

Research Article

Intelligent Recognition English Translation Model Based on Embedded Machine Learning and Improved GLR Algorithm

Lei Lei 

School of Foreign Languages, Xi'an Shiyou University, Xi'an 710065, Shaanxi, China

Correspondence should be addressed to Lei Lei; leilei010702@xsyu.edu.cn

Received 7 May 2022; Revised 12 July 2022; Accepted 25 July 2022; Published 13 September 2022

Academic Editor: Imran Shafique Ansari

Copyright © 2022 Lei Lei. 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.

Most of the popular translation models are based on encoder-decoder architecture and belong to the autoregressive translation model. When autoregressive translation models decode, they generate the current sequence according to the sequence generated before. This process is not parallel. The generalized maximum likelihood ratio detection (GLR) algorithm model cannot effectively guarantee the overlapping and accurate results of English translation detection. To improve the recognition rate of English phrases and meanings, this paper proposes an intelligent model for English translation recognition based on embedded machine learning and an improved GLR algorithm. A corpus of 520000 English phrases is used for training. And we compare and analyze different corpora and compare GLR algorithm with other traditional algorithms. Words and phrases are based on analytic linear structure. The syntactic function of the table corrects the ambiguity between English and Chinese structures in some speech recognition results and finally retains the recognition content. The research shows that the recognition accuracy based on the improved GLR algorithm is more than 96.58%, which is 23% higher than the classical GLR in semantic recognition. Statistical algorithm and dynamic storage algorithm make it more suitable for intelligent translation and provide a new model method for intelligent machine translation.

1. Introduction

Although most of the current translation models are embedded machine learning models, which are also widely used in many generation tasks, there are also significant shortcomings in the process of the English-Chinese translation of autoregressive translation models. For example, the fitting degree of data in regression model is not high, and sentences cannot be translated completely. Using the current translation models will easily lead to data sparsity and overfitting problems, and the lack of semantic information will affect the dependency of context information; thus, the quality of English-Chinese translation is not high. The decoding process of the embedded machine learning model is serial output, the current sequence depends on the previously generated sequence, which leads to high generation time complexity, and it usually takes a long time to generate with GPU. Therefore, it is necessary to propose an intelligent English translation model based on embedded machine

learning and improved GLR algorithm to improve translation efficiency.

Chiang translates the original Hindi documents into English through a machine translation system and then uses an interactive system based on English for multi-document summarization and title generation to realize cross-language summarization generation [1]. Bai considers that applying translation regularization to the model can effectively improve the robustness of the model. Therefore, language constraints are imposed on the RC model through translation regularization [2]. Ren et al. first use the monolingual summarization method to generate source language text summarization and then use machine translation to generate target language summarization [3]. To avoid the influence of error accumulation, Su et al. proposed a joint optimization model of machine translation and summarization and realized the cross-language summarization of zero-shot [4]. Its core idea is to construct a linear system from translation to abstract or from abstract to translation. Liu et al. first

proposed to use the GLR framework to generate cross-language abstracts. That is, input the original text of the source language directly to decode the cross-language abstract, and use the cross attention from the encoder to the decoder to achieve the alignment of the two languages [5]. Sutskever et al. further improve the way of translation integration by combining neural network model with an external probabilistic bilingual dictionary to improve the performance of cross-language summarization [6]. It can be found that many scholars' researches only optimize the framework of translation model, but there is no real algorithm to refine the translation model.

Based on shared machine translation encoder, Vaswani et al. introduce phrase-based statistical machine translation (SMT) model which is robust to noise data as translation regularization to guide the unsupervised NMT model training in the process of reverse translation; existing Bayesian models, especially nonparametric Bayesian methods, rely on special priors and domain knowledge to discover and improve potential representations [7]. Although a priori can affect the translation distribution through the Bayesian theorem, it may be more direct to directly apply the regularization term to the translation distribution, which is more natural and easier in some cases. Therefore, Luong et al. proposed a Bayesian translation regularization reasoning framework with expectation constraints [8]. Yoon and Alexander proposed a general constrained translation regularization framework, which added additional knowledge as translation constraints to model training. Translation regularization separated model complexity from structural constraint complexity and effectively integrated indirect learning by constraining the translation distribution of probabilistic models with potential variables, thus improving the computational efficiency of the model [9]. In Chris et al.'s model, the super-parameter B is usually real value, but it is difficult to constrain the feature expectation effectively by specific super-parameters [10]. Although the above research based on machine translation cross-language summarization method can use monolingual summarization and machine translation model, it is affected by the error accumulation of two independent subtasks. The error of the previous step will affect the performance of the next step, which restricts the quality of translation.

Based on embedded machine learning and an improved GLR algorithm, this paper proposes an intelligent model to recognize English translation. A corpus of 520000 English phrases is used for training. The innovations of this paper are as follows: (1) this paper proposes a method of Chinese-English cross-language translation generation for embedded machine learning based on word alignment; (2) based on the embedded machine learning method of the Chinese-English bilingual dictionary, bilingual word vectors are aligned in the same semantic space; (3) the target language translation is obtained by decoding the bilingual context vector based on the attention mechanism. Words and phrases are based on analytic linear structure. The syntactic function of the table corrects the ambiguity between English and Chinese structures in some speech recognition results and finally retains the recognition content.

2. Experiment and Method

2.1. Experimental Environment and Settings. This experiment uses CentOS 7.4 operating system, 16 GB memory, i7 8700k processor, and two NVIDIA P40 graphics cards. CentOS Linux distribution is a stable, predictable, manageable, and reproducible platform, which is compiled from the source code released by Red Hat Enterprise Linux (RHEL) according to the open-source code (mostly GPL open-source protocol). This article uses Tensor2Tensor which is open source of Google. The tool trains the model and rewrites the data processing part of the model. For the comparability of experimental results, GLR is used in all systems_Base parameters; using a single GPU training, benchmark model training iterations are 200000 steps, and other subsystem training iterations are 100000 steps. At the end of the training, the last 30 rounds of training model parameters are averaged. The settings of the parameters in the experiment are those commonly used in previous studies. For the samples in this experiment, such parameters can be optimized more effectively.

2.2. Embedded Machine Learning Parallel Translation. The parallel translation is a bilingual or multilingual corpus composed of the source text and target text. Parallel translation can be divided into vocabulary level, phrase level, sentence level, paragraph level, and text level. Parallel translation contains rich linguistic knowledge, which can provide data support for many natural language processing (NLP) tasks, such as machine translation, bilingual dictionary construction, word sense disambiguation, and cross-language information retrieval. Large-scale and high-quality parallel translation can greatly improve the performance of these NLP tasks. Manual construction of a parallel translation library can ensure the quality of data, but the data are not easy to form scale and expensive, so people often use computer technology to collect parallel translation, which can save time and effort and ensure large-scale production [11]. In addition, the emergence of neural machine translation makes the translation quality of machine translation reach a new height. Extracting features from machine translation and building a parallel translation library based on multi-translation engines are also the development trend of the construction of parallel translation libraries [12]. Sentence alignment refers to the process of finding the best mapping relationship by using the characteristics of sentences from texts with the same or comparable content expressed in different languages, to find out the mutual translation sentence pairs [13]. In the construction process of parallel translation library, parallel translation at the next level or paragraph level is relatively easy to obtain. The source language end and target language end of a text or paragraph are divided into sentences, respectively, then the parallel sentence pairs can be obtained by using the sentence alignment technology, and then the parallel translation at the lexical level or phrase level can be obtained by using the word alignment technology [14]. The performance of sentence alignment determines the quality and scale of the

sentence-level parallel translation library [15]. Therefore, sentence alignment has important application value and is of great help to the research of corpus-based NLP tasks [16].

