International Journal of Trend in Research and Development, Volume 10(2), ISSN: 2394-9333
www.ijtrd.com

Machine Translation Model of Chinese Cultural Classics under the Background of BLEU Algorithm-Taking the Text Recognition Framework of the Analects of Confucius as an Example

[1,2]RAO LI
[1]School of Foreign Studies, Nanjing University, Nanjing Jiangsu, China
[2]Jincheng College of Nanjing University of Aeronautics & Astronautics Nanjing Jiangsu, China

***Abstract*—**Based on the similarities between the two, this paper proposes an automatic scoring method for mathematics subjective questions based on the machine translation scoring index BLEU. The paper proposes a semi-supervised neural network translation model based on the sentence-level bilingual evaluation surrogate (BLEU) index to select data. China's The Analects of Confucius translation seminars and regular and irregular academic seminars at various levels and other representative related academic activities, to find an effective paradigm for the English translation and dissemination of The Analects of Confucius, in order to provide the current and future foreign dissemination of Chinese cultural classics. Sexual reference. Use statistical machine translation and neural machine translation models to generate candidate translations for unlabeled data, respectively, and then select monolingual translation candidates through sentence-level BLEU metrics.

Keywords*—Machine Translation Model, Chinese Cultural Classics, BLEU Algorithm, Text Recognition Framework, the Analects of Confucius*

I. INTRODUCTION

In the teaching of primary and secondary schools, examination is an important means for teachers to understand the degree of students' mastery of classroom [1] knowledge, among which subjective questions are compulsory. However, due to the large number of examinations and the high frequency of examinations under normal circumstances, teachers' workload of marking papers is very large [2], and the status of teachers can easily affect the marking results. The end-to-end NMT adopts a pure neural network model, and very little feature engineering greatly simplifies the model complexity of machine translation [3]. Less linguistic feature design also makes more and more researchers join the research of machine translation. All kinds of academic activities and academic research have different themes [4], but almost all kinds of themes originate from various translations. Therefore, this paper firstly examines the attention to several different translations in academic activities (from the perspective of translators) from the CNKI database platform [5].

On the one hand, the development of NMT benefits from the introduction of the Encoder-Decoder framework in the literature and the [6] attention mechanism in the literature. There are also some works that try to explore the implicit learning ability of the neural machine translation model and stimulate its use in the neural machine translation model [7]. potential for related tasks. In the task of word alignment (Word Alignment), Li et al. tried to use the attention learned by the neural machine translation [8] itself as a supervision signal. And combined with the Transformer model to train the Mongolian-Chinese neural machine translation model to prove that the translation model trained by [9] the unregistered word processing method based on semantic similarity has the best effect. In order to alleviate the problem of limited vocabulary mentioned above [10].

The term machine translation was first proposed by Warren Weaver, the pioneer of machine translation technology in the United States [11], in 1949, which made the concept of machine translation receive widespread attention. In the next decade, the popularity of machine translation continued to rise [12]. The United States, the Soviet Union and some European countries generally attached great importance to machine translation, and set off a wave of machine translation for a period of time. The model uses a convolutional neural network [13] (Convolution Neural Network, CNN) to encode the provided source language into a continuous vector, and then uses a recurrent neural network to decode the vector and convert it into the target language [14].

It solves the long-distance reordering problem in statistical machine translation [15], and lays the foundation for the subsequent pure neural network for machine translation. The Analects of Confucius is a classic of Chinese cultural classics, a treasure of Chinese culture, and is included in China's "Four Books". The Analects of Confucius contains rich traditional Confucian theories [16]. In the cultural exchange between China and the West, the influence and value of The Analects of Confucius is immeasurable. In the centuries since the 16th century, the English translation of The Analects of Confucius has never stopped [17], and there have been more than 70 English translations so far. The main idea of subjective questions is to calculate the similarity between the student's answer and the standard answer. The existing calculation method of text similarity is to divide the text into short text and long text [18].

Phrase-based statistical machine translation system mainly adopts log-linear model, and uses development set to train based on minimum error rate for multiple feature functions. and foreign language journals [19]. The current mainstream neural machine translation adopts an end-to-end (end-to-end) training method based on the "encoder-decoder" framework, which can be implemented by a variety of model architectures, including recurrent neural networks [20]. To construct a phrase translation table, phrase extraction should be performed first, and the parallel sentence pairs that have been word-aligned obtained in the previous step are extracted from two directions [21]. Parallel sentences in source language - target language. The dimension of the word embedding represented by the one-hot encoding is determined by the length of the dictionary [22]. In other words, the dimension of the word embedding is the

same as the length of the dictionary. The 4th word in the dictionary, the 4th dimension represented by the word embedding has a value of 1, and the rest of the dimensions have a value of 0 [23].

