Research Article

Translation Mechanism of Neural Machine Algorithm for Online English Resources

Yanping Ye

Foreign Language Department, Xi'an University of Finance and Economics, Xi'an 710100, China

Correspondence should be addressed to Yanping Ye; yeyanping666@xaufe.edu.cn

Received 6 February 2021; Revised 20 March 2021; Accepted 26 March 2021; Published 7 April 2021

Academic Editor: Wei Wang

Copyright © 2021 Yanping Ye. 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.

At the level of English resource vocabulary, due to the lack of vocabulary alignment structure, the translation of neural machine translation has the problem of unfaithfulness. This paper proposes a framework that integrates vocabulary alignment structure for neural machine translation at the vocabulary level. Under the proposed framework, the neural machine translation decoder receives external vocabulary alignment information during each step of the decoding process to further alleviate the problem of missing vocabulary alignment structure. Specifically, this article uses the word alignment structure of statistical machine translation as the external vocabulary alignment information and introduces it into the decoding step of neural machine translation. The model is mainly based on neural machine translation, and the statistical machine translation vocabulary alignment structure is integrated on the basis of neural networks and continuous expression of words. In the model decoding stage, the statistical machine translation system provides appropriate vocabulary alignment information based on the decoding information of the neural machine translation and recommends vocabulary based on the vocabulary alignment information to guide the neural machine translation decoder to more accurately estimate its vocabulary in the target language. From the aspects of data processing methods and machine translation technology, experiments are carried out to compare the data processing methods based on language model and sentence similarity and the effectiveness of machine translation models based on fusion principles. Comparative experiment results show that the data processing method based on language model and sentence similarity effectively guarantees data quality and indirectly improves the algorithm performance of machine translation model; the translation effect of neural machine translation model integrated with statistical machine translation vocabulary alignment structure is compared with other models.

1. Introduction

Machine translation refers to the process of converting two natural languages through a computer on the premise of keeping the semantics unchanged, involving knowledge in many fields including linguistics, computer science, and mathematics [1]. With the increasingly close international exchanges, large-scale cross-language communication scenarios are increasing. These scenarios have high requirements for real-time translation, involve a wide range of languages, and have a huge amount of translation tasks. Manual translation cannot meet the above requirements. The birth of machine translation technology provides a theoretical basis and foundation for solutions to similar scenarios. Machine translation, as one of natural language

processing and one of its important downstream tasks, is the embodiment of humans' ultimate pursuit of letting machines understand human languages [2]. Using machines to achieve high-quality translation between languages has very important practical significance. It can also promote the development of computer science and artificial intelligence [3].

For the online English course education platform, its users pay for a specific course on the course education platform. This behavior is not only a technology acceptance behavior of information systems but also a consumption behavior of Internet learning products and services. In addition, the difference between education products and other products is that online education products are usually paid at a time. If the user's first course learning experience is

not good, when they have a new course demand, there is basically no need to buy other required courses on the same platform again. According to marketing theory, the operating cost of developing new customers is five times the cost of maintaining old customers. Therefore, enhancing user stickiness is the only way for the development of an online English course platform. Only a platform that meets the needs of users, understands the pain points of users, and can maintain the willingness of users to continue to use can remain alive, and a unicorn in the field of online language education can be born.

Traditional translation work is mostly done by manual translation, and its accuracy and translation quality are guaranteed. However, in the context of extremely close global exchanges, the speed and cost of manual translation are far from meeting the corresponding needs. The translation quality of machine translation is slightly inferior to manual translation, but relying on the powerful computing power of computers and the rapid development of the Internet has greatly increased the speed of translation and reduced costs. It has been used by many companies in large-scale translation scenarios. Therefore, the research on machine translation technology and the continuous improvement of the quality of machine translation can make it better serve the economic development and social progress, which has important practical significance.

This paper proposes a neural machine translation model incorporating statistical machine translation vocabulary alignment structure. First, the problem is brought out, the difficulty of the problem is analyzed, and the related work is introduced. Then, we introduced the proposed model architecture in detail. Specifically, based on the neural machine translation “Encoder-Decoder” as the main body, the model adds a statistical machine translation vocabulary knowledge recommendation module to complete the fusion of statistical machine translation vocabulary knowledge. From the aspects of data processing methods and machine translation technology, we carried out experimental comparisons to evaluate the effectiveness of data processing methods based on language model and sentence similarity, as well as machine translation models based on fusion principles. Experimental results show that data processing methods based on language models and sentence similarity can improve data quality to a certain extent, and neural machine translation models incorporating statistical machine translation vocabulary alignment structures can effectively improve machine translation results.

