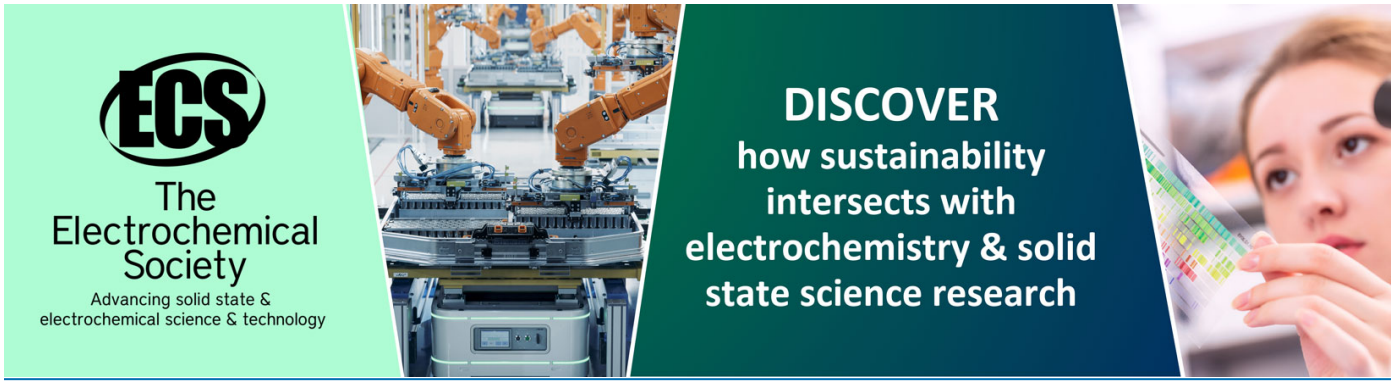
Investigating Back-Translation in Tibetan-Chinese Neural Machine Translation



Investigating Back-Translation in Tibetan-Chinese Neural Machine Translation

Ding Liu [a], Yachao Li*, Dengyun Zhu, Xuan Liu, Ning Ma, Ao Zhu

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

[a]y182730594@stu.xbmu.edu.cn *Corresponding author's e-mail: harry_lyc@foxmail.com

Abstract. In recent years, the proposal of neural network has provided new idea for solving natural language processing, and at the same time, neural machine translation has become the frontier method of machine translation. In low-resource languages, due to the sparse bilingual data, the model needs more high-quality data, and the translation quality fails to achieve the desired effect. In this paper, experiments on neural network machine translation based on attention are conducted on Tibetan-Chinese language pairs, and transfer learning method combined with back translation method is used to alleviate the problem of insufficient Tibetan-Chinese parallel corpus. Experimental results show that the proposed transfer learning combined with back translation method is simple and effective. Compared with traditional translation methods, the translation effect is significantly improved. From the analysis of translation, it can be seen that the citation of Tibetan-Chinese neural machine translation is smoother, which is greatly improved compared to the translation without back translation. At the same time, there are common deficiencies in neural machine translation such as inadequate translation and low translation loyalty.

1. Introduction

Machine translation studies how to use computer to automatically translate between natural languages. Machine translation methods can be divided into rule-based machine translation, case-based machine translation, statistical machine translation and neural machine translation which are widely used at present.

With the development of machine learning technology, neural machine translation based on deep learning is gradually emerging. Since 2014, it has gained obvious advantages in most tasks in just a few years[3]. In neural machine translation, word strings are represented as real vectors, that is distributed vectors. In this way, the translation process is not carried out on the discrete words and phrases, but on the real vector space, so the expression of word order has undergone essential changes. Compared with statistical machine translation, the advantage of neural machine translation is that it does not need Feature Engineering, and all information is automatically extracted from the original input by neural network. Moreover, compared with the continuous model, the method can provide more information for sentence representation. In addition, natural network storage devices are suitable for small applications. However, there are also problems in NMT. First of all, although it is separated from feature engineering, the structure of neural network needs to be designed manually. Even if the structure is well designed, the optimization of the system and the setting of super parameters still rely on a large number of experiments. Secondly, neural machine translation is lack of interpretability, and its process and human

cognition are quite different, and the degree of intervention through human prior knowledge is poor; Thirdly, neural machine translation relies heavily on data, and the scale and quality of data have great influence on performance. Especially in the case of data scarcity, it is challenging to fully train neural networks.