2.3. Improved GLR Algorithm for English Sentence Alignment

2.3.1. Length-Based Method. The length-based method is an earlier and more widely used method, and its calculation principle is mature. This method calculates the similarity of bilingual sentence pairs according to the length of characters, the number of words, or the number of bytes in the sentence. The feature extraction is simple, and the alignment performance is good on the language pairs of the same language family, but other semantic features are ignored. Here, it is defined according to the constraint feature $f(x, y)$ and its expectation [17]. Here, it should be noted that constraint features never appear in the model. Consider a specific example; in the process of natural language learning, we hope that the model can be biased learning; that is, each sentence is marked with at least one verb. Therefore, the information is encoded, the constraint feature is defined, and the expectation of the feature is at least R [18]. Here, to keep consistent with other discussion and standard optimization literature, the equivalent representation is used here.

$$R = -\frac{1}{T} \ln(1 + \beta). \quad (1)$$

The expected value is less than k , and the specific expression is as follows:

$$k_{t1}[i] = \sum_j \cos(w_i^1, w_j^2). \quad (2)$$

In practice, applying this constraint to the model will break the first-order Markov property of the distribution. Therefore, each sentence is required to contain at least one noun, so additional constraints are usually imposed in the model [19]. To ensure the efficiency of the algorithm, the constraint feature u is usually required as the sum decomposition.

$$Q_{jh} = \frac{\sum_{Z=1}^{h_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}|}{n_j n_h (u_j + u_h)},$$

$$U_t = \sum_{j=2}^k \sum_{h=1}^{j-1} G_{jh} (p_j s_h + p_h s_j) D_{jh} (1 - D_{jh}). \quad (3)$$

2.3.2. Method Based on Mutual Translation Information.

This method focuses on the mutual translation information of words in the dictionary or the mutual translation information of sentences obtained by the machine translation engine for sentence alignment. Although the accuracy is high, the alignment speed is slow, and the alignment effect depends heavily on the scale and quality of the bilingual dictionary or the performance of the translation engine. Although the GLR model achieves the parallel output of the decoder by improving the encoder structure, which greatly

improves the speed of machine translation, the decoding stage is completely independent, and in the actual translation, the final translation sentence is not conditional independent. The conditional independence hypothesis stops the translation model from correctly capturing the highly multimodal distribution of the target translation. Ignoring the context information of the target language in the model will lead to insufficient semantic information of the target language and reduced translation effect [20]. Knowledge machine learning methods can solve this problem to a certain extent. Knowledge machine learning processing of data sets can maintain the original generation rate and further improve the effect of translation [21]. Therefore, we apply the knowledge machine learning method to improve the defects of the GLR model. The specific process is as follows: first, we need to train an autoregressive machine translation model as a GLR model. To set up conveniently, we use the GLR model as a GLR model; next, with the help of the GLR model, cluster search is carried out on the training set to build a new corpus, namely, machine learning data set; finally, the machine learning data set is applied to the GLR model, and the negative log-likelihood function is used for training to improve the GLR model translation effect [22]. The implementation process is shown in

$$p(t|s) = \alpha + \beta \ln T_{it} - 1 + v_i + \mathfrak{F}_t, \quad (4)$$

$$q(t|s) = \alpha + \beta \ln T_{it} - 1 + \phi X_{it} - 1 + v_i + \tau_t, \quad (5)$$

where $p(t|s)$ is the sequence-level distribution, $q(t|s)$ is the sequence distribution of GLR model on the possible samples and takes the approximate value, and U_y is the machine learning data set generated by cluster search with GLR model [23]. This paper analyzes the role of the knowledge machine learning method in machine translation and the reasons why knowledge machine learning can promote it. Understanding the principle of knowledge machine learning has a great positive effect on this experiment. With the help of cluster search of GLR model, knowledge machine learning of data sets can reduce the complexity of data sets and the dependence of target end on context, better help GLR translation model simulate the change of output data, and then improve the translation effect of GLR model [24].

2.3.3. Hybrid Method. This method combines the length method and the mutual translation information method. Firstly, several candidates' bilingual sentence pairs are obtained by using the sentence length method. Then, the mutual translation information method is used to test the similarity of the candidate bilingual sentence pairs, and the highest score is selected as the optimal sentence. Indirect learning is realized by constraining the translation distribution of the structure. The basic idea of the model is to use the KL deviation between the expected distribution containing prior knowledge and model translation to punish the log-likelihood of the neural translation model. The translation regularization likelihood is defined as

$$KL_{jh} = \int_0^\infty dF_j(y) \int_0^y (y-x) dF_h(x),$$

$$f(x) = \frac{1}{Nh} \sum_{i=1}^N k\left(\frac{X_i - x}{h}\right), \quad (6)$$

where x and H are the super-parameters of the model to balance the differences between the likelihood function and the translation regularization. Q is the set of constrained translations.

$$\frac{H_{nb}}{Q} = \sum_{j=2}^k \sum_{h=1}^{j-1} G_{jh} (p_j s_h + p_h s_j) X_{jh},$$

$$X_{jh} = \frac{d_{jh} - B_{jh}}{d_{jh} + B_{jh}}, \quad (7)$$

$$(x, y)_j = \sum_i B_{ij} u_{(ji)}.$$

where (x, y) is the constraint feature and B is the expected bound of the constraint feature. In this paper, constraint features are used to encode structural deviation, a set of effective distributions are defined according to the expectation of constraint features, and then effective reasoning is carried out. However, the main difficulty in applying translation regularization directly to neural machine translation is that in machine translation, the super-parameter B is usually real value, but it is difficult for specific super-parameters to effectively constrain feature expectation [25]. It is difficult to set an appropriate threshold NMT for all sentences in the training data because the penalty values of different sentences are quite different.

$$NMT = \frac{1/2u_j \sum_{i=1}^{n_j} \sum_{r=1}^{n_j} |y_{ji} - y_{jr}|}{n_j^2}, \quad (8)$$

$$Qt = \sum_{j=1}^k G_{jj} p_j s_j.$$

In addition, the proposed min-max algorithm contains an additional step to calculate QT, which significantly increases the training time compared with the standard NMT, and any prior knowledge is still challenging. Suppose there is an X satisfying

$$X_{ij} = \frac{e^{b_{ij}}}{\sum_k e^{b_{ik}}}. \quad (9)$$

2.3.4. Method Based on Neural Network. This kind of method is mainstreamed at present. Firstly, the source language sentence and the target language sentence are mapped into a fixed-length vector representation at the sentence level, and then the similarity of the vector is calculated to determine whether the source language sentence and the target language sentence are aligned. This method benefits from the strong representation ability of neural

networks, and the accuracy and efficiency of sentence alignment are better than traditional methods, but the vector representation of sentences depends on the pretraining model [26, 27]. In this paper, the best pretraining is introduced into the sentence alignment method, and features are extracted by bidirectional GLR. Each word is composed of three vectors: position information vector, word information vector, and sentence information vector. These three vectors are superimposed as the final word vector to represent the semantic information of words and then implement two-way measurement between the source language sentence and the target language sentence, and integrate BLEU score, cosine similarity, and Manhattan distance for sentence alignment [28]. Experimental results show that the proposed method can achieve better sentence alignment and high-quality parallel translation. Make XW the closest to Z , and use Euclidean distance to evaluate as follows:

$$XW_t = \tanh(w_c x_t + u_c (r_t \Theta h_{t-1}) + Z_c),$$

$$B_t = z_t \Theta h_{t-1} + (1 - z_t) \Theta h_t, \quad (10)$$

$$Z = \frac{\sqrt{1/n \sum_{i=1}^n (I_{it} - B_{it})^2}}{I_{it}}.$$

Because each word consists of three vectors: position information vector, word information vector, and sentence information vector, it is necessary to normalize the maximum cosine length of word vector.

$$W = \frac{T}{XW}$$

$$= \frac{\alpha}{1 - \beta}, \quad (11)$$

$$\alpha = \frac{\sum_{j=1}^k \sum_{h=1}^k \sum_{t=1}^{n_j} \sum_{r=1}^{n_h} |y_{ij} - y_{hr}|}{2n^2 u}.$$

Through the above operation, the matrix W is obtained to embed and map the two languages, and the cross-language word vector is obtained, which reduces the reasoning distance between the two languages and improves the effect of machine translation. Among them, the inference distance of translation indicates the difficulty and task of translation. Reducing the inference distance can improve the effect of machine translation.