II. The Proposed Methodology

A. BLEU Algorithm

It has three main characteristics: first, the encoder-decoder architecture can handle both input and output sequences of variable length, so it is suitable for sequence conversion problems, such as machine translation [24]; second, the encoder uses variable-length sequences. The sequence is taken as input, and it is transformed into a state with a fixed shape. In the centuries after this [25], the translation and introduction of The Analects of Confucius has been continuously developed. Especially in the 20th and 21st centuries, the translation and introduction of The Analects of Confucius has received unprecedented attention, and the English translation has reached an unprecedented grand occasion. Mathematical subjective questions are a special kind of text, whose length is between long text and short text, with many mathematical symbols, relatively monotonous sentences, and less text ambiguity. It is necessary to formulate targeted text similarity calculation according to the specific characteristics of the question method.

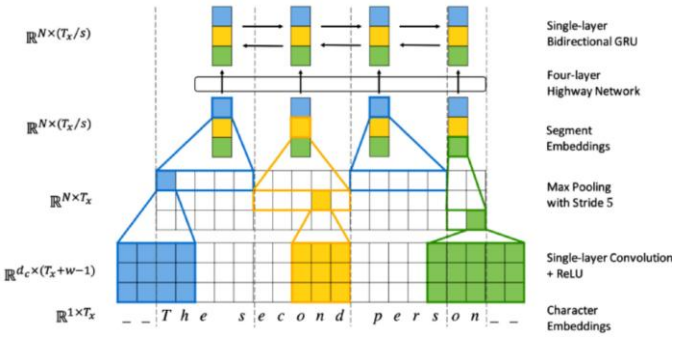

Fig. 1. BLEU algorithm

The model parameters can only be trained by using the expensive and accurate bilingual training corpus, and it is difficult to effectively utilize a large number of easy-to-obtain monolingual corpora. Therefore, this paper studies an effective scheme to use monolingual corpus to improve system performance. To sum up, there is no doubt that the English translation of The Analects of Confucius has developed vigorously in recent years and its achievements, but the deficiencies reflected in the attention of academic activities are also obvious. First, research topics tend to be methodological, such as cognitive translation of various languages, and there is insufficient research on the meaning of the ontology of The Analects of Confucius. Different from these related works, the method proposed in this chapter utilizes the diverse structural factors within the model, which is lighter and more explanatory, and the subsequent experiments also prove that the effect is better. Because the multi-head attention mechanism has the potential to select translation candidates.

B. Machine Translation Model of Chinese Cultural Classics

Then, the corpus division and its preprocessing scheme used in the experiment are introduced. Then, the induction process of the Mongolian-Chinese bilingual dictionary based on the machine translation model is described. Based on the bilingual dictionary, combined with the language model and the back-translation method, an unsupervised Mongolian-Chinese neural machine translation system is constructed. It is used for the training of the bilingual dictionary and the training of the machine translation model; then, for the training of the word embedding, this chapter uses the self-learning method to align the two languages to form a Uyghur-Chinese bilingual dictionary. Attention mechanism originally a strategy used in computer image vision, it was gradually used in the field of natural language processing and really flourished. The Transformer proposed by Google researchers brought out the practical effect of the attention mechanism, and then researchers began to focus on it.

$$L(\theta) = \sum_{i=1}^{N} \log P(y_i) \tag{1}$$

$$D(G_o^t, G_i^g) = \sqrt{\sum_{i=1}^{n} (T_k - T_i)^2} \tag{2}$$

$$q(x) = \exp\{b_a^{0,g} + b_a^{1,g} k_1^g\} \tag{3}$$

Reception theory organically links the reader's appreciation with the author's creation, and positions the relationship between the two as interrelated, conditional, and influencing each other. The main points of acceptance theory include: (1) The social significance and aesthetic value of literary works need to be realized through reading. Realize the automatic marking of mathematics subjective questions. The experimental results of the ten-fold cross-validation method show that the BLEU automatic scoring results are completely equal to the manual scoring results, accounting for 49.5%, and the deviation score of 1 accounts for 39.6%. It is a language model, through the fake source, the sentence and the corresponding target monolingual corpus form a corpus pair, and the model parameters are updated. Scheme 2 mainly translates the target sentence back into the source sentence. The profound meaning of the content and value of The Analects of Confucius, such as social and political philosophy, humanistic and ethical values, and educational ideology, have not been translated to a certain height and sufficiency. development, especially in the traditional humanities.

Important heads have many interpretability pathways, including linking to adjacent words, tracking specific syntactic relations, etc. Michel et al. also found a similar phenomenon. The removal of some attention heads has little effect on the performance of the model. GAN (BLI) + DAE + NMT represents the Mongolian-Chinese bilingual dictionary induction method based on adversarial learning (BLI), combined with DAE-based language. Model training and back-translation method. After preprocessing and training on the Uyghur-Chinese weakly parallel corpus, Uyghur word embedding representation and Chinese word embedding representation are obtained respectively.

