by

$$Z_{i+1} = \sigma(c, u_i, z_i). \quad (9)$$

Among them, z_0 is an all-zero vector, and s represents a nonlinear activation function.

- (2) The softmax function is used for normalization to obtain the probability distribution of the Z_{i+1}^{th} word
- (3) Calculate the cost function according to the obtained probability, and repeat the above sequence processed.

Statistical machine translation usually designs translation models based on linguistic knowledge. Neural machine translation focuses on the design of neural network mechanisms in the model. The encoder and decoder frameworks involve the use of neural networks to build models. In actual use, there are many types of neural networks, and choosing a suitable neural network has an important impact on the machine translation results. As the earliest network structure proposed in the field of neural machine translation, the recurrent neural network is a neural network that can store time series information. It stores historical information through hidden states. This structure makes it theoretically able to input sequences of arbitrary length processed. It evolved on the basis of the feedforward neural network. Each layer of the recurrent neural network not only outputs the next layer but also outputs a hidden layer state, which is used when processing the next sample for the current layer. The state of the hidden layer is a function representation

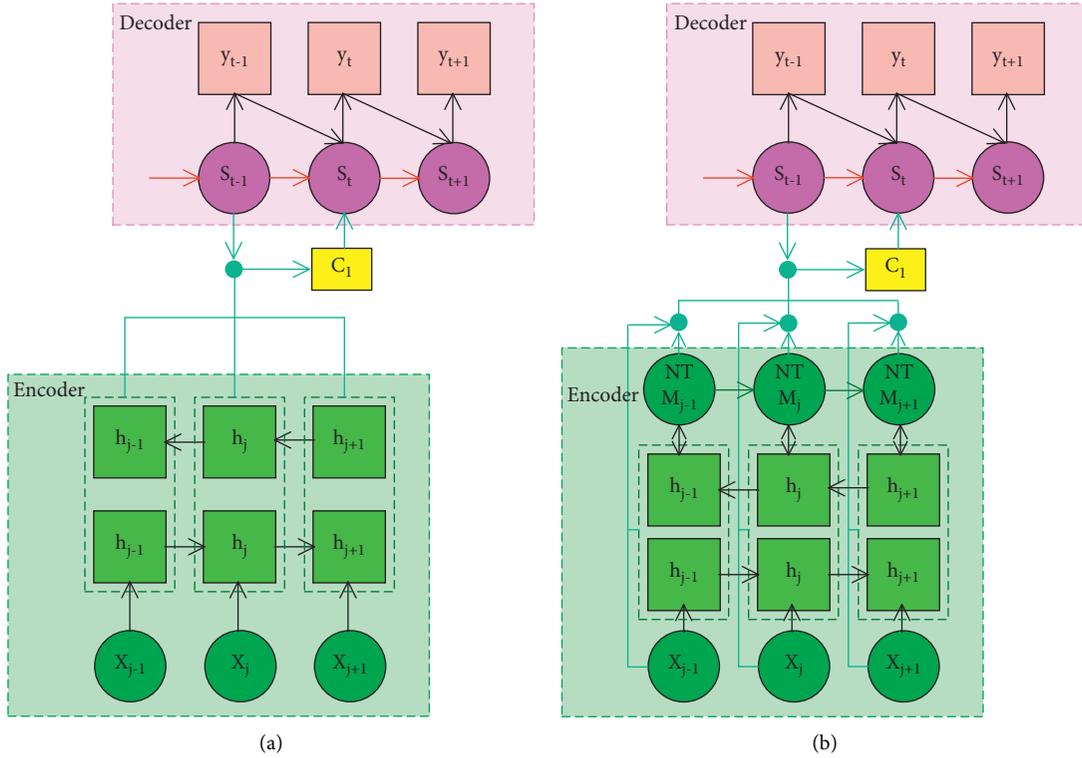


FIGURE 3: The structural principle of the NMT model based on attention. (a) Conventional NMT. (b) NMT with multichannel encoder.