2. Related Work

With the development of neural networks, the birth of neural machine translation technology has opened up a new path in the field of machine translation [4]. The emergence of neural machine translation technology has solved many of the abovementioned shortcomings of statistical machine translation. Neural machine translation, like statistical machine translation, has lower requirements for professional domain knowledge of linguistics. It is a translation method based on a parallel corpus of source and target

languages [5]. Aiming at the locality problem of statistical machine translation, the architecture of neural machine translation adopts the overall sentence-to-sentence translation, which can obtain longer-distance dependencies as much as possible; for the problem of the complicated process of statistical machine translation and numerous functional components, the neural machine translation system is single in the structure; that is, a complete neural network structure can complete the entire translation process and does not need to consider the cooperation and coordination between various components like statistical machine translation [6, 7]. Because neural networks can automatically capture useful features in data, neural machine translation also avoids complicated manual design features. At the same time, because of its single structure and end-to-end training, it is compared with statistical machine translation and neural machine translation [8–10]. The translation process has been greatly simplified.

Related scholars have proposed a semisupervised learning method that uses monolingual corpus to train neural machine translation models [11]. They use different models from the source language to the target language and the target language to the source language to build an automatic encoder to restore the source language sentences according to the target language encoding vector. Related scholars took the lead in introducing phrase-level syntactic knowledge into the RNN-based translation model [12]. They believe that the correspondence between languages is not only the correspondence between words but also the correspondence between words and phrases. The attention mechanism in existing neural translation models is also limited to the word level. The researchers successfully added the source language syntactic structure to the neural translation model. However, their model still has some flaws. For example, they only encode the syntactic structure in a bottom-up manner, which means that the node at the top of the tree sees the whole picture. It is difficult for the bottom node to obtain complete syntactic structure knowledge; for example, when the syntactic structure is added, the hidden state of the encoder doubles, and these states contain overlapping information, so it is difficult for the attention mechanism to avoid such redundant information [13]. Researchers proposed a two-way tree coding structure and, at the same time, introduced a tree structure-based overlay mechanism in the encoder [14]. The two-way tree coding method not only utilizes the bottom-up information flow but also uses the top-down information flow and makes up for the lack of information of some tree nodes. The tree-based coverage mechanism can effectively extract the necessary source language context knowledge in the decoding process. In terms of translation performance, the model has been further improved. Related scholars have proposed a forest-sequence-based neural translation model, which uses source language syntactic forest to avoid errors in the automatic annotation of syntactic structures [15]. Related scholars use a simple way to use the source language dependency syntactic structure [16]. They take the dependency relationship, part of speech, root, and other information in the dependency structure as features and use different

vectors to represent them and stitch them together with the word vectors. The input vectors of the source language words, attention model, and decoder remain unchanged.

Relevant scholars have proposed a classic Encoder-Decoder model for machine translation problems [17]. An encoder composed of a recurrent neural network is used to encode source language sentences into a fixed-length vector, and then the target word is used to predict the next target word, achieving translation effects comparable to traditional translation methods. But this way of encoding variable-length sentences into a fixed-length vector is unreasonable, because different sentences contain different amounts of information, and encoding them into fixed-length vectors will make the vector carry different amounts of information. The average information density in each dimension is also different, so this method becomes a bottleneck to improve the translation effect. In response to this problem, the researchers proposed the use of an attention mechanism to solve this problem and added a bidirectional cyclic neural network to the translation model to improve the translation quality of long sentences [18]. Related scholars have proposed a neural machine translation model Transformer based on the attention mechanism [19]. The model does not use cyclic neural networks and convolutional neural networks but uses the attention mechanism to generate hidden layer states. The distance of the word in the sentence directly learns the dependence of the word, so its long-distance dependence path length is 1 [20].

3. The Key Technology of Neural Machine English Translation Mechanism

3.1. Online English Resource Information Processing for Neural Machine Translation. Figure 1 shows a schematic diagram of online English resource information processing for neural machine translation. Unlike general deep learning tasks, natural language processing tasks are generally sequential learning. The mainstream neural machine translation model is “sequence to sequence” learning. The entire machine model is composed of two neural networks: the first neural network is called the encoder, and the function of the encoder is to encode the source language sequence that needs to be translated and then output it into a fixed-length vector representation; the second neural network is called a decoder, and the vector representation after being encoded by the encoder is decoded into a sequence of the target language by the decoder. Therefore, this model is also called the “Encoder-Decoder” model.

