Deep Learning Model-Based Machine Learning for Chinese and Japanese Translation

Qiang Wang¹ and Hongfeng Wang²

¹School of Foreign Languages, Dezhou University, Dezhou 253023, China
²College of Computer and Information, Dezhou University, Dezhou 253023, China

Correspondence should be addressed to Qiang Wang; wangqiang@dzu.edu.cn

Received 14 January 2022; Revised 8 February 2022; Accepted 21 February 2022; Published 9 March 2022

Academic Editor: Mohammad Farukh Hashmi

Copyright © 2022 Qiang Wang and Hongfeng Wang. 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 paper takes deep learning as the background of researcher design, combines the relevant cutting-edge research results in recent years, addresses the linguistic characteristics of Japanese and the problems faced by completing Japanese machine translation system, and determines the neural network structure of encoding-decoding for Japanese translation based on the characteristics of high similarity between Japanese and Chinese and after referring to the neural network architecture of English translation, and the basic structure and the corresponding improvement of the hidden layer unit calculation are carried out. The training model is optimized and an integrated Japanese machine translation system is implemented. Finally, the translation models of Japanese and Chinese intertranslation and Japanese and Chinese intertranslation are tested, and the optimal model fusion achieves a BLEU value of 39.52.

1. Introduction

With the rapid development of globalization, the exchange of information in different countries and languages is becoming more and more frequent, and the global translation market scale is steadily growing. According to statistics, the global translation industry market scale reached $37.19 billion in 2014, an increase of 6.23% over the same period of the previous year, and the global translation industry scale is expected to reach about $53 billion by 2020 [1]. More and more technology companies have started to step into this huge market through machine translation. Machine translation (MT) is a science that uses computers to translate one language into another [2] to solve the communication barriers between different languages. With the rise of neural network machine translation in the past two years, the quality of machine translation has been significantly improved, and many technology companies such as Google, Baidu, Youdao, Tencent, and Sogou have carried out the research and development of machine translation products, and some real-time machine translation systems have started to be applied in sports, tourism and other fields [3, 4].

These homographs have become an important part of the vocabulary of both countries. There are many homonyms in the same kanji (homomorphic synonyms), but there are also many words with different meanings and usages (homomorphic heteronyms). For Chinese learners of Japanese and Japanese learners of Chinese [5], especially problems such as misuse and mistranslation of homonyms often arise. Therefore, I believe that language education in Japan and China must pay attention to the study of homonyms and dissimilar words in China and Japan [6].

In recent years, with the development of computer technology, a large amount of linguistic data can be used as an electronic corpus, and objective analysis of word and sentence usage has become possible, and lexical research has developed...
by leaps and bounds [7]. At the same time, corpus-based studies of Japanese vocabulary have become increasingly prevalent, and as part of this, studies of Japanese-Chinese homonyms are also underway, from a variety of perspectives including linguistics and education [8].

On the one hand, Japan and China are neighbors who share a close relationship in terms of technology and economy. With the need for information exchange between Japan and China, research on machine translation is flourishing as an alternative to translation methods for which the demand for translation has increased. However, using the free Japanese-Chinese machine translation software now available on the Internet, many mistranslations can still be seen when translating Sino-Japanese homonyms into Chinese. Therefore, I believe that the translation processing of Japanese homonyms in Japanese-Chinese machine translation is one of the important research topics [9, 10].

In addition, from the perspective of machine translation, the Japanese-Chinese machine translation method of “break,” “disseminate,” and “take” and its machine translation rules are proposed, and the Japanese-Chinese machine translation system is used to conduct The translation experiments are conducted using the Japan-China machine translation system, and the translation results and problematic points are analyzed [11].