Due to the problem of data sparsity, the related researches of Tibetan Chinese machine translation are mainly focused on statistical machine translation and Tibetan machine translation. On the whole, the research on Tibetan machine translation is lagging behind, and the application and actual effect of neural network in Tibetan language are rarely published.

2. Neural Machine Translation

Neural machine translation is a new machine translation method proposed by kalchbrenner et al. [12], sutskever et al. [5] and Li. [13]. In neural machine translation, the process of sequence to sequence conversion can be realized by the encoder decoder framework, in which the encoder encodes the source language sequence and extracts the information in the source language for distributed representation, and then the decoder converts the information into another language expression [14]. In modeling, neural network is used to complete the direct translation from source language to target language, without the steps of word alignment, translation rule extraction and order adjustment. This section mainly introduces the transformer model based on self attention and back translation data enhancement technology.

The traditional seq2seq model of neural network has some shortcomings: CNN can't directly process long sequence samples, RNN can't parallel computing. Although the seq2seq model completely based on CNN can be implemented in parallel, it takes up a lot of memory, and it is not easy to adjust parameters in large amount of data. In view of the shortcomings of the seq2seq model based on CNN and RNN, in June 2017, Google released a new machine learning framework transformer, which constructs encoders and decoders based on self attention, and builds a seq2seq model based on attention mechanism[11]. The RNN does not use the previous structure of CNN, or it does not retain any inherent structure of CNN[6]. The model can work in high parallel, and has a great improvement in task performance, parallel ability and easy training[15]. The performance of this model in machine translation and other language understanding tasks is far beyond the existing algorithms[7]. Its design is to process all the words in the sequence in parallel, at the same time, it can combine the context with the distant words with the help of self attention mechanism[16]. In each step, the information of each symbol (such as a word in a sentence) can communicate with all other symbols by means of self attention mechanism. The training speed of transformer is much faster than that of RNN, and its translation result is much better than that of RNN. The transformer framework has achieved the best translation performance.

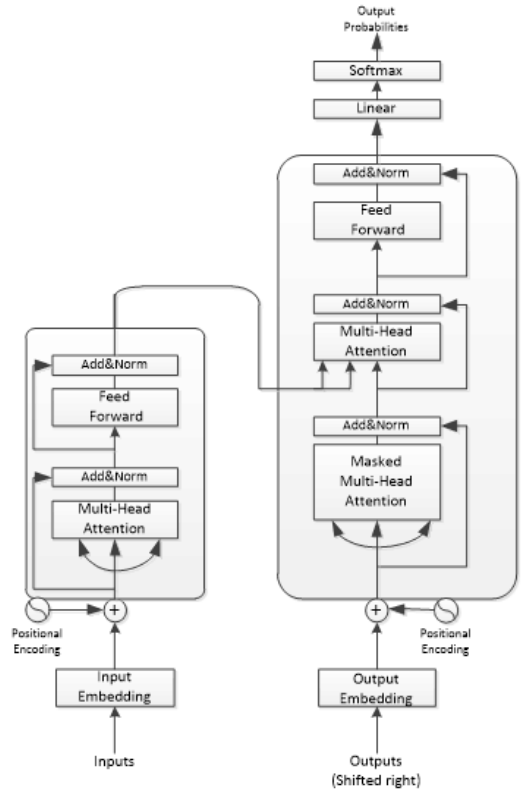


Figure 1 transformer framework

As shown in Figure 1, the transformer framework is still an encoder decoder structure in general, which is a completely attention based encoder decoder model. In a network block of the encoder, it is composed of a multi attention sub layer and a feedforward neural network sub layer, and N blocks are built in the encoder stack. Similar to the encoder, only one more multi attention layer is added to one network block of the decoder. In order to optimize the deep network better, the whole network uses residual connection and add norm[17].

First of all, multiple attention is used to connect the encoder to the decoder. K, V, q are the layer output of the encoder (where k = V) and the input of the multi attention in the decoder. In fact, just like the attention in the mainstream machine translation model, the decoder and encoder attention are used for translation alignment. Then, multiple self attention self attention is used in both encoder and decoder to learn the representation of text. Self attention is K= V= Q. for example, if you input a sentence, then every word in the sentence must be attached to all the words in the sentence. The purpose is to learn the word dependence within a sentence and capture the internal structure of the sentence. The difference of multi attention is that it calculates h times instead of once, which allows the model to learn relevant information in different representation subspaces.