3. Results and Analysis

3.1. Machine Learning English Translation Recognition Analysis. As shown in Figure 1, the complexity of the model is reflected in many aspects, and there are more connections in content, while in structure, it is more on the translation model of words, as follows: the translation regularization framework can add additional knowledge to the model training as the translation constraints of the model and separate the model complexity from the structural constraint complexity in this way. Through decomposable regularization of translation moments of potential variables in the

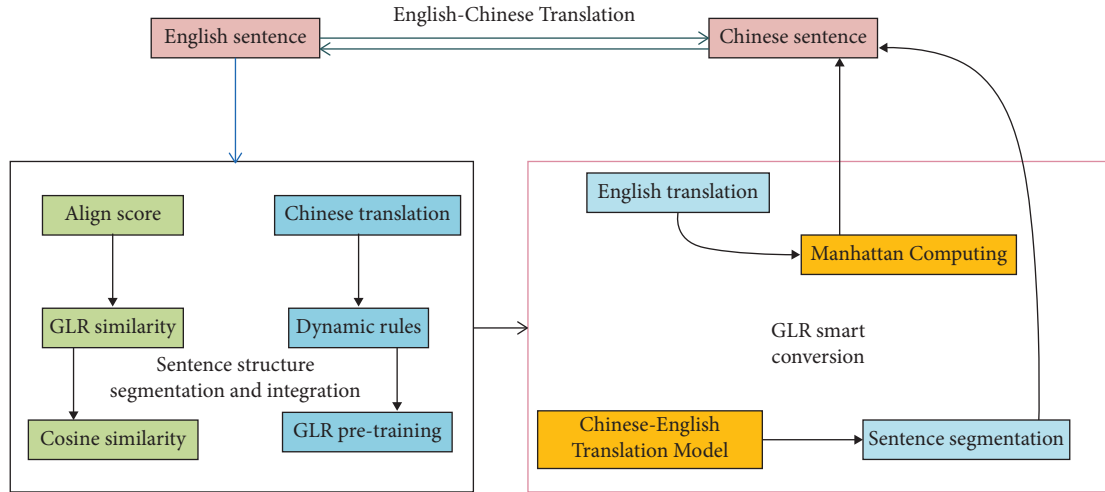


FIGURE 1: Model complexity and structural constraint complexity.

learning process, the computational efficiency of the unconstrained model is maintained under the premise of expectation constraints. Since the application of translation regularization is usually more direct, it is also more natural and easier in some cases. The structured and weak machine learning probabilistic translation regularization framework effectively integrates indirect learning by constraining the translation distribution of probabilistic models with potential variables. The improved method based on this method, unsupervised translation regularization, and Bayesian-based translation regularization are proposed one after another, and several kinds of algorithms are summarized in detail in the following.

As can be seen from Figure 2, the F1 value and recall rate of natural language processing methods are kept in a good level range, which also indicates that supervised machine learning technology has achieved great success in the fields of natural language processing, computer vision, and computational biology. Unfortunately, they often need to create large problem-specific training corpora and make the application of these methods expensive. In addition, it is often necessary to access some external problem-specific information that is not easy to use directly, such as training data with noise for different languages. Or there may be a domain expert to guide human learners, rather than simply creating an IID training corpus. A key challenge for weak machine learning is how to integrate the auxiliary information generated by indirect supervision effectively. It is very important to develop a new framework that can effectively integrate any prior knowledge into neural machine translation (NMT).

As shown in Figure 3, the corpus is preprocessed, the vector is generated by the encoder, the vector and the source language information are processed by the decoder, and the translation result of the target language is obtained. Traditional NMT models, whether based on recurrent neural network (RNN) or GLR, can be classified as embedded machine learning models. The autoregression here means that the model is translated word for word from left to right

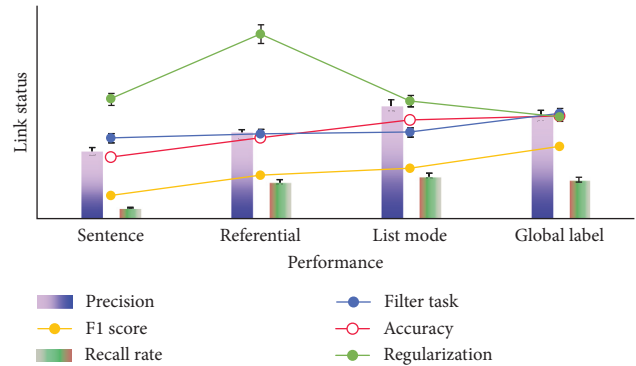


FIGURE 2: The cost of natural language processing methods.

when the target statement is translated, and the current submodel needs to rely on the last generated word, which leads to the fact that the encoder-decoder model can only translate word for word and the decoder cannot output in parallel. Compared with the autoregressive translation model, the non-autoregressive translation model overcomes the defect that the embedded machine learning model relies on context output. With the improvement of an encoder, the parallel output is realized, which greatly improves the generation rate.

As shown in Figure 4, with the help of crossword embedding and knowledge machine learning technology, English and Chinese corpora are processed accordingly to alleviate the dependence between source and target, data sparsity, and overfitting problems. The non-autoregressive translation model has both high accuracy and a high generation rate. Based on the improvement of the GLR translation model, this paper proposes the application of the non-autoregressive GLR model (GLR) in machine translation. Under the same conditions, compared with the GLR model, the GLR translation model is much faster than the GLR model in terms of generation rate. This is because the encoder is improved and a fertility module is added to count the occurrence times of each word in the source sentence,

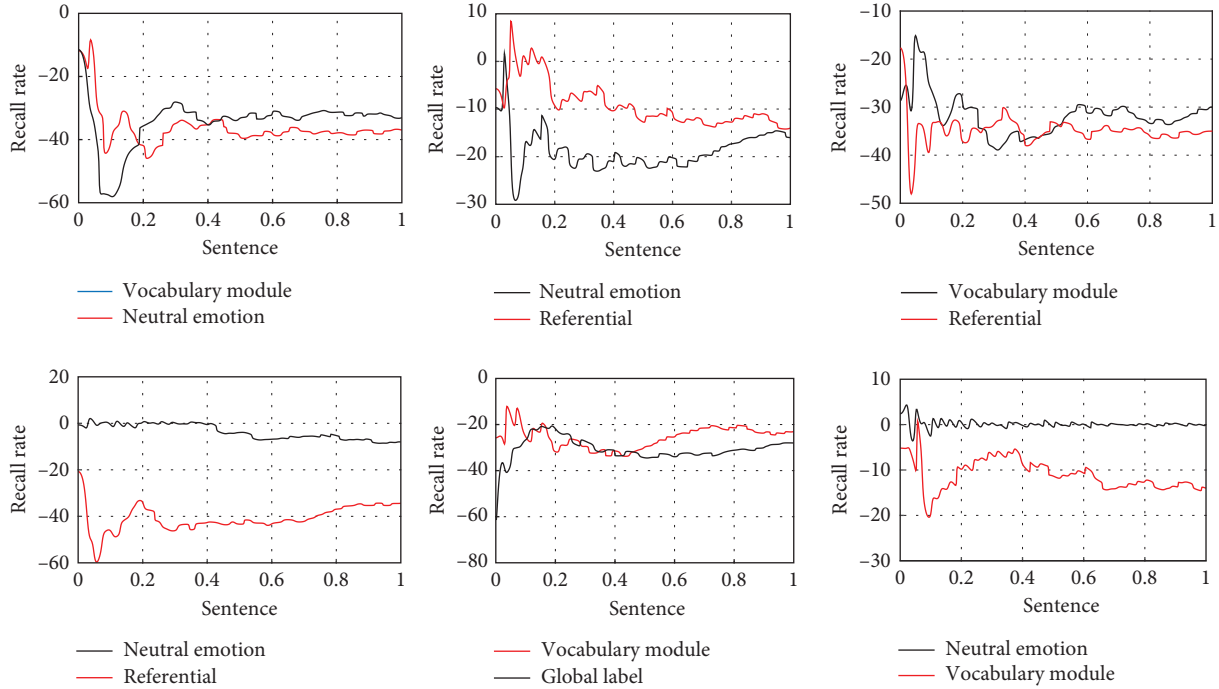


FIGURE 3: Encoder-decoder model generation vector.