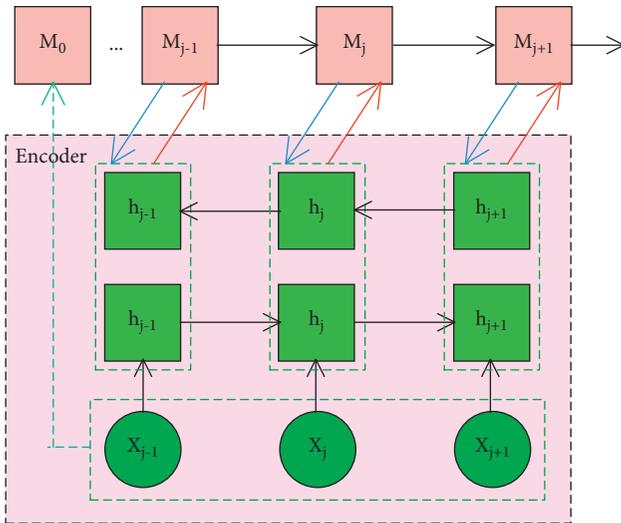


FIGURE 4: Coding layer of the neural translation system.

of the state of all the hidden layers at the previous moment. This structure enables the recurrent neural network to process data samples with front-to-back dependencies.

In a single hidden layer feedforward neural network, the vectorization method is used to calculate, where s represents the sigmoid function, that is, the activation function of the hidden layer; assuming that the length of the hidden layer is h , the number of samples is n , and the feature vector dimension is for batch data of x ; the output of the hidden layer is represented as

$$H = \sigma(XW_{xh} + b_h),$$

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}. \quad (10)$$

Figure 5 shows the basic processing flow of constructing an encoder-decoder machine translation model using a recurrent neural network. The encoder end input is a source language sequence with a start mark and an end representation, which is converted into a distributed representation of word vector data and then passed into the encoder end neural network, and the background vector containing the source language information is passed. As the input of the neural network on the decoder side, finally input the target language sequence through the neural network calculation. Add “<bos>” (beginning of sequence) in front of the input sequence to indicate the beginning of the sequence, and add the character “<eos>” (end of sequence) after the input and output sequences to indicate the termination of the sequence and character to determine the termination condition of the current sequence.

In the recurrent neural network, the hidden layer state at the current moment records the network information at the previous moment, and the output of the network is calculated based on this state. In actual use, the word represented by one-hot is first mapped into a word vector representation using a distributed representation method and then used as the input of the recurrent neural network at each time to achieve sequence output. In the training process, when the time sequence length of each time series training data sample is large or time t is small, the hidden layer variables of the