2.1. data enhancement technology via back translation

At present, neural network method has achieved good translation results in low resource machine translation[20]. However, in the face of language resources and lack of corpus size, there is not enough corpus for neural network training, so neural machine translation model is difficult to learn more useful information. Therefore, how to make full use of existing data to alleviate the problem of resource shortage has become an important research direction of neural machine translation. As the acquisition of monolingual data is easier and faster than that of bilingual data, this paper will combine the back translation data enhancement technology to improve the performance of the translation system. The process is as follows:

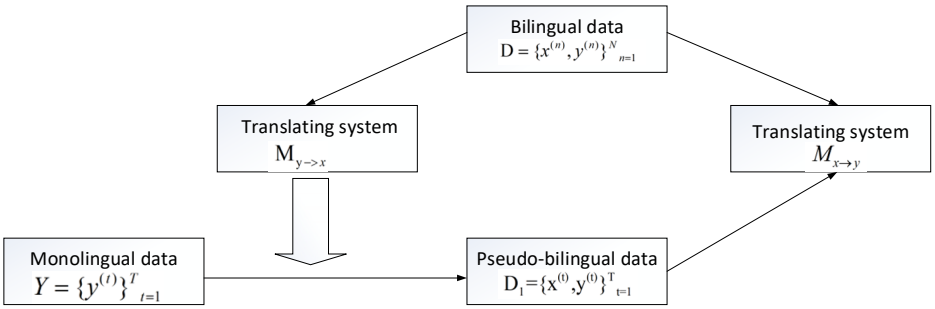


Figure 2 flow chart of back translation

As shown in Figure 2, for example, in the training of Chinese English translation, there are some Chinese English parallel corpora and some unmarked English corpora. First, we train English translation with Chinese English parallel corpus, and then use en_ ZH translation system to get the Chinese translation of unlabeled English corpus, finally put these as ground truth, synthesize parallel data and apply it to training to achieve the effect of expanding corpus.

2.2. Tibetan Chinese neural machine translation under the condition of scarce resources

At present, neural machine translation has made remarkable achievements in large-scale conditions, but in the use of scarce language resources, statistical machine translation can improve performance by using language models, while neural machine translation is difficult to effectively use the only corpus to complete the task. As a data-driven method, the performance of neural machine translation (NMT) is highly dependent on the size, quality and domain coverage of parallel corpora. Only when the training corpus reaches a certain scale, can NMT significantly surpass SMT.

There is also a problem of insufficient corpus in Tibetan-Chinese machine translation. Different from Li et al. [13], the method in this paper not only uses English Chinese model parameter initialization, but also combines Tibetan Chinese material with reverse translation data enhancement [20] technology to carry out experimental research on Tibetan Chinese machine translation. When we train Tibetan Chinese translation, we have some Tibetan Chinese parallel corpus and some unmarked Chinese corpus. First, the Tibetan Chinese parallel corpus is used to train Chinese Tibetan translation, and then used ZH_TI translation system to get the Tibetan translation of unlabeled Chinese corpus, and finally put these as ground truth, synthesize parallel data and apply it to training to achieve the effect of expanding corpus.

3. Experiment And Analysis

3.1. Experimental data and parameters

The parallel corpus used in this experiment is the Tibetan Chinese machine translation evaluation corpus of the 2011 machine translation Seminar (CWMT). The Tibetan Chinese corpus is mainly from the government field. The training data is 100000 sentences and the test data set is 650 sentences.

The deep learning framework used in the experiment is pytoch. Compared with other deep learning frameworks, pytoch has great advantages in performance. Its architecture is more flexible and can complete training and Application on multiple platforms, which has been widely recognized and applied in the industry.

The relevant parameter settings in this experiment are shown in Table 1.

Table 1. Parameter settings

parameter type	**Value**
Vocab size	100000
Vector dim	512
Hidden layer	1024
Learning rate	0.1

Drop out	0.1
Batch_size	4096
optimizer	Adam

3.2. Experimental Results Comparative

This paper uses the baseline to develop and release the transformer neural machine translation system for Google [18], which is represented by "NMT (bi-text)". On the basis of transfer learning, we use the back translation method to express it as "NMT (bi-text + pseudo bilingual data)".

In order to evaluate the translation model, the international machine translation evaluation script and insensitive Bleu value are used to evaluate the translation quality automatically. The experimental results are shown in Table 2.