3.2. Long- and Short-Term Memory LSTM. The cyclic neural network theoretically connects the past information with the current task. But there is a big problem in actual operation [21, 22]. When the task that the model needs to deal with is relatively simple, for example, the information related to the predicted word currently to be output is relatively close to itself, the cyclic neural network can make good use of the previously memorized information learning. When the task to be processed is more complicated, the relevant

information required by the current prediction word to be output is far away from itself, and it is difficult for the cyclic neural network to use the previously memorized information, and poorness can even cause translation failure. Suppose that the circular link formed by the hidden layer is

$$h_t = h_{t-1} \cdot W^T. \quad (1)$$

The hidden layer at time t can be expressed as

$$h_t = h_0 \cdot (W^t)^T. \quad (2)$$

The cyclic neural network that introduces long- and short-term memory has the same chain structure as the standard cyclic neural network and can also handle sequence problems well. The difference lies in the internal structure of the hidden layer. There is only one network layer in the hidden layer unit of the standard recurrent neural network (usually a tanh layer), and the long- and short-term memory is much more complicated. In this neural network, each hidden unit structure uses three gate structures to control forgetting or memory information, as shown in Figure 2.

In the cyclic neural network-based model, the context representation vector C is of a fixed length. During the encoding process, all the information in X needs to be compressed into a fixed-length vector. If the input sequence X is long, it will overcompress. At the same time, each x_i in sequence X has no discriminative degree. In the actual language situation, each x_i has a different degree of influence on each y_j , and this degree of influence cannot be reflected in the traditional model.

3.3. Word Segmentation Technology. Word segmentation is very important in the processing of natural language data, and it is often the first step in text processing [23–25]. No matter plagiarism detection, question answering system, or machine translation cannot do without word segmentation, it has a very high status. The simplest method of Chinese word segmentation should be the maximum matching method (forward and reverse). Figure 3 shows the dictionary-based forward maximum matching method.

The word segmentation algorithm in ICTCLAS proposes an improved version of the multilayer Hidden Markov Model (CHMM) on the basis of HMM. In short, it is a simple multilayer HMM combination method. Specifically, this means that the HMM modules in each layer use the N-Best algorithm, the next level of the model will use the best of these generated results, and adjacent HMM levels share the same segmentation words.

3.4. Probability Estimation of Phrase Translation Table

3.4.1. Probability of Two-Way Phrase Translation. When it comes to big data, the word alignment data file is already relatively large, and the phrase translation information extracted from this is likely to reach the scale of GB. Therefore, the general processing method is to apply an external sorting algorithm. Only after processing, the extracted phrase results will be sorted by phrases in the

FIGURE 1: Schematic diagram of online English resource information processing for neural machine translation.

FIGURE 2: LSTM network structure.

source language and appear in sequence. In this case, you read the phrase pairs of each part into the memory in order and calculate their probability distribution.

In the phrase-based method, usually, only the two-way phrase translation table probability is used. In this scenario, once the data is noisy or the extracted phrases are less, the translation result is likely to be affected. Under normal circumstances, the phrase will be broken down into the translation of the corresponding word, and the smoothing method of lexicalization weighting will be used, so that it can check whether some occasional phrase pairs are really correct.

3.4.2. Probability of Bidirectional Lexicalization Translation.

$$P_r(e, f) = \frac{\text{count}[P_r(f, e)]}{\sum_{f_j} \text{count}[P_r(f, e_i)]}. \quad (3)$$

In the above formula, there are inner and outer loops. e_i in the innermost layer is the word in the target sentence. The purpose of the inner layer is to calculate the probability that different f_j sentences are translated into e_i , and then add and average them. The outer layer will traverse all the words in the target sentence and multiply their calculation results together.

FIGURE 3: Maximum forward matching.

3.5. Alignment Mechanism. In the processing of traditional statistical machine translation, word alignment is a very basic and important step. Generally speaking, alignment is divided into several different levels of forms, such as chapters, paragraphs, sentences, phrases, and vocabulary. However, the goal of this step is to find out the parallel prediction database that can be translated into each other.