2. Related Work

This section focuses on the current research status of Japanese-Chinese machine translation in Japan and abroad. In Japan, research on machine translation was started in 1955 by Kyushu University and the Electrical Testing Laboratory. Since 1987, the research on Japan-China machine translation has become popular in Japan through the joint research of “Research Cooperation on Machine Translation Systems in Neighboring Countries,” which was established with the assistance of the Japanese government development aid. The research on Japanese-Chinese machine translation at Japanese universities, research institutes, and companies has been flourishing. [12] developed a Japanese-Chinese machine translation system based on the decomposition of text using a family model. Experiments confirmed the effectiveness of this proposed Chinese-Japanese machine translation technique. A pilot pattern conversion type machine translation system called jaw was developed by [13], which is a machine translation engine from Japanese to multilingual, and jaw/Chinese is a Japanese-Chinese machine translation system with Chinese as the target language. [14] describes a Chinese-Japanese machine translation system developed with the main purpose of supporting companies entering China for business activities such as information collection and delivery. It is now working to strengthen the performance geared toward commercialization through experiments with translation services on the Web. The research project to develop a practical Japanese-Chinese mechanical translation system is introduced, targeting Japanese-Chinese scientific and technical documents. As a translation method, use-case-based translation that takes into account the structure of the language more deeply is used [14]. A Japanese-Chinese machine translation system for travel conversations was developed [15]. The system is roughly divided into the morphological element parsing section, parsing section, grammar generation section, and morphological element generation section. In addition, it was installed on a commercially available PDA terminal with Windows mobile action and evaluated using a cost example text of about 1200 texts. Experiments of the Japanese-Chinese machine translation system were conducted. That system consisted of a Japanese dictionary, a Japanese-Chinese dictionary, a Chinese dictionary, and a rule system, using a language implemented in a computer made in the United States. In the experiment, the system tried to translate 102 sentences that were used as experimental materials and 10 sentences that the author made at random within a certain range and got an accuracy rate of 90% [16]. For the Japanese-Chinese machine translation, the similarity was calculated using the method of concept classification and subsidiary language [17]. The method fully reflects the conjugation of word-word combinations in the text, can appropriately reduce the example-based scale, can simplify the calculation of values, and shows good results.

3. Cross-Reference of Word Meaning Items in the Dictionary

This subsection examines the Japanese word meaning of “break” in Japanese and “break” in Chinese dictionaries, and the author consulted five dictionaries, namely, the Mei-kei Kokugo Dictionaries, the Guang Jiyuan (6th edition), the Shin Meikei Jyutsu Kokugo Dictionaries, the Nihon Kokugo Daidai Dictionaries, and the New Japanese-Chinese Dictionaries. The five dictionaries are listed below.

In the above dictionaries, the interpretation of “break” is generally consistent. Unlike the “Hiroshiken (6th edition),” [18] other Japanese dictionaries have two meanings of “break.” In the case of Chinese, I have consulted five dictionaries: Modern Chinese Dictionary (5th edition), Modern Chinese Dictionary Hai, Modern Chinese Dictionary, Chinese Dictionary, and Crown Chinese and Japanese Dictionary, and called it “break” [18].

In the abovementioned dictionaries, the number of lexical items related to the Chinese word “break” differs considerably. The Modern Chinese Dictionary (fifth edition) only expresses one meaning of “break,” while the Modern Chinese Dictionary of the Sea divides “break” into six meaning items in detail. In fact, the meaning of “break” in Modern Chinese Dictionary (Hai) is ① “break the standard, order or state”, ② “exceed”, and ③ “break the limit or bondage.” Break the limits or “fetters” and “break” in the Modern Chinese Dictionary, ② “break the original limits or conventions,” ⑦ “to bring to an abrupt end” and “to fail to follow the cycle or comply with” are combined into one meaning item, which is equivalent to the meaning item ② of “break” in the Modern Chinese Dictionary (fifth edition) break the original limitation, agreement, etc. In addition, the meaning of “break” in the Modern Chinese Dictionary (Hai) and the meaning of “break” in the Dictionary of Chinese (Da Dian) are not found in other Chinese dictionaries.
4. Neural Network Structure for Japanese Translation

Neural network translation is different from traditional statistical machine translation, which accomplishes the translation task through multiple modules, while neural machine translation uses the encoder-decoder architecture. The encoder-decoder architecture was first proposed by Oxford University in 2013 [19]. The whole translation process of this architecture is divided into two phases: the encoder part and the decoder part. The encoder step is the process of encoding the source language. The decoder step decodes the target language based on the results of the encoding. The whole system inputs a large parallel corpus, and the training process is performed by the model to save the appropriate encoding and decoding parameters to translate the source language into the target language.

