Optimization of Dimension-Han Machine Translation Model Based on Neural Network

Jiatong Liu¹, Zhonghao Wang², Yanchang Cui³, Minghong Fan⁴* and Beizhan Wang⁵

¹²⁵Institute of information College, Xiamen University, Xiamen City, Fujian Province, China
³College of software, Heilongjiang University, Harbin, Heilongjiang Province, China
⁴Computer Engineering College Jimei University, 185 Yinjiang Rd., Jimei District, Xiamen City, Fujian Province, China
Corresponding author: Minghong Fan

Abstract. Machine translation is a natural language (source language) converted into another natural language by computer processing. As the main branch of artificial intelligence, machine translation based on neural network has significantly exceeded the traditional statistical machine translation in small language translation. Resource-poor language has become a hot topic, such as Uygur translation, as a typical resource-poor language, although there are some achievements in Uygur information processing, its basic available tools are very limited. This paper first expounds the background of the topic, briefly introduces the background knowledge of related languages. This paper describes the machine translation method based on neural network, and analyzes the existing neural machine translation methods on Uygur-Chinese translation tasks. The analysis of Uygur, Turkish, Arabic and Japanese finally determines the transfer translation of Uygur in Japanese. The parallel corpus of Uygur-Chinese and Japanese-Chinese are used to experiment.

1. Introduction
Machine translation is the main research field of artificial intelligence, and language is the bridge of communication. During the construction of the "Belt and Road" core area, China has had in-depth exchanges with the surrounding Central Asian countries and ushered in a new language life. In this region, cross-border languages carry multiculturalism and can play many important functions in the construction of Belt and Road in China[1]. However, at present, the study of cross-border language in the core area of Xinjiang has started late and lacks scientific and reasonable planning, which makes them unable to play an ideal role in serving social development and national strategy. In some minority areas, most young rural workers do not have Chinese communication skills and production skills, and these young people can not enter labor-intensive processing enterprises for employment. Although bilingual education has been launched in Xinjiang Uygur Autonomous region, it is still a long way from universalization. In the face of remote areas in Xinjiang, it is still very difficult for minority people to use Chinese. Language translation is particularly important for the statistics of some basic information and the development of daily work. In this era of "information explosion ", machine translation is the best choice to improve translation efficiency. The emergence of high-quality translation systems has largely overcome language barriers. On the one hand, it can help government departments to improve their work efficiency, on the other hand, it can promote economic development in minority areas and strengthen national unity.
This paper starts with the research background and significance, selects the appropriate model to modify according to the language characteristics of Uyghur language, and then uses the same data set to analyze the performance of machine translation of different models. The following chapters are arranged as follows: Section 2, the neural network based Uyghur Chinese machine translation. This chapter introduces the principle of neural network machine translation, and then introduces the existing Uyghur Chinese machine translation model and similar language machine translation model. Section 3, the experimental design is explained based on the relevant theoretical information introduced in the previous chapter. Analysis of experimental results. First, the experimental data and experimental environment are explained, and then the translation performance is analyzed, and finally the conclusion is drawn. Section 4 summarizes the research content of this paper. Section 5 shows the references used in the process of writing the article.

2. Research Theory and Methods

2.1. Basic Principles

The main idea of neural machine translation is to directly experiment the automatic translation between natural languages through neural networks[2]. The commonly used encoder-decoder model (encoder-decoder) implements automatic sequence conversion. The encoder encodes the source language (Fig 1), extracts the information, and then converts the information sequence into the target language to complete the translation of the language. Suppose that the given source language is $X = \{X_1, X_2, X_3, \ldots, X_n\}$ Generate map sequence $Z = \{Z_1, Z_2, Z_3, \ldots, Z_n\}$, then decodes the production target language $Y = \{Y_1, Y_2, Y_3, \ldots, Y_n\}$.