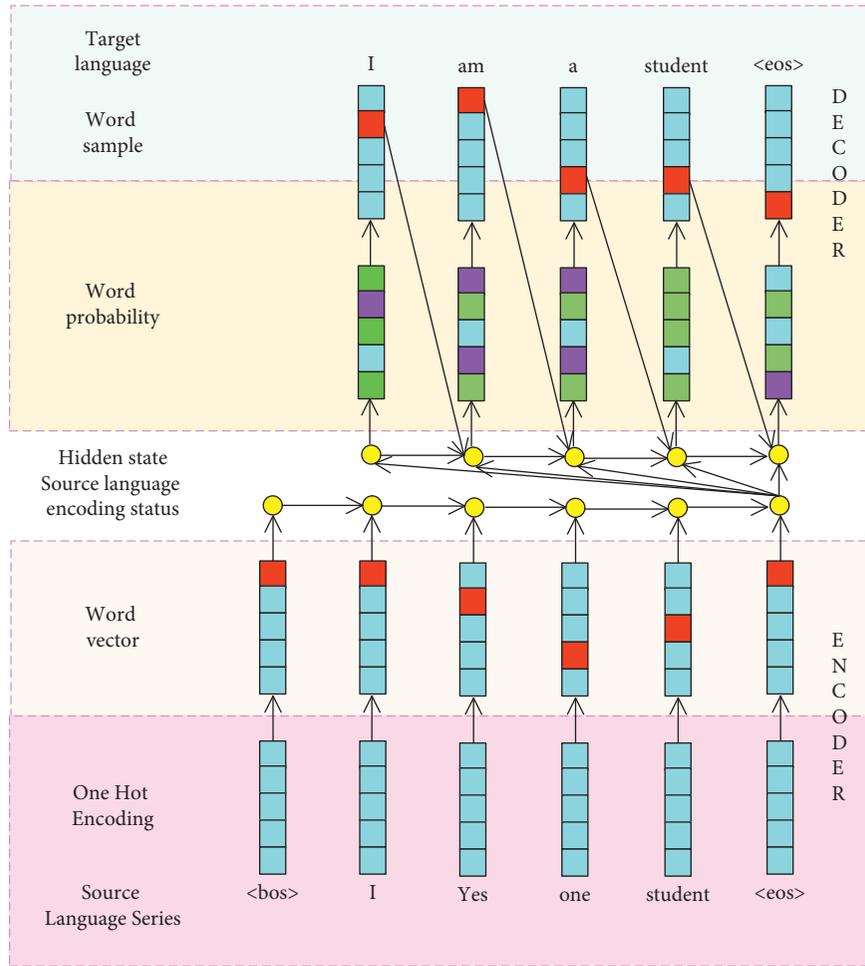


FIGURE 5: Encoder-decoder framework.

objective function at time t may have gradient decay or gradient explosion. Since the deep neural network training adopts the method of back propagation, chain derivation is required. The gradient calculation of each layer will involve continuous multiplication operation. In the face of multilayer networks, if the continuous multiplication factor is mostly less than 1, the final product may tend to 0, that is, gradient attenuation occurs, resulting in the subsequent network layer parameters not changing, and updated iterations cannot be performed; similarly, if the multiplication factor is mostly greater than 1, the final product may tend to infinity, that is, a gradient occurs and explodes.

In the face of gradient explosion, gradient clipping is required. All gradients are spliced into one vector, and the clipping threshold is set to θ . According to (10), it can ensure that the magnitude of the gradient vector does not exceed the threshold:

$$g = \min\left(\frac{\theta}{|g|}, 1\right)g. \quad (11)$$

According to the translation model constructed in the previous content, it can be found that the output of the

encoding stage is a fixed-dimensional vector, which leads to the use of the same-dimensional vector to encode the semantic syntax and other information, regardless of the length of the source language sequence. The input involved in the translation process is a sequence of variable length, and the constructed model may not fully fit the source language input sequence, and the processing effect is not ideal when dealing with long sentence sequences. At the same time, from the perspective of human translation, while translating, more attention will be paid to the part of the source language that is more closely related to the current translation. As the translation progresses, the part of concern will also change, but the use of fixed-dimensional vectors here means all parts of the source language sequence are given the same degree of attention, which is obviously not conducive to the improvement of machine translation performance. The decoder model strengthens the representation ability of the source language sequence. As shown in Figure 6, the neural machine translation model consists of three parts: the encoder, the decoder, and the attention mechanism, using a set of multiple vectors instead of a fixed vector to represent the source language information, the target sequence generation process dynamically selects the background vector, and the