4.1. Basic Network Structure for Japanese Translation. The encoder-decoder model, as the name implies, is the encoder-decoder model. The main point of this model is that the encoder converts the input language sequence into an intermediate vector of fixed length, and then, the decoder converts this generated fixed vector into the output language sequence [20]. Figure 1 shows the overall encoder-decoder model structure for translating Japanese into English. It can be seen that when we translate English to Japanese, English is used as the input language sequence, and the fixed vector C is obtained by the encoder, and the output Japanese sequence is obtained by the decoder based on the fixed vector C.

Among them, neural networks such as CNN, RNN, GRU, and LSTM can be used for the specific implementation in the encoder and decoder parts, respectively. For the specific experiments of Japanese translation, we use BIRNN in the encoder part and GRU in the decoder part.

4.2. Improvement of the Underlying Network Structure. The network structure of coding and decoding will have some problems that cannot be solved even after the improvement of the design warp elements, because these problems are caused by the structural characteristics of this model. In the code-decode structure, the encoding part encrypts the input sequence, that is, encodes it, and outputs a fixed-length intermediate vector, which contains the information of the whole input sequence. Then, the decoding part then decrypts through this fixed vector and outputs the target.

In order to finally get a neural network translation model with good translation performance, we need to load the bilingual material from the training set into the model so that the model learns the appropriate parameters. However, because the length of the parallel utterances in the training is not unique, and there will be sequence information of the words in the utterances, in order to get a robust translation system, a recurrent neural network (RNN) is used, whose structure expansion diagram is shown in Figure 2:

According to the expansion diagram, it can be seen that the RNN has a memory function, and its output depends not only on the current input but also on the current memory. This memory is sequence-to-sequence, which means that the current moment is influenced by the previous moment. This feature is very applicable to the sequence of language, because no matter it is Japanese, English, or Chinese, the meaning of the word in the sentence will depend on the context, for example, banks in English can be interpreted as river bank or bank; then, the specific translation in the context as which meaning will be judged according to the context [21].

The reason why the encoder part uses BIRNN is that the difference is that RNN reads sequence information from \( x_0 \) to \( x_n \), while BIRNN not only reads sequence information from front to back but also reads information from back to front, that is, from \( x_n \) to \( x_0 \), so as to better ensure the integrity of the input sequence information. The simple structure of the RNN has two main problems: one is the long distance dependence problem, that is, the RNN will not learn the information because of the long distance when facing the predicted word and the word at the farther position above; the second is the gradient dispersion problem, because the RNN model is trained using the back propagation algorithm. The gradient explosion or gradient disappearance problem will occur in the process of error backpropagation, and the gradient explosion will make the parameter update of model learning cause oscillation, while the gradient disappearance will make the learning very slow and lead to the final learning ineffective. Therefore GRU is used to improve these two problems [22].

In short, what are the similarities and differences in the meanings of "break" in Japanese and "break" in Chinese. I think there is a corpus of verification.

Figure 1: Encoder-decoder model structure.

![Figure 1: Encoder-decoder model structure.](image)

![Figure 2: RWJ structure expansion diagram.](image)
language sequence. In order to get a more accurate translation, the dimension of the intermediate vector obtained by the encoding part should be as large as possible in order to contain more information about the content of the source sentence and compress longer sentences.

Therefore, in the encoder-decoder model, the intermediate generated fixed-length vector \( C \) is improved. The improvement is that when the decoder generates a single target language word, the word is independent of the majority of source language words and is relevant to only a portion of the source language words. Therefore, the context vector \( c \) at the source language end is dynamically generated for each current target language word generated, and the fixed-length vector representing the entire source language sentence is not used. In this way, the translation effect is improved. The improved model structure is shown in Figure 3.

Still, when translating English Iamastudent to Japanese Sayapelajar, when parsing the Japanese word Saya, it is found that only the two words lam in English have strong correlation, and this time after learning of the neural network, a vector is generated according to the different weight values of the four words in English, and then, when calculating the Japanese pelajar, it only has a strong correlation with the English word astudent, so that after learning of the neural network, a vector is generated.