C. The Analects of Confucius Text Recognition Framework

At this time, the row of Uyghur-Chinese word embedding represents the column representation of Uyghur-Chinese word embedding There is no logical correspondence between and. The output of the decoder part is a vector represented by floating-point numbers, which is projected into a vector of the size of the target language vocabulary through a linear fully connected network layer. In this chapter, we use a subword encoding vocabulary of size 40,000 dimensions. Each element in the projected vector uniquely represents the score of a word. The effect of expectation horizon on target language readers. The reader's participation in the text makes a literary work truly meaningful.

$$PE=\sin(pos/10000) \quad (4)$$

$$FFN(x)=\max(0,xw_1+b) \quad (5)$$

$$M_{t+1}=S_t^*D_t \quad (6)$$

It is only an ideal state that the "aesthetic scale" and "directional expectation" formed by readers before reading the work are fully realized in the reading of the work. The automatic scoring of subjective mathematics questions can be attributed to the calculation of the similarity between the standard answer and the student's answer. If the text of the student's answer and the text of the standard answer of the test question are the same, the similarity is 1; if the text of the test taker's answer and the text of the standard answer of the test question are completely different. The automatic scoring of mathematics subjective questions can be attributed to the calculation of the similarity between the standard answer and the student's answer question.

If the text of the student's answer and the standard answer of the test question are the same, the similarity is 1; if the text of the examinee's answer and the text of the standard answer of the test question are completely different. Overseas dissemination, research in this category itself is insufficient, such as the characteristics of overseas translators, translation ideas and culture (2) The research on "domestic translation" is not fully integrated with "overseas communication". Due to the high subjectivity of human dialogue, it is difficult to evaluate automatically. Therefore, in addition to the inter-sentence BLEU and the reference translation BLEU in the above sections, we also evaluate the quality of the generated dialogues through several human-evaluated indicators.

III. EXPERIMENT

The text recognition framework of "The Analects of Confucius" is shown in the figure.

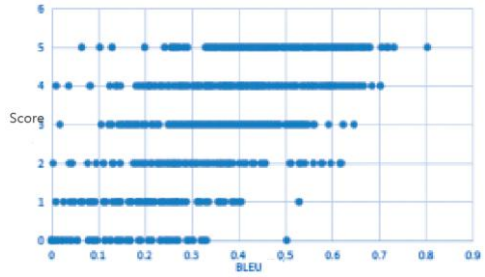

Fig.2. The Analects of Confucius Text Recognition Framework

The machine translation model of Chinese cultural classics is shown in the figure.

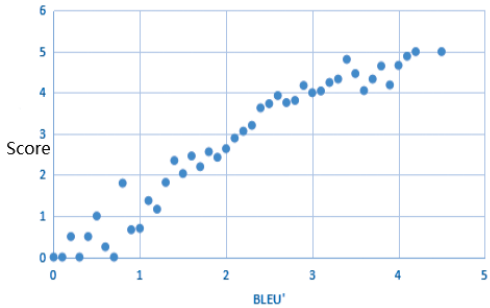

Fig. 3. Machine translation model of Chinese cultural classics

The BLEU algorithm is shown in the figure.

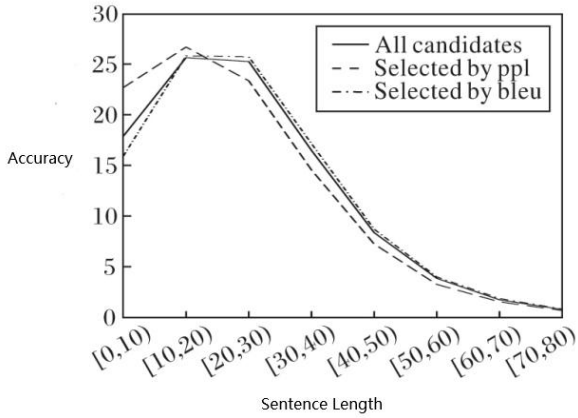

Fig. 4. BLEU algorithm

CONCLUSION

Based on sentence-level BLEU, the candidate translations with complementary features are selected for higher quality translations and integrated into the semi-supervised scheme of the refined training set. Experiments show that the algorithm in this paper can effectively use monolingual corpus. The sign and way of Chinese culture finally going to the world must be the translation and introduction of classics and their overseas dissemination. The translation and introduction of Chinese classics marked by the banner of the English translation of The Analects of Confucius still has a long way to go. On the task of document translation, this paper proposes large-scale new datasets and fine-grained evaluation indicators for the task itself, and uses the ability of neural machine translation to model long-distance texts.