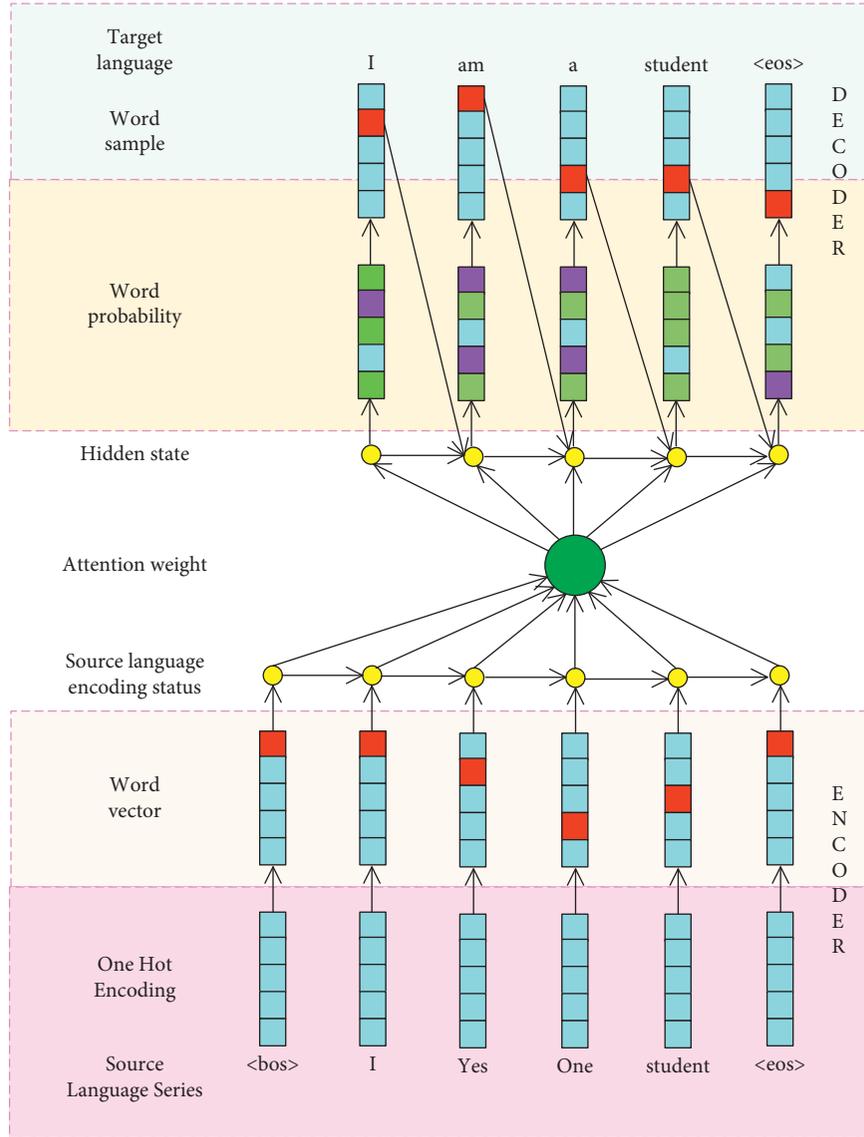


FIGURE 6: Translation model using the attention mechanism.

decoding process pays more attention to the part of the source language that is more relevant to the target sequence.

Similar to the decoder part proposed in this part, the calculation formula for the next hidden layer state of the target end is

$$Z_{i+1} = \sigma(c_i, u_i, z_i). \quad (12)$$

At this time, the background vector is no longer uniform and fixed, and C_i is the i^{th} word in the source language sentence, which makes it possible to find a unique background vector C_i corresponding to each word.

The calculation formula is

$$C_i = \sum_{j=1}^T \alpha_{ij} h_j. \quad (13)$$

This formula actually performs a weighted average calculation on the hidden layer state. The calculation formula is

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^T \exp(e_{ik})}, \quad (14)$$

$$e_{ij} = a(z_i, h_j).$$

The attention mechanism allows the translation model to better solve the problem of long-distance dependence by adding different weights to the source language. According to the experimental results of Bahdanau et al., as shown in Figure 7, through the introduction of the attention mechanism, the translation model's translation effect for long sentences has been significantly improved compared to the model without the attention mechanism. The exact model is shown in Figure 8.