The specific dynamic vector \( c \) is calculated as follows:

\[
c_i = \sum_{j=1}^{T_x} a_{ij} h_j,
\]

(1)

where \( T_x \) represents the length of the input sentence, i.e., the length of the English sentence, plus the stop character, \( T_x \) here is 5.

\( a_{ij} \) are the weight values, which are continuously optimized during the learning process of the neural network.

So, the \( c_i = g(a_{i1} \ast h_1, a_{i2} \ast h_2, a_{i3} \ast h_3, a_{i4} \ast h_4, a_{i5} \ast h_5) \) function is the weighted summation.

Therefore, each word of the input sequence has a different weight value for each output target word, so that the dynamically generated \( c_1, c_2 \cdots c_n \) can reduce the dimensionality of the fixed-length vector and further improve the translation effect.

4.3. Improvement of Decoder Hidden Layer Unit. We adopt the encoder-decoder neural network structure to conduct experiments on Japanese machine translation. After the source and target language texts are pretrained with word vectors, the encoder part reads the input source language sequence from front to back and from back to front and summarizes the input information, that is, to obtain the encoder. The output of the encoder is obtained.

Figure 4 shows the specific structure of the encoder-decoder’s hidden layer unit.

In the Japanese translation experiment in the calculation of decoder decoding, an improvement is made to the hidden layer unit in it: that is, the GRU module is composed of the prehidden layer state \( S_{t-1} \), all the source annotations (encoder output of all the input hidden layer states), and the result of the previous decoding together to update the current hidden layer state [3].

The GRU module now consists of 3 parts: two GRU units and an attention mechanism between them, which is computed three times to obtain the implicit state at the current moment.

For the first time, the first GRU unit, which is the result of the previous decoding and the previous hidden state \( S_{t-1} \) in the figure above, is used to generate an intermediate representation.

For the second time, the result of the first computation and the context combined with \( C \) to compute the context vector.

For the third time, the second GRU unit generates the current hidden layer state from the results of the first two calculations. The last hidden layer value of the RNN in der.

5. Structural Design of Japanese Translation

Based on the above analysis, an integrated translation system can be designed, which is based on file configuration and can complete the whole translation system, including the processes of data preprocessing, word vector pretraining, model training, and model testing, and at the end of the final training, the \( n \) models with the best performance can be saved, and the \( n \) values can be set manually in advance, and finally, these models can be fused and then evaluated.

Such a system structure will alleviate the labor cost, and the whole training test result is very simple, only need to
start loading our organized bilingual material directory, and after running the script, no longer need too much labor cost to test the BLEU value of each model in real time, and finally also can get the result.

The architecture of the Japanese machine translation system is shown in Figure 5.

Before the training starts, the parallel training corpus is loaded, and the model first pretrains the word vectors of the parallel corpus to get the word vectors of the source and target languages, respectively, and then, all the pairs in the form of these word vectors are input to the model for training. When all the corpus is trained once, it is called a round of epoch. At this time, in order to ensure that the model can learn the language features better, the corpus needs to be reordered, that is, shuffle operation. To ensure that the model can learn the language features better, we need to reorder the corpus, that is, shuffle, to disrupt the grouping order of the training corpus, and then proceed to the next epoch [23, 24].

Before the training starts, a model saving frequency saVe_freq is set, and during the training process, after one iteration of saving frequency, the error value is checked to determine whether it is within the set range, and if it is, the model is saved, and then, the model is tested by BLEU.

Training cannot be done all the time, when our training iterations or the number of training epochs, exceeds the set threshold, or the error value accumulates to the preset threshold; then, our training will be over.

6. Experimental Results

The Japanese research sector is also very concerned about the research trends in China and hopes to obtain valuable
As important science and technology information institutions in China and Japan, CITIC and JST share the common goal of providing foreign science and technology trends to researchers and decision makers in their respective countries, which is the basis for the cooperation project on machine translation in science and technology.

As shown in Figure 6, Chinese and Japanese machine translation developers have natural advantages in syntactic analysis technology of their respective languages. The Japanese side is the international leader in example-based machine translation engine and Japanese lexical and syntactic analyses, while the Chinese side has made great progress in statistical machine translation in recent years, and the Institute of Automation of Chinese Academy of Sciences, which participates in this project, has ranked top in IWSLT and NIST machine translation evaluation, and Harbin Institute of Technology has a deep accumulation in Chinese syntactic analysis research. The cooperation between the two sides can make full use of each other’s existing technology to reduce their own input and improve their own service level.