Acknowledgement

The study was supported by 2021 of School-level Education and Teaching Reform of Jincheng College of Nanjing University of Aeronautics and Astronautics" Research on the Blended Teaching Practice of Listening and Speaking Course with Ideological and Political Education Construction in Private Colleges and Universities "(2021-Y-18).

References

[1] Zhang Tianfei. Investigation and Research on Academic Activities of "The Analects of Confucius" under the Background of Translation and Introduction of Domestic Classics [J]. Overseas English, 2018(10):2.

[2] Zhang Xuelin. Comparative Analysis of Russian Translations of The Analects from the Perspective of Pragmatic Equivalence: Taking Popov, Semyonenke and Belleromov's Translations as Examples [J]. Journal of Heihe University, 2022, 13(2): 4.

[3] Cai Xinjie, Wen Bing. Statistical Analysis of Error Types in Machine Translation from Chinese to English: An Example of Chinese-English Translation of Foreign Announcement Texts [J]. Journal of Zhejiang Sci-Tech University: Social Science Edition, 2021, 46(2):8.

[4] Mo Chang, Zhang Mengdi, Wang Ruibing, et al. The performance of machine translation in literary texts: Taking the first lesson of advanced English "Middle East Market" as an example [J]. Research on Communication Power, 2019(33).

[5] Ansuyala, Wang Siri Gulen. Research on the translation of Chinese and Mongolian institution names based on transformer neural network [J]. Journal of Chinese Information, 2020, 34(1):5.

[6] Hou Weili. Design of automatic error checking system for translation robot English text [J]. Automation and Instrumentation, 2022(4):5.

[7] Zhao Huijun, Lin Guobin. Research on Corpus Context Strategy for Missing Words in Machine Translation [J]. Foreign Language Teaching and Research, 2022, 54(2):12.

[8] Cui Zihan. The quality comparison of machine translation translations—taking Google Translate and DeepL as examples [J]. 2021.

[9] Toudan Cairang, Renqing Dongzhu, Nima Zhaxi, et al. Research on Chinese-Tibetan Machine Translation Based on Improved Byte Pair Coding [J]. 2021.

[10] Huo Huan, Wang Zhongmeng. A Reading Comprehension Model Based on Deep Hierarchical Features [J]. 2018.

[11] Huang Yanhua, Chen Ping. Cultural Presuppositions and Translation of Chinese Classics--Taking Yang Daiying's Translation of "A Dream of Red Mansions" as an example [J]. Jingu Wenchuang, 2022(17):3.
[12] Liu Ding. Research on Chinese word and sentence segmentation technology and machine translation evaluation method [J]. 2004.
[13] Lai Wen. Research on key technologies of low-resource language neural machine translation.
[14] Cai Hua. Column: "Translators are promising".
[15] Liu Yupeng, Ma Chunguang, Zhang Yanan. A Deep Recursive Hierarchical Machine Translation Model [J]. Chinese Journal of Computers, 2017, 40(4):11.
[16] Li Boze. Research on machine translation modeling method based on deep learning [D]. Xi'an University of Science and Technology, 2019.
[17] Ye Shaolin, Guo Wu. Semi-supervised neural machine translation based on sentence-level BLEU index selection data [J]. Pattern Recognition and Artificial Intelligence, 2017, 030(010):937-942.
[18] Zhang Xiyuan. Research on Statistical Machine Translation Sequencing Model [D]. Xi'an University of Technology, 2015.
[19] Liu Jiqiang, Zhang Ruiqing, He Zhongjun, et al. Machine translation model acquisition and text translation method, device and storage medium: CN111859994A[P]. 2020.
[20] Jiang Jialiang, Li Xiang, Cui Jianwei. Training methods, devices and systems for machine translation models: CN110941966A[P]. 2020.
[21] Mei Yangchun. The text construction strategy of Chinese sci-tech classics from the perspective of Western readers' expectations [J]. Journal of Xi'an International Studies University, 2018, 26(3):5.
[22] Chen Lijiang. Research on the teaching method of improving Chinese-English statistical machine translation model [D]. Nanjing Normal University.
[23] Tao Yuanyuan, Tao Dan. Research on Machine Translation Algorithms Based on DNN and Rule Learning [J]. Computer Measurement and Control, 2021.
[24] Wang Hongli. Optimization of machine translation algorithm for complex long sentences in English under semantic relation [J]. Mechanical Design and Manufacturing Engineering, 2020, 49(12):3.
[25] Su Jinsong, Dong Huailin, Chen Yidong, et al. Introducing a statistical machine translation model based on topic retelling knowledge [J]. Journal of Zhejiang University (Engineering Science Edition), 2014, 048(010):1843-1849.