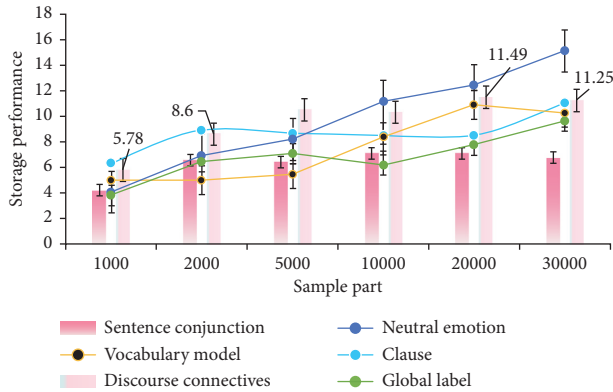


FIGURE 4: Alleviation of data sparseness and overfitting problems.

then the length of the target sentence is obtained, and the translation parallel output is realized by taking the fertility and the copy of the source sentence as the input of the decoder. However, it will also lead to the decline of translation accuracy. With the help of knowledge machine learning, the dependence between the source language and the target language will be reduced, and the translation effect and generation rate will be improved simultaneously.

As shown in Figure 5, we have 1.2 million English-Chinese parallel translations as experimental data and do a comparative experiment on general BPE processing and knowledge machine learning processing of English-Chinese corpus. It makes word vectors with similar meaning but from different languages has similar word vector representation, and narrows the representation gap between the two languages. The model can infer the meaning of words

with the help of each other's semantic information and transfer knowledge between bilinguals. The language with fewer data can infer with the help of a high resource corpus, to improve the effectiveness of the translation model. To map from the source space to the target common space, a common method is to learn a linear mapping to minimize the distance between the corresponding word pairs in bilingual dictionaries. The general approach is to train the bilingual corpus independently to get the embedded vectors of them and then train them to get a kind of linear mapping, minimize the distance between the equivalents listed in the bilingual dictionary, and get the general vector space.

As shown in Figure 6, the self-attention layer encoder-decoder model needs to generate an embedding vector for the input data. After the embedding vector is obtained, it is input to the coding layer. After the self-attention layer processes the data, it inputs the data into the feedforward neural network. The calculation process can be processed in parallel, and the output will be input to the next encoder. After the final encoding is completed, it is input into the decoding layer. After all the decoding layers are executed, a full connection layer and softmax layer are used at the end to get the corresponding word with the maximum probability value, which is our translation result.

As shown in Figure 7, all models are trained according to the environment and parameters, and the loss and BLEU values of the training set and verification sets are visualized. Although the loss of the baseline system decreases from 6 to 2 with the training of the model, the overall loss range is still greater than 2. After using the partition strategy, the loss effect of the interval submodel is the same, and the training loss of the interval submodel is the smallest. The training loss

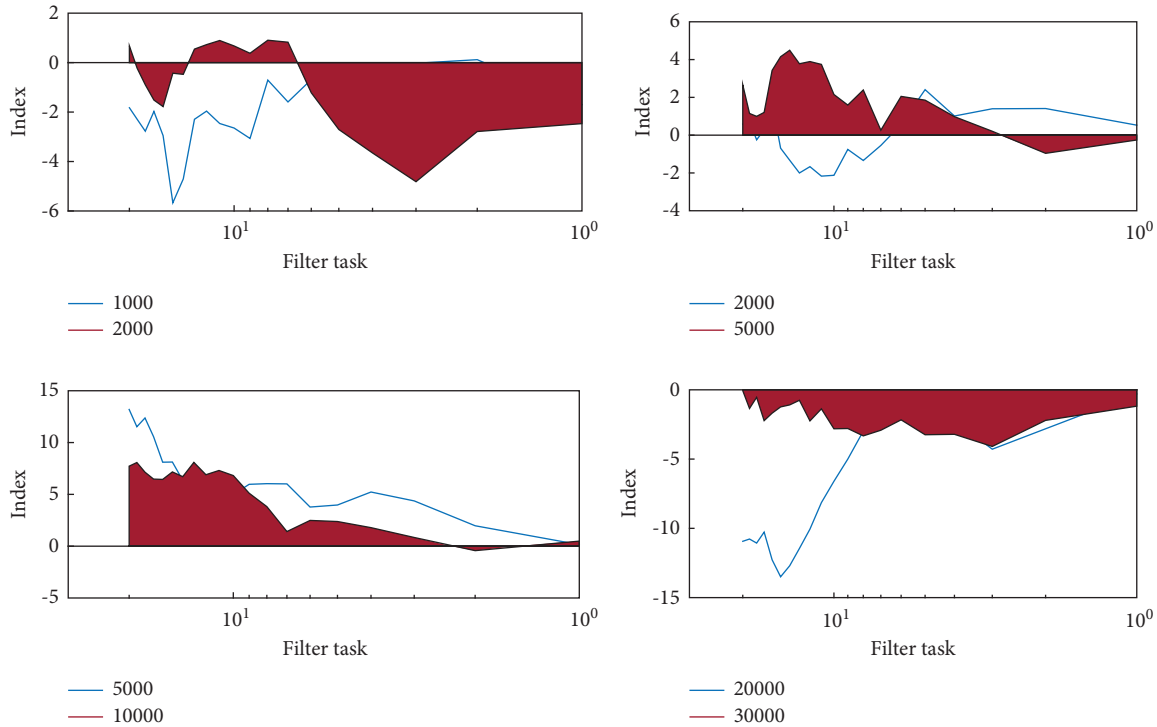


FIGURE 5: The distance between the equivalents listed in the bilingual dictionary.

of each submodel after the partition strategy is less than that of the benchmark system, which shows that the training of each submodel is better than that of the benchmark system.

3.2. GLR Translation Training. Training loss represents the training situation of the model, while BLEU value represents the efficiency and effect of training. As shown in Table 1, the training loss and BLEU value of the interval submodel fluctuates greatly, because the interval submodel generally contains a large number of short sentence, and with the increase of training times, the loss increases greatly, and the large attenuation of BLEU means that the interval submodel has the problem of overfitting. Although the submodel of interval does not change as much as the submodel of an interval, it can be observed that the BLEU value decays with the increase of training times. However, the two submodels of long sentence intervals show a better effect, and the submodel of the interval is better than other models in terms of loss and BLEU value.

Sentence length is reflected in the number of words in the data set. As shown in Table 2, sentence length does have a certain impact on training and testing. In the training set, the interval submodel with a large number of short sentences has a prominent effect, while in the verification set, the short sentence model will be overfitted due to too many training times, and the effect will decay seriously. The submodel with a large number of long sentence intervals has a good translation. In this paper, we use different sentence lengths to test different model translations, to improve the translation effect.

As shown in Figure 8, most of the translation models have high computational costs, and the time consumption increases rapidly while obtaining a high translation effect. The non-autoregressive GLR translation model can effectively solve this problem and shorten the time consumption, but it will lead to the decline of the translation effect. Therefore, this paper focuses on the GLR translation model, studies the role of cross-language word embedding and knowledge machine learning in the translation process, and makes a comparative analysis of the experimental results of each method. The experimental results show that the GLR translation model with knowledge machine learning processing can improve the effect of English translation compared with the GLR translation model while maintaining low time consumption. However, knowledge machine learning also has some shortcomings. Knowledge machine learning reduces the dependence between the source language and target language, which may cause the problem of insufficient semantic information in translation generation. Therefore, in the next study, we consider whether unsupervised knowledge machine learning method and source-target corpus alignment can be added to model training, and integrating more semantic information to further improve the effectiveness of the English translation is the focus of future work.

As shown in Figure 9, the characteristics of machine learning technology make it difficult to explain the result, which leads to the “trust” problem, which is contradictory to the high-reliability operation requirements of the GLR system. With the complexity of the current neural network architecture, the problem becomes more and more