As shown in Figure 7, the different modeling methods section is the focus of the cooperative research, covering various modules of the machine translation system, including lexical analysis, syntactic analysis, and translation engine research for Chinese and Japanese. In the specific implementation process, the CITIC Institute and JST are responsible for the coordination between China and Japan, respectively, and the specific work content of each collaborative unit is divided, and the work progress is monitored regularly. In addition to email exchanges, annual exchange visits are arranged for researchers from both countries to meet face-to-face, so that they can better understand each other’s research work and discuss solutions to projects and difficulties on the one hand, and enhance the friendship between project members on the other.

The example-based machine translation engine reached the world leading level in Chinese-Japanese translation at the end of the project; the Chinese side’s research on Chinese syntactic analysis in science and technology also adopted the most advanced integrated model of word division and lexical annotation and syntactic analyses, as shown in Figure 8; the accuracy rate of syntactic analysis increased from 80% to 90%; the Chinese side’s research on translation engine proposed various innovative ideas on how to make full use of the syntactic knowledge of source and target languages, how to effectively introduce the idea of example machine translation, and how to use semantic information, respectively. In terms of translation engine research, the Chinese side has proposed various innovative ideas on how to make full use of the syntactic knowledge of source and target languages, how to effectively introduce the idea of example machine translation and how to use semantic information, and combined with the application of deep learning in the machine translation engine, the translation effect of the machine translation engine has been greatly improved.
7. Conclusions

Cross-language research like machine translation is suitable to be carried out through international cooperation, especially the science and technology-oriented machine translation is more inclined to the nature of public services, and the cooperation between the two countries can, on the one hand, reduce the investment of individual countries in resource construction, and on the other hand, make full use of the expertise of both sides in their own language processing to complement each other’s strengths, accelerate the practicalization in science and technology information services, and enhance the science and technology and cultural exchanges between the two countries.

The Sino-Japanese cooperation project on machine translation has been designed and negotiated in detail from the beginning in terms of cooperation objectives, cooperation mode, cooperation contents, intellectual property rights, etc., which has laid a good foundation for the smooth development of cooperative research. During the project, both sides have been adhering to the principle of equality and mutual benefit, and researchers from both sides have built up a deep friendship during the cooperation process. The cooperation model, the results of the project, and the problems and solutions encountered in the process of cooperation have provided valuable experience for more transnational collaborative research on machine translation.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

[1] K. Xiong, R. G. Verdonschot, and K. Tamaoka, “The time course of brain activity in reading identical cognates: an ERP study of Chinese-Japanese bilinguals,” *Journal of Neurolinguistics*, vol. 55, p. 100911, 2020.

[2] E. A. Burlingame, M. McDonnell, G. F. Schau et al., “SHIFT: speedy histological-to-immunofluorescent translation of a tumor signature enabled by deep learning,” *Scientific Reports*, vol. 10, no. 1, article 17507, pp. 1–14, 2020.

[3] T. Ananthanarayana, P. Srivastava, A. Chintha et al., “Deep learning methods for sign language translation,” *ACM Transactions on Accessible Computing (TACCESS)*, vol. 14, no. 4, pp. 1–30, 2021.

[4] A. Eriguchi, K. Hashimoto, and Y. Tsuruoka, “Incorporating source-side phrase structures into neural machine translation,” *Computational Linguistics*, vol. 45, no. 2, pp. 267–292, 2019.

[5] Y. Liu and J. Zhang, “Deep learning in machine translation,” in *Deep Learning in Natural Language Processing*, pp. 147–183, Springer, Singapore, 2018.

[6] W. Yang and Y. Lepage, “Improving automatic Chinese–Japanese patent translation using bilingual term extraction,” *IEEE Transactions on Electrical and Electronic Engineering*, vol. 13, no. 1, pp. 117–125.

[7] S. He, “Research of a sign language translation system based on deep learning,” in 2019 *International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*, pp. 392–396, Dublin, Ireland, 2019.