The current mainstream machine translation system is a phrase-based method. Therefore, word alignment is a fairly basic process. Not only that, in the subsequent phrase extraction module, the result of word alignment is input as basic data. When it comes to word alignment, one has to say GIZA++, which is a typical alignment software that implements models 1~5 designed by IBM and an improved HMM model. The main process is to iteratively train the parallel bilingual library through the maximum expectation (EM) algorithm and obtain the corresponding vocabulary alignment information from the aligned bilingual pairs.

4. Neural Machine Translation Model Incorporating Statistical Machine Translation Vocabulary Alignment Structure

4.1. Statistical Machine Translation Vocabulary Alignment Structure. The statistical machine translation vocabulary recommendation module is responsible for perceiving and using the attention information of neural machine translation and neural machine translation target language to generate historical information to make vocabulary recommendations. Given the historical information y generated in the target language of neural machine translation, the statistical machine translation model is ideally aimed at the untranslated part of the source language sentence, and the phrase table is looked up and combined with various features to evaluate and score the next translation candidate:

$$SMT_{score}(y|x, y_{\leq t}) = \sum_m^M \theta \cdot h_m(x_t, y_t), \quad (4)$$

where y_t is a translation candidate, and x_t is its corresponding source language word. $H_m(y_t, x_t)$ is the translation

feature, and θ is the weight of the feature. The statistical machine translation model will recommend appropriate vocabulary knowledge based on the evaluation score for neural machine translation.

The above formula represents that the neural machine translation alignment model generates alignment probabilities for all words in the source language sentence according to the attention mechanism. Different from the explicitly specified word alignment in statistical machine translation, this alignment probability of all words in the source language sentence is called “soft alignment.”

The “soft alignment” in neural machine translation brings two problems to the vocabulary knowledge recommendation module of statistical machine translation. The first problem is the evaluation and scoring of the ordering model in the statistical machine translation model. The ordering model is a model that characterizes the difference in the word order of different languages and plays a very important role in statistical machine translation. However, the ordering model of statistical machine translation is designed for word alignment in statistical machine translation. The scoring formula of the ordering model is as follows:

$$d(y_t) = sp_{y_t} + sp_{y_{t+1}} + 1. \quad (5)$$

Among them, sp_{y_t} is the position of the source language word corresponding to the target language word y_t , and $sp_{y_{t+1}}$ is the position of the source language word corresponding to the target language word y_{t+1} . The position information is obtained based on the word alignment information in statistical machine translation.

This article uses the probability distribution in “soft alignment” to evaluate and score the ordering model based on distance:

$$d(y_t) = \sum_{j=1}^{T_x} \alpha_{j,t-1} (j + 1 + sp_{y_t}). \quad (6)$$

According to the word alignment information in the phrase table, the statistical machine translation model can obtain the position of the source language word corresponding to the target language word.

In the initial stage, the contents of the coverage vector are all set to 0. In the decoding stage, if the final word generated by the model appears in the statistical machine translation vocabulary recommendation set of the decoding stage, the corresponding cover vector is set according to the word alignment information in the statistical machine translation phrase table.

4.2. Gating Mechanism Model. In order to integrate the vocabulary knowledge of statistical machine translation, the gating mechanism model sets a classifier to estimate the probability of the statistical machine translation vocabulary recommendation at the current decoding moment and then uses a neural network-based gate control adjustment model to estimate the probability and the neural machine. The estimated probability of translating its own words is reestimated.

Specifically, at decoding time t , the classifier that processes statistical machine translation vocabulary recommendations combines the decoding information at the current decoding time to reevaluate and score each statistical machine translation vocabulary recommendation:

$$\text{score}_{\text{smt}}(y|x, y_{\leq t}) = g_{\text{smt}}(c_t, s_t, y_t, y_{t+1}), \quad (7)$$

where y is the word recommendation being evaluated, s_t is the implicit state of the neural machine translation decoder at the current moment, y_{t-1} is the word representation of the word generated by the model at the previous moment, and c_t is the context vector at the current moment. g_{smt} is a nonlinear activation function.

