Tibetan-Chinese Neural Machine Translation Combining Attention Mechanism

Tao Jiang*, Hao Sun, Yu Gang Dai, Ding Liu

Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, Gansu, 730000, China

*Corresponding author’s e-mail: xinxiyuanjt@126.com

Abstract: Neural Machine Translation (NMT) has developed rapidly in recent years, and translation effects in most languages have surpassed statistical machine translation methods. The Seq2Seq framework has brought great advantages to machine translation, but the model still has great limitations in the ability to capture long-distance information. Recurrent neural networks (RNN) and LSTM networks are proposed to improve this problem, but the effect is not obvious. However, the proposal and application of the attention mechanism effectively compensated for this defect. This paper uses the attention mechanism as the basis to construct the encoder-decoder framework. This method analyzes the mechanism and principle of the attention model to implement the Tibetan-Chinese neural machine translation system. Compared with the previous neural network translation models, the results show that the translation model that incorporates the attention mechanism can obtain better translation results.

1. Introduction

Currently, the research of Tibetan-Chinese machine translation has received high attention, and the application requirements of Tibetan-Chinese and Chinese-Tibetan machine translation systems are also increasing. However, the existing research results of Chinese-Tibetan machine translation are few, and its research is even less. As an important research content of Tibetan information technology, the research of Chinese-Tibetan machine translation is still less developed. With the rapid advancement of informatization in my country's Tibetan areas, Chinese-Tibetan machine translation technology has shown a wide range of needs in many fields such as e-government, education, academic research, publishing, and information security management. Under this background, it is very important to study the key technologies of Chinese-Tibetan statistical machine translation, which will not only help enrich the theory of machine translation, but also promote the substantial development of Tibetan information technology. It laid the foundation for the final development of the Tibetan-Chinese bidirectional machine translation system.

2. Related Work

Machine Translation (Automatic Translation) is a process that uses a computer to convert one natural language (source language) into another natural language (target language). It is one of the most important research directions in the field of natural language processing[2]. In recent years, with the progress of deep learning research, Neural Machine Translation based on deep learning has also made breakthrough progress. Both in terms of translation efficiency and translation quality, they have gradually surpassed the traditional statistical-based machine translation methods. With the
development and improvement of technology, the difficult task of machine translation will gradually be overcome.

Deep learning is a new field in machine learning research. It establishes and simulates the neural network of the human brain for analysis and learning, and interprets the data by mimicking the mechanism of the human brain. At present, deep learning has made breakthroughs in many core problems of natural language processing. Machine translation tasks are one of the typical examples.

In machine translation research, using neural network machine translation to achieve direct translation from source language to target language, which greatly improves the translation effect. This method is the frontier hotspot of current machine translation research. Neural network-based machine translation methods originated in the 1990s[3] and have not become a mainstream method due to resource constraints. After the rise of deep learning, neural networks are usually used for word alignment and translation rule extraction based on statistical machine translation[4]. Since 2014, Sutskever, Cho, Jean and others [5-8] have proposed a new method for implementing machine translation simply by using neural networks, which is called neural network machine translation. This method realizes the direct translation from the source language to the target language, and its translation effect surpasses the traditional statistical machine translation method in multiple language pairs. This makes it gradually become the current mainstream machine translation method.

Research on Tibetan machine translation mainly focuses on statistical machine translation and basic research on Tibetan machine translation. Nima Zhaxi [9] proposed the implementation method of Chinese-Tibetan machine translation system based on neural network in 2014. In terms of Tibetan word segmentation, the system uses the discriminant model-based Tibetan word segmentation method to study the correlation between the results of Tibetan word segmentation and the minimum word formation granularity. Li Yachao [10] proposed a method of using transfer learning to solve the problem of scarcity of Tibetan Chinese materials. Experiments show that this method improves 3 BLEU values compared with the phrase statistical machine translation method. At present, the online Chinese-Tibetan translation system includes ‘Tencent Tibetan-Chinese Translation’ developed by Tencent, ‘Tibetan-Chinese Intelligent Translation System’ developed by China National Language and Translation Bureau, the ‘Maverick Translation Open Platform’ developed by the Natural Language Processing Laboratory of Northeastern University and Xiamen University and ‘Cloud Translation’ developed by Natural Language Processing Lab. These Tibetan-Chinese machine translation systems all use translation technology based on neural networks.

Neural machine translation represents a novel machine translation model. Through continuous improvement, its performance level has been significantly improved, making this method widely used [11]. The characteristic of neural machine translation is that the difficulty of application is low, and it can comprehensively apply learning and memory technology to deal with long-distance dependence problems, so it shows a high performance advantage in this field. The disadvantage of this method is the high dependence on the training algorithm, we should continue to improve and optimize.