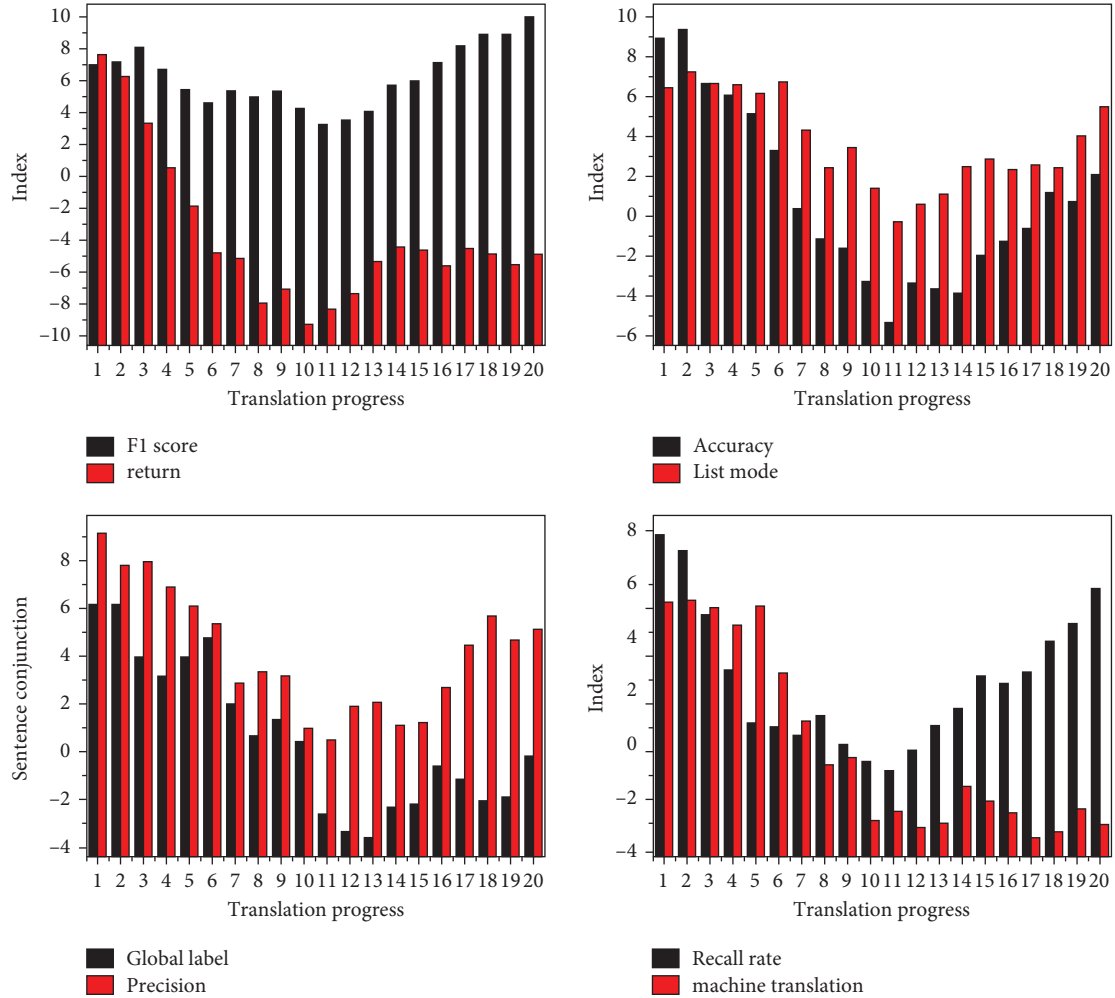


FIGURE 6: Self-attention layer encoder-decoder model.

prominent. For this reason, some classical machine learning techniques (such as decision tree and regression analysis) are still widely used in engineering practice. This is because the maturity, stability, and interpretability of these technologies are relatively high. Compared with the “black box” deep network, these technologies are friendly to operators, which is also one of the important differences between engineering practice and theoretical research. Therefore, in engineering practice, we should not blindly pursue the complexity of the model, we should consider the complexity of practical problems, the availability of data, and other factors, and reasonably select the machine learning model.

As shown in Figure 10, in addition to the interpretability problem, reinforcement learning technology continuously interacts with the environment to learn knowledge, which may bring large trial and error costs in the GLR system. In addition, the model training of reinforcement learning is time-consuming, parameter adjustment is difficult, and the generalization ability of the model is weak, which restricts its engineering application in the GLR system to a certain extent. At present, reinforcement learning technology is mainly used in the scenario of a high fault tolerance rate at the end of the power grid; even if there is a wrong decision, it

will not cause significant economic losses. But at the same time, it should be noted that compared with the traditional physical model, the reinforcement learning model has stronger adaptability to environmental changes, and it still has in-depth research value in the field of power distribution and consumption.

As shown in Figure 11, the GLR system operates in a stable state in most cases, with fewer fault states and corresponding fault sample data. And these few fault samples are often more valuable for research. On the contrary, machine learning technology usually needs a large number of sample data for training to learn the rules, which leads to the challenge of dealing with these “outliers.” Therefore, it is of great significance to study the technology of small samples and data enhancement. In addition to prediction technology, machine learning technology in the field of power distribution planning theoretical research and engineering applications are relatively small. Part of the reason is that the real-time requirement of the planning problem is low, the planners have enough time to carry out detailed simulation analysis and optimization model solving, and the dependence on machine learning technology is low. In addition, the planning problem requires the results to be strongly

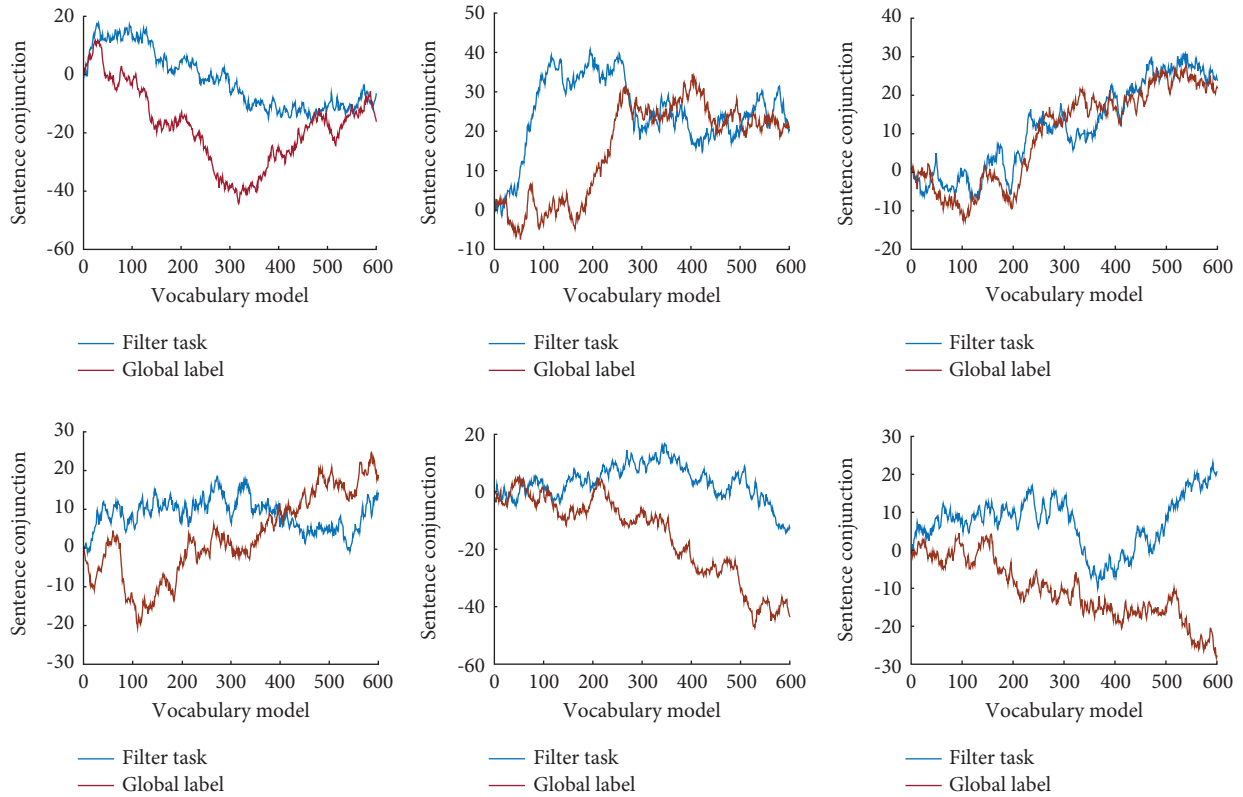


FIGURE 7: Model training and its loss value.

TABLE 1: Training loss and BLEU value.

Item	Accuracy	Precision	Recall rate	F1 score	Filter task	Regularization
Sentence	4.91	5.41	0.9	1.87	6.43	9.53
Referential	6.42	6.78	2.97	3.48	6.75	14.72
List mode	7.83	9.03	3.45	4.02	6.9	9.38
Global label	8.19	8.22	3.15	5.76	8.41	8.12
Translation	8.21	10.21	4.76	8.18	9.47	10.9
Return	9.14	12.65	6.87	7.78	8.78	8.22

TABLE 2: The impact of sentence length on training and testing.