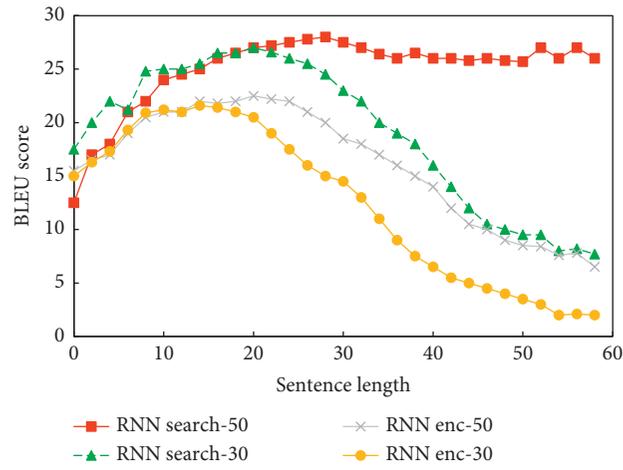


FIGURE 7: Attention mechanism performance.

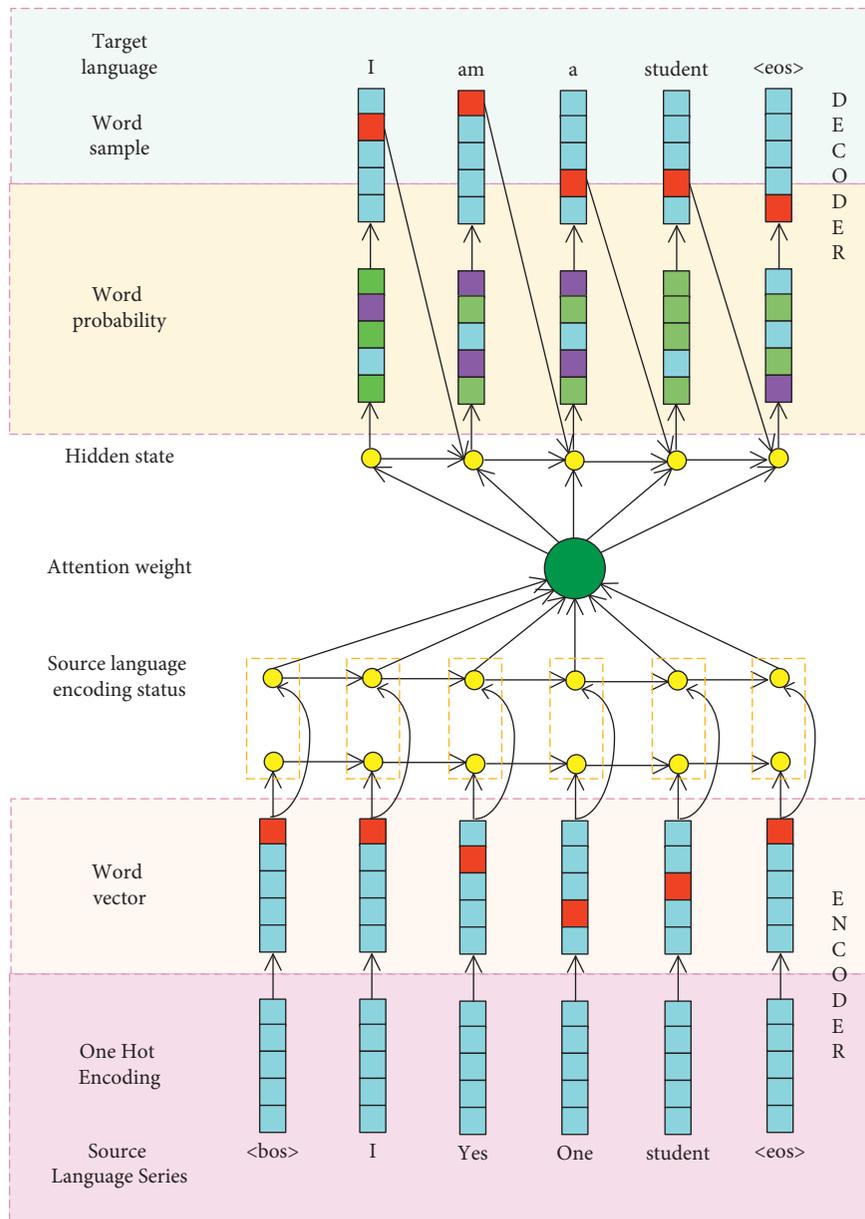


FIGURE 8: BiGRU model structure.