3. Neural Network Machine Translation

Neural machine translation is a machine translation method based on deep learning, which uses an encoder-decoder framework. The basic idea is to use a neural network to input sentences, which is called an encoder (Encoder). The method is to compress the entire sentence information into a fixed-dimensional encoding; then it uses another neural network, which is called a decoder (Decoder). This method reads this code and "decompresses" it into a sentence in the target language [12]. Neural machine translation uses neural networks to complete the direct translation of the source language to the target language in modelling. It does not require steps such as word alignment, translation rule extraction, and sequence adjustment based on statistical machine translation. This section mainly introduces the model structure of Tibetan-Chinese neural network machine translation incorporating attention mechanism.
3.1. Neural network machine translation model

Neural network machine translation originates from sequence-to-sequence learning. This article takes the translation model proposed by the University of Montreal as an example [9~10]. The encoder-decoder model is one of the neural network machine translation models. The encoder reads the source language sentences and encodes it into a vector with a fixed dimension. The decoder reads the vector and generates the target language word sequence in turn. The model is shown in Figure 1. The main parts of the neural network machine translation model are as follows: \( x \) represents input, \( h \) represents hidden state, and \( y \) represents output.

\[
\text{Encoder:} \quad h_i = f(x_i, h_{i-1}) \\
c = q(h_{i-1}, h_i) 
\]

(1)

(2)

c is the sentence representation of the source language, and \( f \) and \( q \) are nonlinear functions.

Given the source language representation \( c \) and the preamble output sequence, the decoder generates target words in sequence, as shown in formula (3).

\[
p(y) = \prod_{t=1}^{T} p(y_t | \{y_1, \ldots, y_{t-1}\}, c) 
\]

(3)

\( y = (y_1, y_2, \ldots, y_T) \), The decoder usually also uses a recurrent neural network in the form of formula (4).

\[
p(y_t | \{y_1, \ldots, y_{t-1}\}, c) = g(y_{t-1}, s_t, c) 
\]

(4)

Model training: The encoder and decoder can be jointly trained, as shown in formula (5).

\[
L(\theta) = \max_{\theta} \frac{1}{N} \sum_{n=1}^{N} \log p(y_n | x_n) 
\]

(5)

Among them, \( \theta \) is the parameter of the model, which represents the bilingual training corpus. It is usually calculated by gradient descent.

3.2. Neural network machine translation incorporating attention mechanism

Traditional neural network machine translation models represent source language sentences as a fixed vector, but this method still has shortcomings. For example, vectors of fixed size cannot fully express the semantic information of source language sentences. Neural network machine translation based on attention mechanism encodes source language sentences into vector sequences. When generating a target language word, the source language word information related to generating the word is dynamically found through the attention mechanism. Therefore, this method greatly enhances the expression ability of neural network machine translation and improves the translation effect. The core idea of the attention mechanism is that the input source sentence is input into the encoder, and the encoder generates an output. Then it adds the weight vector to the output as the input of the decoder.
The decoder uses the hidden state as input to the query at each step of decoding to obtain the encoded hidden state. Then, it calculates a weight reflecting the degree of correlation with the query input at each input position. Finally, the weighted average of the hidden states of each input position is calculated according to the weights [13-14]. The implementation details of the attention mechanism are shown in Figure 2.

![Figure 2 Attention mechanism](image)

Neural machine translation based on the attention mechanism only focuses on a part of the source language sentences during decoding, which can enhance the ability of long sentence representation. Compared with ordinary neural network machine translation, this method integrates more source language information during decoding, which can significantly improve the effect of machine translation. This method is currently the mainstream method of neural machine translation.

4. Experiment And Analysis

4.1. Experimental data and parameters

The parallel corpus used in this experiment is derived from the 2011 Chinese-Chinese Machine Translation Seminar (CWMT) Tibetan-Chinese machine translation evaluation corpus. The Tibetan Chinese material mainly comes from the government field. The corpus includes 100,000 sentences for training data and 650 sentences for test data set. This paper compares and analyzes the neural network translation model incorporating the attention mechanism with the existing neural machine translation tasks.

The deep learning framework used in the experiment is Pytorch, which has a great performance advantage over other deep learning frameworks. Pytorch's architecture is more flexible and can be trained and applied on multiple platforms. At present, it has been widely recognized and applied in industry. The relevant parameter settings in this experiment are shown in Table 1.

| parameter type  | Value   |
|-----------------|---------|
| Vocab size      | 30000   |
| Vector dim      | 512     |
| Hidden layer    | 1024    |
| Learning rate   | 0.1     |
| Drop out        | 0.1     |
| Batch_size      | 4096    |
| optimizer       | Adam    |

4.2. Experimental Results Comparative

The baseline used in this article is the Niutrans phrase statistical machine translation system developed by Northeastern University [15], which is represented by ‘Niutrans’. In this paper, the neural network
machine translation system uses the neural network machine translation system developed by the group based on the work of Bahdana [16], which is represented by ‘NMT’. The Tibetan-Chinese neural network machine translation system with attention mechanism is denoted by ‘NMT+Att’. Since our test corpus is relatively small, the test corpus is also used as the development set of the training phrase translation model in statistical machine translation. This means that the statistical machine translation results are the results on the development set. The test corpus in neural network machine translation is also used as the development set. It is only used to select the optimal translation model, which does not affect the translation model.