Data set	Conjunction	Discourse connectives	Vocabulary	Neutral emotion	Global
1000	4.19	5.78	4.97	4.03	3.78
2000	6.57	8.6	4.98	6.88	6.41
5000	6.4	10.49	5.43	8.2	7.07
10000	7.08	10.32	8.37	11.16	6.15
20000	7.08	11.49	10.9	12.43	7.75
30000	6.74	11.25	10.24	15.14	9.62

interpretable, which also conflicts with the “black box” feature of machine learning. However, a small number of researchers have begun to focus on generating feasible feeder topology through machine learning technology to improve the efficiency of planning. This provides a new idea for research in the field of planning.

As shown in Table 3, when the fusion weight and threshold of BLEU value and cosine similarity are fixed, the accuracy rate, recall rate, and F1 value of Manhattan distance are in the fusion weight χ_1 , so the fusion weight parameter

value of Manhattan distance is 0.1. When the fusion weights of BLEU value, cosine similarity, and Manhattan distance are fixed, the accuracy rate, recall rate, and F1 value of the threshold intersect when the threshold tends to 0.9, and the accuracy rate and F1 value show an upward trend with the increase of the threshold, so the final threshold is set as 0.9. Among the three similarities, cosine similarity has the greatest impact on the quality of sentence alignment, followed by BLEU similarity, and Manhattan distance has the least impact.

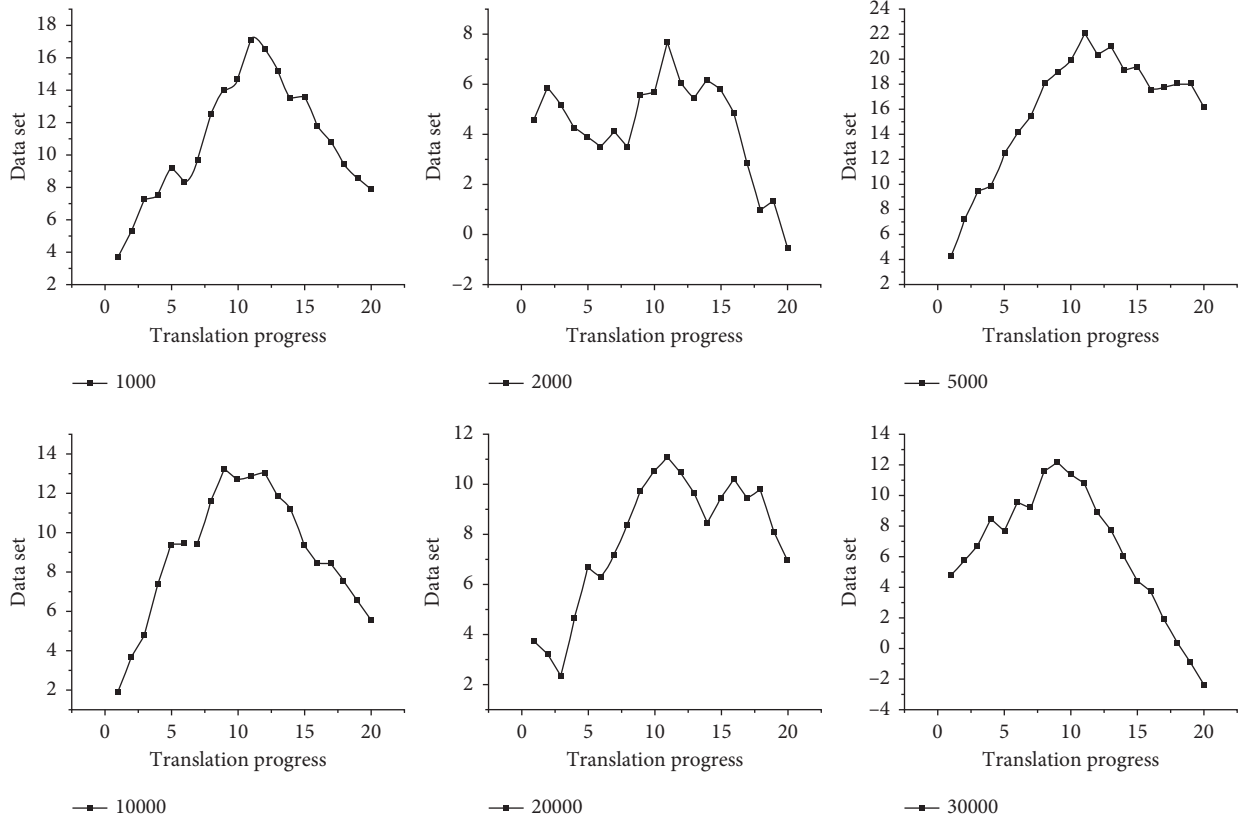


FIGURE 8: GLR translation model English translation effects.

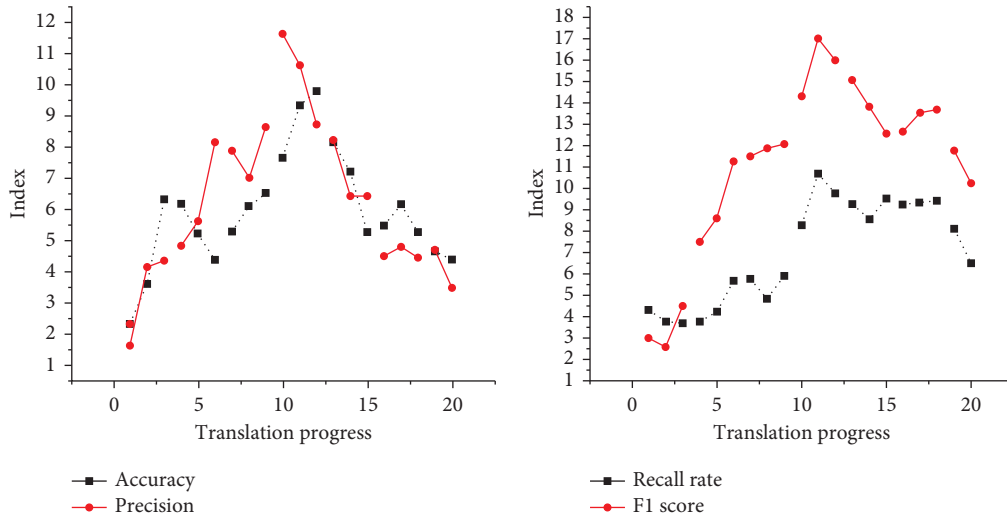


FIGURE 9: The characteristics of machine learning technology make its results.

As shown in Figure 12, this paper combines BLEU similarity, cosine similarity, and Manhattan distance to maximize the advantages of these three similarity measurement methods. For each similarity, the fusion weight of this paper is between 0.1 and 0.9. 90% noise data are added into 1000 English-Chinese parallel sentence pairs to construct a comparable corpus set. The proposed method is used to align sentences, and the optimal fusion weight and

filtering threshold of each similarity are compared. The noise data come from an English-Chinese data set.

As shown in Table 4, when the fusion weight and threshold of cosine similarity and Manhattan distance are fixed, the accuracy rate, recall rate, and *F1* value of BLEU value similarity are in the fusion weight α_7 , and *F1* value reaches the maximum value at this point, so the fusion weight parameter value of similarity is 0.7. When the fusion