[8] M. R. Costa-Jussà, “From feature to paradigm: deep learning in machine translation,” *Journal of Artificial Intelligence Research*, vol. 61, pp. 947–974, 2018.

[9] C. Wei, J. Zhang, X. Yuan, Z. He, G. Liu, and J. Wu, “NeuroTIS: Enhancing the prediction of translation initiation sites in mRNA sequences via a hybrid dependency network and deep learning framework,” *Knowledge-Based Systems*, vol. 212, article 106459, 2021.

[10] X. I. E. Tao, Z. H. A. N. G. Chunjiong, and X. U. Yongjian, “Collaborative parameter update based on average variance reduction of historical gradients,” *Journal of Electronics and Information Technology*, vol. 43, no. 4, pp. 956–964, 2021.

[11] C. H. Liu, C. C. Silva, L. Wang, and A. Way, “Pivot machine translation using Chinese as pivot language,” in *China Workshop on Machine Translation*, pp. 74–85, Springer, Singapore, 2018.

[12] C. H. Cao, Y. N. Tang, D. Y. Huang, W. M. Gan, and C. Zhang, “IIBE: an improved identity-based encryption algorithm for WSN security,” *Networks*, vol. 2021, article 8527068, pp. 1–8, 2021.

[13] D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu, and Y. Xu, “Forest fire recognition based on feature extraction from multi-view images,” *Traitement du Signal*, vol. 38, no. 3, pp. 775–783, 2021.

[14] C. Wei, J. Zhang, and Y. Xiguo, “DeepTIS: improved translation initiation site prediction in genomic sequence via a two-stage deep learning model,” *Digital Signal Processing*, vol. 117, article 103202, 2021.

[15] T. Ranasinghe, R. Mitkov, C. Orásan, and R. C. Quintana, “Semantic textual similarity based on deep learning: can it improve matching and retrieval for translation memory tools?,” in *Corpora in Translation and Contrastive Research in the Digital Age: Recent Advances and Explorations*, pp. 102–124, University of Surrey, 2021.

[16] K. Deep, A. Kumar, and V. Goyal, “Machine translation system using deep learning for Punjabi to English,” in *Proceedings of the International Conference on Paradigms of Computing, Communication and Data Sciences*, pp. 865–878, Springer, Singapore, 2021.

[17] U. R. Ko and H. G. Cho, “SickZIL-machine: a deep learning based script text isolation system for comics translation,” in *International Workshop on Document Analysis Systems*, pp. 413–425, Springer, Cham, 2020.

[18] S. Swathi and L. S. Jayashree, “Machine translation using deep learning: a comparison,” in *International conference on artificial intelligence, smart grid and smart city applications*, pp. 389–395, Springer, Cham, 2020.

[19] Y. Ma and W. Xiang, “Yi-Chinese translation system based on Android platform and deep learning,” in *2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, vol. 5, pp. 1270–1273, Xi’an, China, 2021.

[20] T. Yang, S. Zhao, H. Chen, and B. Chen, “Machine translation system based on deep learning,” in *Journal of Physics:
[21] W. Yang and Y. Lepage, "Improving automatic Chinese-Japanese patent translation using bilingual term extraction," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 13, no. 1, pp. 117–125, 2018.

[22] D. Pesu, Z. Q. Zhou, J. Zhen, and D. Towey, “A Monte Carlo method for metamorphic testing of machine translation services,” in *2018 IEEE/ACM 3rd International Workshop on Metamorphic Testing (MET)*, pp. 38–45, Gothenburg, Sweden, 2018.

[23] Q. Tang, J. Zhou, J. Xin, S. Zhao, and Y. Zhou, “Autoregressive model-based structural damage identification and localization using convolutional neural networks,” *KSCE Journal of Civil Engineering*, vol. 24, no. 7, pp. 2173–2185, 2020.

[24] Y. Ji, H. Hou, J. Chen, and N. Wu, “Training with additional semantic constraints for enhancing neural machine translation,” in *Pacific Rim International Conference on Artificial Intelligence*, pp. 300–313, Springer, Cuvu, Yanuka Island, Fiji, 2019.