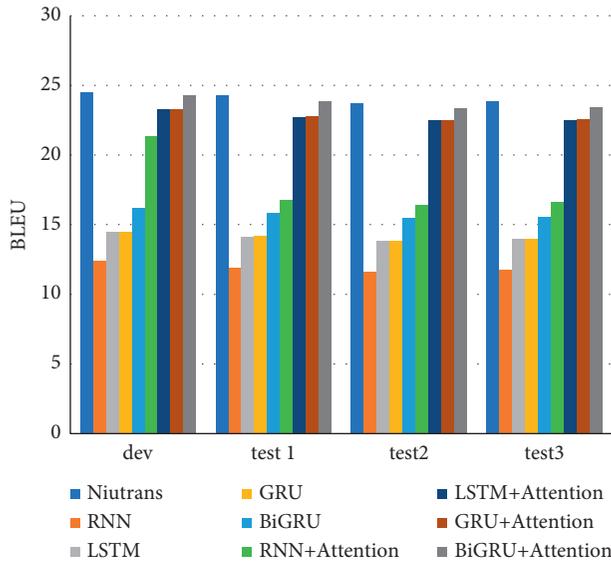


FIGURE 9: BLEU values of different translation models on the development set and test set.

4. Experimental Comparative Analysis

In view of the actual situation of the research in this article, the experimental data in this section use the IWSLT dataset with a smaller data scale for experiments. The International Workshop on Spoken Language Translation (IWSLT) is the most influential spoken language machine translation in the world. Evaluation of the game: a comparative analysis of the various neural network models mentioned in this chapter, and an experimental comparison of Chinese-English machine translation tasks. The experimental data use the IWSLT 2015 dataset, which includes 220000 Chinese-English parallel sentence pairs training data, 1 pair development set data, and 3 pairs of test set data. Build a neural machine translation system on the deep learning framework TensorFlow.

Different neural networks are used in the translation model used for training separately. The experimental results are shown in Figure 9. The first column in the figure shows the translation models of different types of neural networks, and the second to fourth columns show that several translation models are in the BLEU value performance on the development set and test set.

Comparing lines 2–5 in the figure, it can be found that under the premise that the attention mechanism is not used, the performance of the model based on the RNN structure is better than that of the model using the LSTM and GRU structure. The two-way GRU proposed in this study is in the original, improved on the basis. The 5–8th lines in the figure are all translation models that incorporate the attention mechanism on the original structure.

5. Conclusions

This article mainly focuses on Chinese-English machine translation tasks. The encoder and decoder parts are constructed by the neural network, and the word vector method

of distributed representation is adopted, source language and target language sequence. Different translation models are constructed for comparison and analysis of three kinds of neural networks: recurrent neural network, long-term short-term memory, and gated recurrent unit. The attention mechanism is added to improve the translation model's processing of long-distance dependence. A neural machine translation model based on two-way GRU is proposed to improve the contextual representation ability of the source language, and experiments are carried out to verify the effectiveness of the improved method.

Neural machine translation uses neural networks to switch between the source language and the target language. The relatively closed translation process makes it difficult to explicitly use prior knowledge. Aiming at the problem that neural machine translation cannot make good use of linguistic knowledge, proposed a neural machine translation model that adds part-of-speech sequence information. On the basis of the two-way GRU translation model fused with the attention mechanism, Stanford Parser is used for syntactic analysis to obtain the part-of-speech sequence information, and then, it is integrated into the translation model in the form of two-way encoding. The encoder part uses vector splicing to form the background vector together. Verify the effectiveness of the analysis and improvement methods.