The purpose of designing the model in this form is to make full use of the decoding information of the decoder (source language sentence information and target language sequence history information) to evaluate the matching degree of the vocabulary recommendation with the current decoding environment. Then, you use the evaluation score of each word recommendation to estimate the probability of matching the word recommendation with the current decoding environment:

$$p_{\text{smt}}(y|x, y_{\leq t}) = \text{soft max}[\text{score}_{\text{smt}}(y|x, y_{\leq t})]. \quad (8)$$

The above formula indicates that the vocabulary in the current decoding environment is recommended for probability normalization. The gate control adjustment model based on neural network interpolates and sums the probability estimation of word recommendation and the word probability estimation of neural machine translation itself:

$$p(y_t|x, y_{\leq t}) = \alpha_t^* \cdot p_{\text{smt}}(y_t|x, y_{\leq t}) + (\alpha_t^* - 1) \cdot p_{\text{nmt}}(y_t|x, y_{\leq t}). \quad (9)$$

Among them, the word y in the neural machine translation vocabulary does not appear in the recommended word set, and then

$$p_{\text{smt}}(y_t|x, y_{\leq t}) = 0. \quad (10)$$

The interpolation probability α_t is calculated as follows:

$$\alpha_t = g_{\text{gate}}(y_{t-1}, s_t, c_t). \quad (11)$$

Among them, g_{gate} is a nonlinear activation function.

The setting of the gate control adjustment model based on neural network is to let the model automatically learn by judging the decoding environment to determine whether to rely on the vocabulary knowledge provided by statistical machine translation. The door control model for statistical machine translation vocabulary recommendation is shown in Figure 4.

4.3. Direct Competition Mechanism Model. The difference between this model and the gating mechanism model lies in the calculation of statistical machine translation vocabulary probabilities and the probability reestimation with neural machine translation word prediction probabilities. Specifically, at decoding time t , the scoring device that processes statistical machine translation vocabulary recommendations combines the decoding information at the current decoding time to reevaluate and score each statistical machine translation vocabulary recommendation:

$$\text{score}_{\text{smt}}(y|x, y_{\leq t}) = g_{\text{smt}}(c_t, s_t, y_{t-1}, y_t), \quad (12)$$

where y is the word recommendation being evaluated, s_t is the implicit state of the neural machine translation decoder at the current moment, y_{t-1} is the word representation of the word generated by the model at the previous moment, and c_t is the context vector at the current moment. g_{smt} is a nonlinear activation function.

The model uses calculated statistical machine translation vocabulary recommendations and neural machine translation target language word probabilities to reestimate the probability:

$$p(y_t|x, y_{\leq t}) = p_{\text{smt}}(y_t|x, y_{\leq t}) - p_{\text{nmt}}(y_t|x, y_{\leq t}). \quad (13)$$

5. Experiment and Analysis

5.1. Data Processing Method Comparison Experiment. The application data comes from the open data set of machine text translation of online English resources. The original data set contains 10 million parallel sentence pairs in Chinese and English. Considering the actual experimental hardware environment and experimental time conditions, it is specified that 100,000 pairs of parallel sentences are selected to construct a training data set, and 10,000 pairs of parallel sentences are selected to construct a test data set. The test data set and the training data set sentence pairs are mutually exclusive.

In order to verify the effectiveness of data processing methods based on language models and sentence similarity, comparative experiments of different machine translation models under the same parameters and different training data set settings were carried out, and different training data sets were used to train each machine translation model. We compare the BLEU scores of each trained model under the same test data set.

FIGURE 4: Door control model for statistical machine translation vocabulary recommendation.

Randomly, we select 10,000 pairs of parallel sentences as the test data set and apply the data processing method (Similar-Gram) and the basic random screening method (Random) to construct the data containing 100,000 pairs of parallel sentences in the remaining data as the training data set. Among them, the random selection method selects data according to the Random method after segmentation and generalization, and the data set selects the translation direction from English to Chinese.

The training data set constructed using the two data processing methods of Similar-Gram and Random was used to train three machine translation models, including the machine translation model based on the fusion principle (Chunk-based) and the machine based on the reply neural network translation model (Encoder-Decoder) and Encoder-Decoder model based on the attention mechanism. We compare the application effects of the new algorithm model and the classic algorithm model on the data set to ensure that the common parameter settings of the training process are the same. After the model training is completed, the BLEU score obtained by applying the trained model on the same test data set is shown in Figure 5.