![Figure 1. Neural network translation principles](image1)

2.2. Basic Model

Transformer model fully complies with the overall encoder-decoder architecture, but only self-attention[3]. The structure is shown in Figure 2:

![Figure 2. Transformer Schematic diagram](image2)
On the left, the encoder structure consists of six identical layers, each consisting of multi-head attention and Position-wise feed-forward. The decoder is also composed of six identical layers, but compared with the sublayer of the encoder, there is also one more layer to integrate the output of each sublayer of the encoder.

2.2.1. Attention mechanism. The attention mechanism outputs the weighted sum of vector key values and vector mapping query values.

\[
\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{Q K^T}{\sqrt{d_k}} \right) V
\]

(1)

**Figure 3.** Scaled Dot-Product Attention

Figure 3 the left input is composed of query, dimension dk, and dimension dv. Q represents the hidden layer sequence, K the hidden layer sequence of the encoder, v the weight sequence of the hidden layer of the encoder. Its specific calculation:

\[
\text{MultiHead} (Q, K, V) = \text{Concat} \left( \text{head}_1 \text{head}_3 \cdots \text{head}_3 \right) O
\]

(2)

where \text{head}_i = \text{Attention} Q W_i Q W, \text{KVW}, \text{V}

(3)

**Figure 4.** Multi-Head Attention

Multi-Head Attention (Figure 4) allows the model to simultaneously focus on information from different representation subspaces at different locations. When one of them is taken as the main, then the mean processing can suppress this situation.

\[
\text{MultiHead} (Q, K, V) = \text{Concat} (\text{head}_1 \text{head}_3 \cdots \text{head}_3 \text{W}) O
\]

(2)

\[
\text{where \ head}_i = \text{Attention} Q W_i Q W, \text{KVW}, \text{V}
\]

(3)

2.2.2. Position feedforward networks. Each layer of the Transformer encoder and decoder has a fully connected feedforward network consisting of two linear transformations:

\[
\text{FFN} (x) = \max (0, x W_1 + b, 11W)32 + b_2
\]

(4)

2.2.3. Embedding and softmax. Similar to other sequence transduction models, input and output sequences are transformed into vectors of dimensional data models using the learned embedding
method. We also use commonly used linear transformations and softmax functions to convert the decoder output to the predicted next sequence probability. Share the same weight matrix between two embedded and pre-soft layer in the model[4].

3. Research Process and Results

3.1. Experimental Design Description

3.1.1. Structure. The neural machine translation of resource-deficient language is a problem worthy of further study. Can high resource parallel corpus help the research work of low resource? In the face of machine translation, high-quality network models are often implemented through large-scale and annotated data. Because there are enough Japanese-Chinese parallel corpora, but Uygur-Chinese corpus is less, we try to improve the performance of Uygur neural machine translation in Japanese.

3.1.2. Model factors. Self-attention limits connect all positions to operations performed in a fixed number of orders. Simply considering the neighborhood of size r in the input sequence centered on the output position, the computational performance of the long sequence can be improved. Self-attention forces are limited to self-attention, and the computational complexity of each layer is lower. The length of the path between the long-distance dependencies in the network is an important factor in many sequence transformation tasks. The shorter the combination of these paths between any position in the input and output sequences, the easier it is to learn the remote dependencies[5].

Figure 5. Comparison of Translation Performance of Different Models

3.1.3. Design description. A model is trained in a small number of Uygur languages, a model is trained in a large number of Japanese languages, and then a model is trained by mixing Uygur with Japanese. In general, in order to enhance the decoder performance of low resource neural machine translation, the weight of decoder can be shared. However, the decoder corresponds to the Han-dimensional relationship, which lacks control information and can not share the weight of the decoder. Uygur and Japanese have similar word order, so the hidden layer weight of shared encoder can be used to improve the encoder performance of mixed model.