Data Availability

The dataset used to support the findings of this study can be accessed upon request from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [2] M. Yun, J. Zhao, J. Zhao, X. Weng, and X. Yang, "Impact of in-vehicle navigation information on lane-change behavior in urban expressway diverge segments," *Accident Analysis & Prevention*, vol. 106, no. 1, pp. 53–66, 2017.
- [3] S. Kumar Dwivedi, R. Amin, V. Satyanarayana, and C. Rashmi, "Blockchain-based secured event-information sharing protocol in internet of vehicles for smart cities," *Computers & Electrical Engineering*, vol. 86, no. 1, pp. 1–9, 2020.
- [4] Z. Khan and S. Amin, "Bottleneck model with heterogeneous information," *Transportation Research Part B: Methodological*, vol. 112, no. 1, pp. 157–190, 2018.
- [5] J. M. Cairney, K. Rajan, D. Haley et al., "Mining information from atom probe data," *Ultramicroscopy*, vol. 159, no. 1, pp. 324–337, 2020.
- [6] J. Yu and P. Lu, "Learning traffic signal phase and timing information from low-sampling rate taxi GPS trajectories," *Knowledge-Based Systems*, vol. 110, no. 1, pp. 275–292, 2016.

- [7] K. P. Wijayarathna, V. V. Dixit, and L. Denant-Boemont, S. Waller, "An experimental study of the Online Information Paradox: does en-route information improve road network performance?" *PLoS ONE*, vol. 12, no. 9, pp. 184–191, 2017.
- [8] Z. Wang, H. Ren, Q. Shen, W. Sui, and X. Zhang, "Seismic performance evaluation of a steel tubular bridge pier in a five-span continuous girder bridge system," *Structures*, vol. 31, no. 1, pp. 909–920, 2021.
- [9] S. Nakayama and J. Takayama, "Traffic network equilibrium model for uncertain demands," in *Proceedings of the 82nd Transportation Research Board Annual Meeting*, Washington, DC, USA, 2021.
- [10] H. Shao, W. H. K. Lam, and M. L. Tam, "A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand," *Networks and Spatial Economics*, vol. 6, no. 3, pp. 173–204, 2019.
- [11] A. Chen, J. Kim, S. Lee, and Y. Kim, "Stochastic multi-objective models for network design problem," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1608–1619, 2020.
- [12] H. Wang, W. H. K. Lam, X. Zhang, and H. Shao, "Sustainable transportation network design with stochastic demands and chance constraints," *International Journal of Sustainable Transportation*, vol. 9, no. 2, pp. 126–144, 2015.
- [13] S.-M. Hosseininasab and S.-N. Shetab-Boushehri, "Integration of selecting and scheduling urban road construction projects as a time-dependent discrete network design problem," *European Journal of Operational Research*, vol. 246, no. 3, pp. 762–771, 2015.
- [14] S.-M. Hosseininasab, S.-N. Shetab-Boushehri, S. R. Hejazi, and H. Karimi, "A multi-objective integrated model for selecting, scheduling, and budgeting road construction projects," *European Journal of Operational Research*, vol. 271, no. 1, pp. 262–277, 2018.
- [15] M. Johnson, M. Schuster, and Q. V. Le, M. Krikun, Y. Wu, Z. Chen et al., "Google's multilingual neural machine translation system: enabling zero-shot translation," *Transactions of the Association for Computational Linguistics*, vol. 5, no. 1, pp. 339–351, 2017.
- [16] M. D. Moreno, "Translation quality gained through the implementation of the iso en 17100:2015 and the usage of the blockchain," *Babel*, vol. 1, no. 2, pp. 1–9, 2020.
- [17] X. Wang, X. Yu, L. Guo, F. Liu, and L. Xu, "Student performance prediction with short-term sequential campus behaviors," *[J]. Information*, vol. 11, no. 4, p. 101, 2020.