The BLEU score obtained by the algorithm model in the test data set is used to characterize the pros and cons of the training data set and the algorithm model. It can be found from Figure 5 that, for any of the three algorithm models, when the model parameter settings are consistent, the training data set processed by the Similar-Gram data processing method performs better than the training data set processed by the Random data processing method. The experimental results show that the data processing method based on the language model and sentence similarity can improve the application data to a certain extent.

FIGURE 5: Comparison of experimental results of data processing methods.

5.2. Comparative Experiment of Machine Translation Technology. In order to verify the effectiveness of the machine translation model based on the fusion principle, from the comparison of the machine translation model based on the fusion principle under different parameters and the same training data set setting, the comparison experiment is carried out in two aspects, and the cross-entropy loss of the training data set of the machine translation model based on the fusion principle and the BLEU score of the training data set increase with the number of training iterations. Finally, the machine translation model based on the fusion principle is drawn.

5.2.1. Parameter Setting. In this experiment, the data set chooses the translation direction from Chinese to English, and the performance of the model is improved by adjusting the parameters, which is mainly reflected in the accuracy of prediction, training time, and prevention of overfitting.

The dropout parameter is used to control the probability that the neurons in the network are discarded during the model training and tuning process. Assuming that the dropout is set to 0.5, the tuning value of 50% of the neurons is randomly ignored during each training and tuning process of the model. Dropout is an important means to prevent the model from overfitting. In order to select the best dropout value under the current data and the current application scenario, we compare the iterative training of the model when the dropout is set to 0.3, 0.6, and 0.9, respectively. The model cross-entropy loss of the first 5 iterations and the BLEU score of the training data set are shown in Figures 6 and 7. In the comparison experiment, we ensure that the remaining parameters are consistent. We set the parameter cell size to 124, parameter Chinese vocabulary to 20000, parameter English vocabulary to 12500, parameter layer size to 3, and parameter learning rate to 0.0004.

It can be seen from Figures 6 and 7 that dropout can avoid overfitting to a certain extent, but setting the dropout value to a larger value may increase the training time of the model. Since the number of hidden layers of the model is set to 2, the number of model layers is less, and combined with the comparison experiment results, considering the model training time and the risk of overfitting, we set dropout to 0.2 to ensure the model's performance while avoiding model overfitting to a certain extent.

The learning rate parameter is used to control the model training and tuning step length. The learning rate will affect the learning efficiency and accuracy of the model. Setting this value too high may cause the model to miss the optimal solution and reduce the accuracy. Too low may cause the model to obtain the local optimum and greatly extend the training time of the model. The learning rate setting size is related to the optimizer. The optimizer used in this article is Adam. The optimizer requires a small learning rate. Therefore, the model is iteratively trained when the learning rate is set to 0.0002, 0.0004, and 0.0008. The cross-entropy loss of the model and the BLEU score of the training data set are shown in Figure 8 and Figure 9. In the comparison experiment, we ensure that the remaining parameters are consistent.

It can be seen from Figures 8 and 9 that the model training speed is slow when the learning rate value is set small, and the model may miss the optimal solution when the learning rate value is set large. The optimizer used for model training is Adam, which requires a small learning rate and can be adjusted during model training. We compare the model training iteration effects when the learning rate is 0.0001, 0.0005, and 0.001 and set the initial value of the learning rate to 0.0005.

5.2.2. Comparison of Different Machine Translation Models. We specify each parameter value of the machine translation model according to the above parameter setting results and

FIGURE 6: Comparison of cross-entropy loss under different dropouts.

FIGURE 7: Comparison of BLEU scores under different dropouts.

FIGURE 8: Comparison of cross-entropy loss under different learning rates.

compare the machine translation models under the condition that the common parameter settings are consistent. The comparison method selects the Encoder-Decoder model (Encoder-Decoder) in the machine translation technology based on the reply neural network and the Encoder-Decoder

FIGURE 9: Comparison of BLEU scores under different learning rates.

model (Attention-based) based on the attention mechanism and compares the fusion algorithm model in this paper with other algorithm models. The common parameter settings of the training process are the same. After the model training is completed, the BLEU score and cross-entropy loss on the training data set and the results obtained by testing on the same test data set are shown in Figure 10.

It can be seen from Figure 10 that the cross-entropy loss of each translation model on the training data set is different, but the fitting is lacking; the BLEU score on the training data set of each model is similar when the training is stopped. On the same test data set, it can be found that the translation effect of the machine translation model based on the fusion principle is improved compared with the Attention-based model and the Encoder-Decoder model in the classic model. The data in Figure 10 also proves that the translation effect of the model on the test data set is not completely positively correlated with the fitting effect on the training data set.