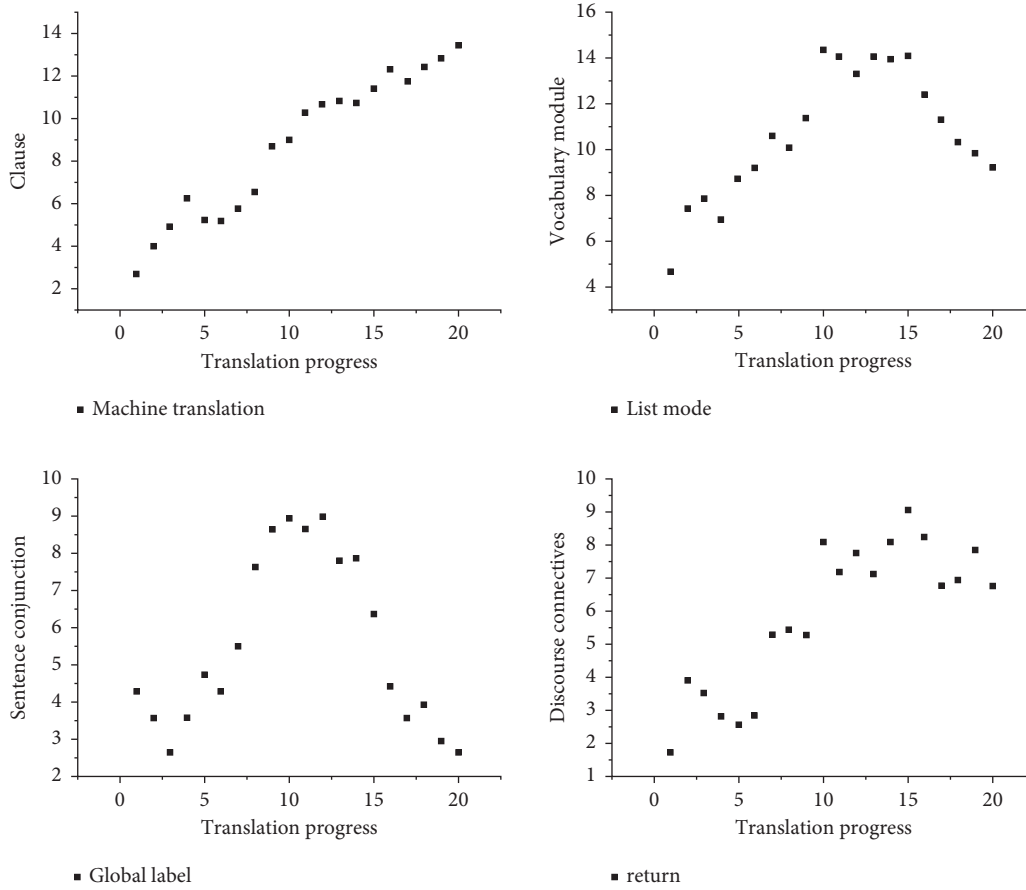


FIGURE 10: Reinforcement learning model training takes time.

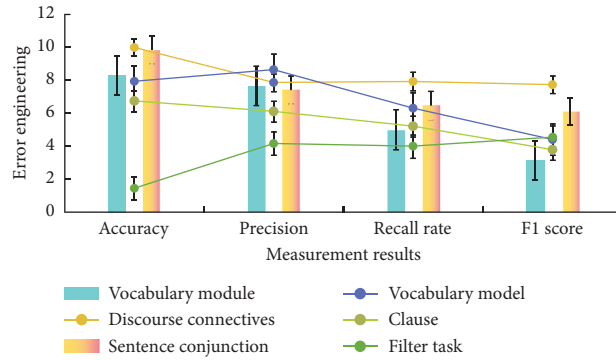


FIGURE 11: The real-time law of sample data planning problem.

TABLE 3: Cosine similarity and Manhattan distance fusion value.

Item	Vocabulary module	Sentence conjunction	Clause	Vocabulary	Filter task
Accuracy	8.28	9.84	6.73	7.92	1.41
Precision	7.66	7.4	6.09	8.63	4.14
Recall rate	4.99	6.46	5.18	6.29	3.98
F1 score	3.14	6.1	3.78	4.39	4.53

weight and threshold of BLEU value and Manhattan distance are fixed, the accuracy rate, recall rate, and F1 value of cosine similarity are higher in the fusion weight β . When it is between 0.8 and 0.9, it intersects, and the recall rate is

between 0.8 and 0.9 β , so the fusion weight parameter value of cosine similarity is 0.9.

Using any one or two methods of BLEU similarity, cosine similarity, and Manhattan distance, the experimental

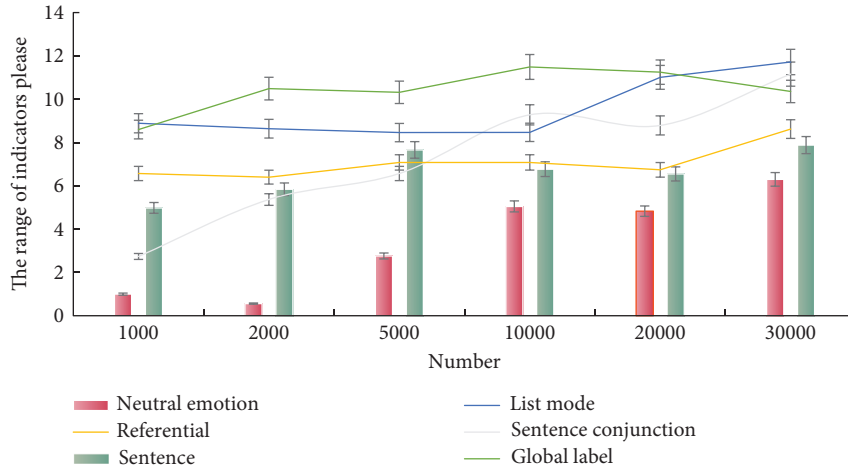


FIGURE 12: The advantages of the three similarity measurement methods.

TABLE 4: Similarity and Manhattan distance fusion weight and threshold.

Data set	Neutral	Conjunction	Sentence	Referential	List mode	Global
1000	1	2.74	4.98	6.57	8.89	8.6
2000	0.56	5.37	5.84	6.4	8.64	10.49
5000	2.76	6.57	7.66	7.08	8.46	10.32
10000	5.05	9.28	6.77	7.08	8.47	11.49
20000	4.83	8.79	6.55	6.74	11.01	11.25
30000	6.3	11.16	7.88	8.62	11.72	10.36

results are not as good as the three similarity methods based on Bert bidirectional fusion proposed in this paper, and the accuracy and $F1$ value of this method are higher than other similarity fusion methods. Experimental results show that compared with the single similarity calculation method between vectors, the proposed method effectively combines the advantages of three kinds of similarity. To evaluate the performance of the proposed method in filtering sentence pairs in the parallel translation library, a parallel translation filtering experiment is carried out in 1000 parallel sentence pairs through the proposed sentence alignment method and the classical sentence alignment method.

4. Discussion

This paper selects the two most classic sentence alignment methods as a baseline for comparison. Because the experiment is the filtering task of a parallel translation library, the accuracy rate is 100%, only the recall and $F1$ values are analyzed. The recall rate and $F1$ value of the BLEU align method are almost equal to the champion method. The recall rate and $F1$ value of the proposed method are 97.84% and 98.91%, respectively, in the parallel translation library filtering task because of the integration of multiple similarities and complementary advantages. The alignment effect of the proposed method is much better than that of the two baseline methods. The experimental results show that the proposed method calculates the similarity between sentence vectors from different perspectives so that more word-level information can be fully utilized in the calculation of

sentence pair similarity, and more high-quality bilingual parallel sentence pairs can be extracted from a large number of bilingual parallel translation. To evaluate the performance of the proposed method in filtering sentence pairs in comparable corpora, this experiment adds noise sentence pairs to 1000 parallel sentence pairs, so that the noise ratio is 20%, 50%, and 90%, respectively. The noise data come from Enzo sentence pair each. The proposed method is compared with the classical method in corpus filtering. In the comparable corpus filtering task, when the noise ratio is 20%, 50%, and 90%, the accuracy of the proposed method is higher than the BLEU align method and champion method. The alignment effect of our method is much better than that of the baseline method in the comparable corpus filtering task. Mining parallel sentence pairs from the large-scale comparable corpus and considering sentence similarity information from multiple perspectives and granularity can also extract high-quality bilingual parallel sentence pairs from a large number of the bilingual comparable corpus.

In this paper, the bilingual dictionary is used to improve the learning ability of the model for two languages, and the bilingual word vector is introduced for antagonistic learning to achieve bilingual alignment in the same semantic space. The experimental results show that this method can improve the effect of low resource cross-language translation. In the future research, we will continue to explore better alignment methods between Chinese and English bilinguals, combine multi-language Bert and multi-language Bart models to achieve cross-language translation tasks, and improve cross-language translation performance.