In order to evaluate the translation model, this experiment uses the internationally accepted machine translation evaluation script and case-insensitive BLEU value to automatically evaluate the translation quality. The experimental results are shown in Table 2.

| Model          | BLEU |
|----------------|------|
| Niutrans       | 33.40|
| LSTM+Att       | 29.19|
| NMT            | 30.28|
| NMT+Att(ours)  | 35.27|

By analyzing the result in Table 2, it can be seen that the NMT model that incorporates the attention mechanism has a greater improvement in translation performance than the traditional LSTM model. It shows that the model we used can better obtain the context information in the language, and get better performance. The translation model used in this article achieved the best performance in the comparative experiment. This shows that the NMT model combined with the attention mechanism can effectively solve the problems of long-term dependence and weak ability to obtain long-distance information in the traditional recurrent neural network.

Although the data size of the Tibetan-Chinese parallel corpus is smaller than that of the traditional English-Chinese and English-French parallel corpus, the neural network machine translation compare to the statistical machine translation have a significant increase in the BELU value. On the one hand, the context information is better used in our method. On the other hand, the field of the training corpus and the test corpus is relatively single. Therefore, the lack of corpus resources is the biggest obstacle to Tibetan-Chinese machine translation. In future research, we will introduce other methods to improve the effect of machine translation of Tibetan-Chinese neural networks under conditions of scarce resources, such as transfer learning methods. And study the application of this method in other languages.

5. Conclusions And Prospects
In this paper, the attention mechanism and neural machine translation are fused to realize the Tibetan-Chinese neural machine translation system based on the attention mechanism. Through the comparative analysis of experiments, it is verified that the attention mechanism can effectively improve the effect of machine translation. Due to the small size of the Tibetan-Chinese parallel corpus, it is difficult to achieve good translation results. In the next step, we will use language pairs with rich corpus resources such as English-French and English-Chinese, and then adopt the method based on transfer learning to further improve the performance of the Tibetan-Chinese machine translation system.

ACKNOWLEDGMENTS
The work was supported by the Ministry of Education Humanities and Social Sciences Research Project (No.18YJCZH065), and National Key Research and Development Plan (No.2017YFB1002103).
REFERENCES

[1] Zhang Jiajun, Zong Chengqing. Application of neural network language model in statistical machine translation [J]. Information Engineering, 2017, 3(03): 21-28.

[2] Liu Yang. Advances in the frontiers of neural machine translation [J]. Computer Research and Development, 2017, 54(06): 1144-1149.

[3] Neco R P, Forcada ML. Asynchronous translations with recurrent neural nets [C]//Proceedings of International Conference on Neural Networks (ICNN'97). IEEE, 1997, 4: 2535-2540.

[4] Zhang J, Zong C. Deep neural networks in machine translation: An overview [J]. IEEE Intelligent Systems, 2015 (5): 16-25.

[5] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]//Advances in neural information processing systems. 2014: 3104-3112.

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

[7] Cho K, Van Merriënboer B, Bahdanau D, et al. On the properties of neural machine translation: Encoder-decoder approaches[J]. arXiv preprint arXiv:1409.1259, 2014.

[8] Jean S, Cho K, Memisevic R, et al. On using very large target vocabulary for neural machine translation[J]. arXiv preprint arXiv:1412.2007, 2014.

[9] Wan Mozhaixi, Nima Tashi. Research on Several Key Issues in Tibetan Automatic Word Segmentation [J]. Journal of Chinese Information Processing, 2014, 28(04): 132-139.

[10] Li Yachao, Xiong Deyi, Zhang Min, Jiang Jing, Ma Ning, Yin Jianmin. Research on Tibetan-Chinese Neural Network Machine Translation [J]. Journal of Chinese Information Processing, 2017, 31(06): 103-109.

[11] Hou Qiang, Hou Ruili. A review of research and development of machine translation methods [J]. Computer Engineering and Applications, 2019, 55(10): 30-35+66.

[12] Dyer C, Kuncoro A, Ballesteros M, et al. Recurrent neural network grammars[J]. arXiv preprint arXiv:1602.07776, 2016.

[13] Kim Y, Denton C, Hoang L, et al. Structured attention networks[J]. arXiv preprint arXiv:1702.00887, 2017.

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

[15] Weiss K, Khoshgoftaar T M, Wang D D. A survey of transfer learning[J]. Journal of Big data, 2016, 3(1): 9.

[16] Zoph B, Yuret D, May J, et al. Transfer learning for low-resource neural machine translation[J]. arXiv preprint arXiv:1604.02201, 2016.