In this paper, the BLEU score obtained by executing the test data set with the algorithm model represents the pros and cons of the algorithm model. It can be found from Figure 10 that, under the same application data set, the machine translation models based on the fusion principle all improve the machine translation effect to a certain extent.

It should be noted that due to the limitation of experimental conditions, the training data set used in the comparative experiment of machine translation technology only contains 100,000 pairs of parallel sentences. The insufficient corpus may lead to insufficient training of the machine translation model; in addition, due to the lack of data set and the lack of quality and the abstractness of the neural network model itself, the experimental results of the actual translation model training and testing process may have certain deviations. However, the setting of the comparison experiment ensures the consistency of the training data set, the test data set, and the parameter settings, and the comparison experiment results can correctly evaluate the effectiveness of the machine translation model.

FIGURE 10: Comparison of experimental results of machine translation models.

FIGURE 11: Cross-entropy loss and BLEU score during model training.

5.2.3. Training and Fitting of Machine Translation Model Based on Fusion Principle. For the machine translation model based on the fusion principle, the training data set constructed by the data processing method is used for training. The data set selects the translation direction from Chinese to English. The cross-entropy loss of the training data set of the model and the BLEU score of the training data set increase with the number of training iterations. The change curve is shown in Figure 11. It can be seen from Figure 11 that as the number of model training iterations increases, the overall cross-entropy loss of the training data set shows a downward trend. The decrease is small in the first few iterations of training and then gradually increases. However, after the model tends to converge, the cross-entropy loss does not drop to a very low level, which may be caused by various reasons such as data characteristics and model training. In addition, as the number of model training iterations increases, the overall BLEU score of the training data set shows an upward trend, with a small increase in the first few iterations of training, and then gradually increases.

6. Conclusion

This paper uses statistical machine translation vocabulary alignment knowledge to alleviate the smooth and unfaithful translation problem of neural machine translation. The model is mainly based on neural machine translation, and the fusion of statistical machine translation vocabulary alignment information is based on neural networks and continuous expression of words. Specifically, at each decoding moment, statistical machine translation provides recommendations containing vocabulary alignment information based on the decoding information of neural machine translation. The neural machine translation decoder uses this vocabulary recommendation to adjust its own target language word prediction probability, in order to make full use of the vocabulary alignment information to more faithfully generate the target language sentence. From the aspects of data processing methods and machine translation technology, experiments are carried out to compare the data processing methods based on language model and sentence similarity and the effectiveness of machine translation models based on fusion principles. Experimental results show that the data processing method based on language model and sentence similarity can ensure data quality to a certain extent, and the neural machine translation model integrating statistical machine translation vocabulary alignment structure can effectively improve the effect of machine translation. Although neural machine translation has achieved breakthrough results, the training of neural networks relies on a large amount of training data. However, there is no bilingual data on a large number of language pairs, so unsupervised machine translation operations were born. Unsupervised machine translation is based on monolingual data and establishes a potential connection between the two languages through a neural network to obtain a coarse-grained alignment relationship. On this basis, the coarse-grained alignment knowledge is gradually refined using ideas such as iteration. The syntactic structure is a manifestation of the internal structure of a language. In the future, we hope to use syntactic knowledge to help unsupervised machine translation to better establish the internal connection between languages.

Data Availability

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

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the General Special Scientific Research Project of Education Department of Shaanxi Provincial Government, 2020—Translation Research Based

on the Logical Semantic Relations of Images and Texts from a Multimodal Perspective (20JK0150).