For example, reasonable initialization and complex data transformation are often used to affect the potential translation distribution of the model. A key challenge of structured and weak machine learning is to develop a flexible and declarative framework to express the structural constraints of potential variables generated by prior knowledge and indirect supervision. A structured model can capture a very rich array of possible relationships, which usually leads to inference problems that are difficult to deal with. In recent years, Bayesian nonparametric methods have relaxed some unrealistic assumptions about data, such as homogeneity and exchangeability. For example, to deal with heterogeneous observations, a prediction-dependent process is proposed; to relax the interchangeability hypothesis, stochastic processes with various related structures have been successfully proposed, such as hierarchical structure, time or space dependence, and random sequence dependence. The common feature of all these methods is that they rely on definitions or in some cases encode some special structures to learn nonparametric Bayesian priors. According to Bayesian rules, this method will indirectly affect the translation distribution of the model through the interaction of likelihood models. The translation regularization framework separates the model from the complex structural constraints. However, different from the parameter regularization in the Bayesian framework, this method also restricts the correlation of the data. These constraints are easy to be used as the model translation information coding on the observation data, but it is difficult to be used as the model parameter information coding through Bayesian prior. Through decomposable regularization of translation moments of potential variables in the learning process, the computational efficiency of the unconstrained model is maintained under the premise of expectation constraints.

5. Conclusions

In this paper, in machine-free learning problems, data have the structure of order, recursion, space, relation, and other types. We often use the structural statistical model with potential variables to obtain the potential dependence and realize the recognition and induction of semantic categories. Unsupervised part of speech and grammar induction, and word and phrase alignment of statistical machine translation in natural language processing are examples. Usually, the hidden variables are first marginalized, and then the observation data are estimated by the expectation-maximization algorithm. Considering the computational and statistical factors, the generative model used in practice is usually a very simple model to estimate the potential structure, such as the syntactic structure of a language or the process of language translation. However, a fatal problem of this kind of model is that marginal likelihood may not guide the model toward the expected results of potential variables, but focus on explaining the correlation in the data. Since the model mainly focuses on learning the distribution of potential variables, it is hoped that they can capture the expected rules without direct learning. Therefore, how to control the potential distribution is very important.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares no conflicts of interest in this study.

References

- [1] D. Chiang, "Hierarchical phrase-based translation," *Computational Linguistics*, vol. 33, no. 2, pp. 201–228, 2007.
- [2] S. Bai, "Research on machine translation of Japanese - Mongolian phrasal verbs based on derived grammar," *Journal of Chinese Information Processing*, vol. 22, no. 2, pp. 47–54, 2018.
- [3] Z. Ren, H. Hou, and Y. Ji, "The application of subword granularity segmentation in Mongolian and Chinese neural machine translation," *Journal of Chinese Information Processing*, vol. 33, no. 01, pp. 90–97, 2019.
- [4] Y. Su, N. Wu, and W. Liu, "Machine translation of Mongolian and Chinese natural language based on statistical analysis," *Journal of Beijing University of Technology*, vol. 43, no. 1, pp. 36–42, 2019.
- [5] J. Li, H. Liu, M. H. Raval et al., "Structure of the MORN4/Myo3a Tail Complex Reveals MORN Repeats as Protein Binding Modules," *Structure (London, England: 1993)*, vol. 27, no. 9, pp. 1366–1374.e3, 2019.
- [6] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks, Proceeding of NIPS 27," *Montreal*, vol. 2, no. 4, pp. 3104–3112, 2019.
- [7] A. Vaswani, N. Shazeer, and N. Parmar, "Attention is all you need," *Proceeding of Advances in Neural Information Processing Systems*, vol. 7, no. 4, pp. 5998–6008, 2019.
- [8] T. Luong, H. Phan, and C. D. Manning, "Bilingual word representations with monolingual quality in mind," *Proceedings of the 1st Workshop on vector space modeling for natural language processing*, vol. 6, no. 4, pp. 151–159, 2018.
- [9] K. Yoon and R. Alexander, "Sequence-level knowledge distillation," *Proceeding of EMNLP*, vol. 7, no. 5, pp. 1317–1327, 2019.
- [10] D. Chris, C. Victor, and S. Noah, "A simple, fast, and effective reparameterization of IBM Model 2," *The North American Chapter of the Association for Computational Linguistics*, vol. 9, no. 45, pp. 644–648, 2019.
- [11] S. Rico, H. Barry, and B. Alexandra, "Neural machine translation of rare words with subword units," *Proceeding of the 54th Annual Meeting of the Association for Computational Linguistics*, vol. 6, no. 4, pp. 1715–1725, 2019.
- [12] A. Mikel, L. Gorka, and A. Eneko, "Learning bilingual word embeddings with (almost) no bilingual data," *Proceeding of the 55th Annual Meeting of the Association for Computational Linguistics*, vol. 1, no. 5, pp. 451–462, 2019.
- [13] H.-B. Liu and I. Lee, "Bridging the gap between training and inference for spatio-temporal forecasting," *Proceeding of the 57th Annual Meeting of the Association for Computational Linguistics*, vol. 2, no. 4, pp. 4334–4343, Florence, Italy, 2019.
- [14] C. Zhou, N. Graham, and J. Gu, "Understanding knowledge distillation in non-autoregressive machine translation," *Proceeding of ICLR*, vol. 7, no. 4, pp. 186–195, 2020.
- [15] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," *Proceedings of the 54th Annual Meeting of the Association for Computational*

- Lin-guistics (Volume1: Long Papers)*, vol. 3, no. 1, pp. 1715–1725, 2019.
- [16] Ru Xu and B. Zeng, “Convolutional self-attention encoding for reinforced automatic summarization model,” *Journal of Chinese Computer Systems*, vol. 41, no. 2, pp. 271–277, 2020.
 - [17] J. Wu, B. Cao, and J. Fan, “Event summarization method combined with bigram semantic extension,” *Journal of Chinese Computer Systems*, vol. 40, no. 7, pp. 1380–1385, 2019.
 - [18] A. Leuski, C. Y. Lin, L. Zhou, U. Germann, F. Och, and E. Hovy, “Cross-lingual CSTRD: English access to Hindi information,” *ACM Transactions on Asian Language Information Processing*, vol. 2, no. 3, pp. 245–269, 2003.
 - [19] Ayana, S. q. Shen, Y. Chen, C. Yang, Zy Liu, and Ms Sun, “Zero-shot cross-lingual neural headline generation,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 12, pp. 2319–2327, 2018.
 - [20] J. Zhu, Q. Wang, and Y. Wang, “NCLS: neural cross-lingual summarization,” *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, vol. 9, no. 14, pp. 3045–3055, 2019.
 - [21] S. Ilya, Vinyals Oriol, and Le V. Quoc, “Sequence to sequence learning with neural networks,” *Proceedings of Advances in Neural Information Processing Systems*, vol. 27, no. 4, pp. 3104–3112, 2019.
 - [22] Q. D. Nguyen and T. A. Nguyen, “PhoBERT: pre-trained language models for Vietnamese,” *Proceedings of Findings of the Association for Computational Linguistics (EMNLP)*, vol. 6, no. 4, pp. 1037–1042, 2020.
 - [23] V. Ashish, S. Noam, and N. Parmar, “Attention is all you need,” *Proceedings of Advances in Neural Information Processing Systems*, vol. 4, no. 15, pp. 5998–6008, 2019.
 - [24] B. Hu, Q. Chen, and F. Zhu, “LCSTS: a large scale Chinese short text summarization dataset,” *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, vol. 9, no. 45, pp. 1967–1972, 2019.
 - [25] T. Vu, N. Q. Dat, and N. Q. Dai, “VnCoreNLP: a Vietnamese natural language processing toolkit,” *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, vol. 5, no. 3, pp. 56–60, 2018.
 - [26] K. Shankar, M. Elhoseny, S. K. Lakshmanaprabu, I. M. V. Rm, and M. Alkhambashi, “Optimal feature level fusion based ANFIS classifier for brain MRI image classification,” *Concurrency and Computation: Practice and Experience*, vol. 32, no. 1, Article ID 24887, 2020.
 - [27] M. Elhoseny, “Multi-object detection and tracking (MODT) machine learning model for real-time video surveillance systems,” *Circuits, Systems, and Signal Processing*, vol. 39, no. 2, pp. 611–630, 2020.
 - [28] M. Ali, L. T. Jung, A.-H. Abdel-Aty, M. Y. Abubakar, M. Elhoseny, and I. Ali, “Semantic-k-NN algorithm: an enhanced version of traditional k-NN algorithm,” *Expert Systems with Applications*, vol. 151, Article ID 113374, 2020.