Table 2. BLEU values

Model	BLEU
NMT(bi-text)	48.02
BackTrans(bi-text+ pseudo bilingual data)	50.10

Through the analysis of the data in Table 2, it can be seen that the performance of the translation system after the fusion of the back translation method is greatly improved compared with the original NMT. It shows that the pseudo data created by the back translation data enhancement method is helpful to the training of the system and can make the machine translation show better performance.

Due to the lack of Tibetan Chinese parallel corpus, corpus resources are the biggest obstacle to Tibetan Chinese machine translation. Although back translation data enhancement technology can effectively alleviate this problem, its translation performance has not achieved the expected effect. In the follow-up study, we will use other methods to improve the effect of Tibetan Chinese neural network machine translation under the condition of scarce resources, such as zero resource method, and study the application of this method in other languages.

4. Related Work

Machine translation, namely automatic translation, is a process of transforming a natural language (source language) into another natural language (target language) by using a computer [1], which is one of the most important research directions in the field of natural language processing [2]. In recent years, with the great progress in the research of deep learning, neural ma chine translation based on deep learning has also made a breakthrough[4]. In terms of translation efficiency and translation quality, it has gradually surpassed the traditional statistical based machine translation method.

In the research of machine translation, neural network machine translation is used to achieve the direct translation from the source language to the target language, which greatly improves the translation effect, and it is the forefront of machine translation research[8]. At present, there are relatively few research results on Tibetan machine translation. Nima Tashi [9] proposed a Chinese Tibetan machine translation technology and system based on neural network in 2014. In terms of Tibetan word segmentation, the system uses the Tibetan word segmentation method based on discriminant model to study the correlation between the Tibetan word segmentation results and the minimum word formation granularity[19]. Li Yachao [10] proposed the method of using transfer learning to solve the problem of the scarcity of Tibetan and Chinese materials, and proved through experiments that this method improved three Bleu values compared with phrase statistical machine translation method. These Tibetan and Chinese machine translation systems all use the translation technology based on neural network. However, due to the lack of Tibetan Chinese bilingual corpus, the translation effect has not been able to achieve the desired effect. In order to improve the effect of Tibetan Chinese neural machine translation, we will use the back translation method to make pseudo data to expand the corpus appropriately.

5. Conclusions And Prospects

This paper studies the application of neural network machine translation in Tibetan Chinese translation. The transfer learning method is used to transfer the parameters of the English Chinese neural network machine translation model to the Tibetan Chinese neural network machine translation model, and the back translation data enhancement technology is used to expand the corpus. This method significantly improves the effect of the baseline neural network machine translation system.

Through the comparative analysis of the experiment, it shows that the successful application of back translation data enhancement technology in low resource neural machine translation can further improve the effect of machine translation. However, the results obtained in this paper are still unsatisfactory because the data set used in the training model is relatively small. In addition, the word vector dimension in the model, the number of training rounds, and the improvement of the model by means of fusion transfer learning may affect the final results. How to adjust the network structure to achieve the optimal effect of the model is a problem that needs to be further studied and solved.

ACKNOWLEDGMENTS

This research was funded by The National Natural Science Fund (NO. 61762076), Fundamental Reserch Funds for the Central Universities (Grand No,31920190114) and the Northwest Minzu University Graduate Research Innovation Project (No.Yxm2020109) .

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 Mozhaxi, 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] Kalchbrenner N, Blunsom P. Recurrent continuous translation models[C]//Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013: 1700-1709.
[13] Li Y, Li J, Zhang M, et al. Improving Neural Machine Translation with Linear Interpolation of a Short-Path Unit[J]. 2020.
[14] Li, Yachao, Junhui Li and Min Zhang. “Adaptive Weighting for Neural Machine Translation.” *COLING* (2018).
[15] 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.

[16] Kim Y, Denton C, Hoang L, et al. Structured attention networks[J]. arXiv preprint arXiv:1702.00887, 2017.
[17] Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation[J]. arXiv preprint arXiv:1508.04025, 2015.
[18] Weiss K, Khoshgoftaar T M, Wang D D. A survey of transfer learning[J]. Journal of Big data, 2016, 3(1): 9.
[19] Zoph B, Yuret D, May J, et al. Transfer learning for low-resource neural machine translation[J]. arXiv preprint arXiv:1604.02201, 2016.
[20] Zoph B, Yuret D, May J, et al. 2016. Transfer Learning for Low-Resource Neural Machine Translation[C]. //Proc of EMNLP. Stroudsburg, PA: ACL, 2016:1568-1575.