- [18] Q. Guo, Z. Zhu, Q. Lu, D. Zhang, and W. Wu, "A dynamic emotional session generation model based on Seq2Seq and a dictionary-based attention mechanism," *Applied Sciences*, vol. 10, no. 6, pp. 1–10, 2020.
- [19] H. Ren, Xi Mao, W. Ma, J. Wang, and L. Wang, "An English-Chinese machine translation and evaluation method for geographical names," *ISPRS International Journal of Geo-Information*, vol. 9, no. 3, pp. 193–201, 2020.
- [20] J. Arús-Pous, T. Blaschke, S. Ulander, J.-L. Reymond, H. Chen, and O. Engkvist, "Exploring the GDB-13 chemical space using deep generative models," *Journal of Cheminformatics*, vol. 11, no. 1, pp. 20–29, 2019.
- [21] T. Moon, T. Ahn, and J. Son, "Long short-term memory for a model-free estimation of macronutrient ion concentrations of root-zone in closed-loop soilless cultures," *Plant Methods*, vol. 15, no. 1, pp. 1–12, 2019.
- [22] N. Pourdamghani and K. Knight, "Neighbors helping the poor: improving low-resource machine translation using related languages," *Machine Translation*, vol. 33, no. 3, pp. 239–258, 2019.
- [23] L. Bote-Curiel, S. Muñoz-Romero, A. Gerrero-Curieses, and J. L. Rojo-Álvarez, "Deep learning and big data in healthcare: a double review for critical beginners," *Applied Sciences*, vol. 9, no. 11, pp. 1–11, 2019.
- [24] J. Zhang and T. Matsumoto, "Corpus augmentation for neural machine translation with Chinese-Japanese parallel corpora," *Applied Sciences*, vol. 9, no. 10, pp. 1–12, 2019.
- [25] Y. Chen, Y. Ma, X. Mao, and Q. Li, "Multi-task learning for abstractive and extractive summarization," *Data Science and Engineering*, vol. 4, no. 1, pp. 14–23, 2019.
- [26] Pu Zhou and Z. Jiang, "Self-organizing map neural network (SOM) downscaling method to simulate daily precipitation in the Yangtze and Huaihe River Basin," *Climatic and Environmental Research*, vol. 21, no. 5, pp. 512–524, 2016.
- [27] X. Xiao, "Analysis on the employment psychological problems and adjustment of retired athletes in the process of career transformation," *Modern Vocational Education*, vol. 5, no. 12, pp. 216–217, 2018.
- [28] S. Sahoo and M. K. Jha, "Pattern recognition in lithology classification: modeling using neural networks, self-organizing maps and genetic algorithms," *Hydrogeology Journal*, vol. 25, no. 2, pp. 311–330, 2016.
- [29] Y. Zhou and B. Yang, "Sports video athlete detection using convolutional neural network," *Journal of Natural Science of Xiangtan University*, vol. 39, no. 1, pp. 95–98, 2017.
- [30] J. Pang, "Research on the evaluation model of sports training adaptation based on self-organizing neural network," *Journal of Nanjing Institute of Physical Education*, vol. 16, no. 1, pp. 74–77, 2017.
- [31] G. Querczola, C. Lovati, C. Mariani, and L. Pantoni, "A semi-quantitative sport-specific assessment of recurrent traumatic brain injury: the TraQ questionnaire and its application in American football," *Neurological Sciences*, vol. 40, no. 9, pp. 1909–1915, 2019.
- [32] J. Wang, X. Luo, and H. Yan, "Correlation analysis between injuries and functional movement screening for athletes of the National Shooting Team," *Journal of Capital Institute of Physical Education*, vol. 5, no. 4, pp. 352–355, 2016.
- [33] G. Ma, "Research on the design of juvenile football players' sports injury prediction model," *Automation Technology and Application*, vol. 277, no. 7, pp. 141–144, 2018.