References

- [1] Y. Fan, F. Tian, Y. Xia, T. Qin, X.-Y. Li, and T.-Y. Liu, "Searching better architectures for neural machine translation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1574–1585, 2020.
- [2] D. Datta, P. E. David, D. Mittal et al., "Neural machine translation using recurrent neural network," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 4, pp. 1395–1400, 2020.
- [3] X. Zhang and K. Duh, "Reproducible and efficient benchmarks for hyperparameter optimization of neural machine translation systems," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 393–408, 2020.
- [4] B. Zhang, D. Xiong, J. Xie, and J. Su, "Neural machine translation with GRU-gated attention model," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4688–4698, 2020.
- [5] M. Yang, S. Liu, K. Chen, H. Zhang, E. Zhao, and T. Zhao, "A hierarchical clustering approach to fuzzy semantic representation of rare words in neural machine translation," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 992–1002, 2020.
- [6] C. Park, C. Lee, Y. Yang, and H. Lim, "Ancient Korean neural machine translation," *IEEE Access*, vol. 8, pp. 116617–116625, 2020.
- [7] X. Liu, W. Wang, W. Liang, and Y. Li, "Speed up the training of neural machine translation," *Neural Processing Letters*, vol. 51, no. 1, pp. 231–249, 2020.
- [8] R. Dabre, C. Chu, and A. Kunchukuttan, "A survey of multilingual neural machine translation," *ACM Computing Surveys*, vol. 53, no. 5, pp. 1–38, 2020.
- [9] M. Yang, M. Zhang, K. Chen, R. Wang, and T. Zhao, "Neural machine translation with target-attention model," *IEICE Transactions on Information and Systems*, vol. E103.D, no. 3, pp. 684–694, 2020.
- [10] D. Mareček, H. Celikkanat, M. Silfverberg et al., "Are multilingual neural machine translation models better at capturing linguistic features?" *The Prague Bulletin of Mathematical Linguistics*, vol. 115, pp. 143–162, 2020.
- [11] K. Chen, R. Wang, M. Utiyama et al., "Towards more diverse input representation for neural machine translation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1586–1597, 2020.
- [12] M. Yang, R. Wang, K. Chen, X. Wang, T. Zhao, and M. Zhang, "A novel sentence-level agreement architecture for neural machine translation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2585–2597, 2020.
- [13] E. Salesky, A. Runge, A. Coda, J. Niehues, and G. Neubig, "Optimizing segmentation granularity for neural machine translation," *Machine Translation*, vol. 34, no. 1, pp. 41–59, 2020.
- [14] Y. La Su and W. W. Liu, "Research on the LSTM Mongolian and Chinese machine translation based on morpheme encoding," *Neural Computing and Applications*, vol. 32, no. 1, pp. 41–49, 2020.
- [15] J. G. Makin, D. A. Moses, and E. F. Chang, "Machine translation of cortical activity to text with an encoder-decoder framework," *Nature Neuroscience*, vol. 23, no. 4, pp. 575–582, 2020.

- [16] S. Stoll, N. C. Camgoz, S. Hadfield, and R. Bowden, "Text2Sign: towards sign language production using neural machine translation and generative adversarial networks," *International Journal of Computer Vision*, vol. 128, no. 4, pp. 891–908, 2020.
- [17] S. Nakagome, T. P. Luu, Y. He et al., "An empirical comparison of neural networks and machine learning algorithms for EEG gait decoding," *Scientific Reports*, vol. 10, no. 1, pp. 1–17, 2020.
- [18] E. Chatzikoumi, "How to evaluate machine translation: a review of automated and human metrics," *Natural Language Engineering*, vol. 26, no. 2, pp. 137–161, 2020.
- [19] D. Grechishnikova, "Transformer neural network for protein-specific de novo drug generation as a machine translation problem," *Scientific Reports*, vol. 11, no. 1, pp. 1–13, 2021.
- [20] Y. Qiao, K. Hashimoto, A. Eriguchi et al., "Parallelizing and optimizing neural Encoder-Decoder models without padding on multi-core architecture," *Future Generation Computer Systems*, vol. 108, pp. 1206–1213, 2020.
- [21] D. Deng and N. Xue, "Translation divergences in Chinese-English machine translation: an empirical investigation," *Computational Linguistics*, vol. 43, no. 3, pp. 521–565, 2017.
- [22] M. S. Maučec and J. Brest, "Slavic languages in phrase-based statistical machine translation: a survey," *Artificial Intelligence Review*, vol. 51, no. 1, pp. 77–117, 2019.
- [23] R. Wang, H. Zhao, S. Ploux, B.-L. Lu, M. Utiyama, and E. Sumita, "Graph-based bilingual word embedding for statistical machine translation," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 17, no. 4, pp. 1–23, 2018.
- [24] 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.
- [25] B. Zhang, D. Xiong, J. Su et al., "Alignment-supervised bidimensional attention-based recursive autoencoders for bilingual phrase representation," *IEEE Transactions on Cybernetics*, vol. 50, no. 2, pp. 503–513, 2020.