3.2. Data

The corpus contains 50000 Uighur-Chinese sentence pairs and 130000 Japanese-Chinese sentence pairs. Both the Chinese-Japanese and Uygur-Chinese databases come from the natural language processing laboratory of the school and have been annotated. The corpus is directly divided into data sets, training sets and test sets.
Table 1. Statistics of Chinese Material Base

| Data sets     | Language     | Number of sentences |
|---------------|--------------|---------------------|
| Training set  | Uighur       | 46584               |
| Training set  | Chinese      | 46584               |
| Development set | Uighur     | 2081                |
| Development set | Chinese    | 2081                |
| Test set      | Uighur       | 1335                |
| Test set      | Chinese      | 1335                |

Table 2. Statistics of Chinese Library

| Data sets     | Language | Number of sentences |
|---------------|----------|---------------------|
| Training set  | Japanese | 124534              |
| Training set  | Chinese  | 124534              |
| Development set | Japanese | 2613                |
| Development set | Chinese  | 2613                |
| Test set      | Japanese | 2853                |
| Test set      | Chinese  | 2853                |

3.3. Experimental Conditions
Ubuntu16.04, is the operating environment of this experiment Python2.7, Tensorflow1.6. Before the experiment started, The existing corpus needs to be preprocessed, The segmentation of Chinese corpus, Morfessor - type segmentation in Uighur, Limit the length of a sentence to less than 50 words. In the experiment, RNNSearch-50 model was selected as the baseline system. After the experiment, the BLEU was used as the performance index to analyze the translation effect of the mixed model. The number of layers in the model is 6, The maximum possible length of the sequence is 500, Unit size 512, The learning rate is 0.2.

3.4. Performance Analysis

Table 3. Bilingual translation tasks

| Model              | BLEU( Uygur - Chinese) |
|--------------------|-------------------------|
| Baseline system    | 19.95                   |
| Uygur language model | 20.11                  |
| Mixed model        | 21.32                   |

As can be seen from Table 3.4, the BLEU value of the trained mixed model is higher than that of the pure Uygur model system, which indicates that the translation model using Japanese is helpful to the training of the mixed model of Uygur translation. It is proved that resource-rich language is helpful to the research of resource-poor language neural machine translation. Although the experiment shows that the translation ability of this design is improved, the performance improvement is not obvious compared with the baseline system, which may be due to the inappropriate pretreatment of Uygur language materials.

4. Summary of Research
This paper first introduces that under the background of promoting the construction of "Belt and Road" in the world, the communication and communication in Uygur autonomous areas are limited by language barriers, and from the point of view of information technology, This paper probes into the late Uygur written translation and tries to improve the translation performance of Uygur-Chinese.
From the perspective of linguistics, this paper studies Uygur language, and finds that Uygur is similar to Japanese in grammatical structure, has a great relationship with Turkish pronunciation and writing, and only uses Latin to write with Arabic. Based on the analysis of the existing machine translation models, it is found that the former dimension-Chinese translation is mainly based on statistical machine translation (such as instance-based dimension-Chinese machine translation). The research of Uygur language started late, and the available text data and other relevant references are scarce. The determination of experimental ideas is mainly influenced by two factors: the first explored language characteristics and the second available corpus. Although Turkish is the most relevant, the available data have no advantage over Japanese, and machine learning depends mainly on data. Therefore, combined with the existing Chinese-Japanese parallel corpus, the machine learning framework is used to train the translation model. The experiment is expected to optimize the translation model, but the actual effect is not ideal, but the optimization of the machine translation model of neural network is an attempt.

5. References
[1] Yan Li. A Study on Cross-border Language Planning in the Core Area of "Belt and Road" from the Perspective of Language Ecology [D].; and Southwest University, 2018
[2] Minh-Thang Luong. NEURAL MACHINE TRANSLATION [D]. STANFORD UNIVERSITY, 2016
[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. Attention Is All You Need [J]. arXiv: 1706.03762v5
[4] O fri r Press and Lior Wolf. Using the output embedding to improve language models [J]. arXiv preprint arXiv: 1608.05859, 2016.
[5] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, et al. Gradient fluflow in recurrent nets: the difficulty of learning long-term dependencies [Z], 2001.