

Retraction

Retracted: Deep Learning-Based English-Chinese Translation Research

Advances in Meteorology

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

Copyright © 2023 Advances in Meteorology. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Y. Huang and Y. Xin, "Deep Learning-Based English-Chinese Translation Research," *Advances in Meteorology*, vol. 2022, Article ID 3208167, 9 pages, 2022.

Research Article

Deep Learning-Based English-Chinese Translation Research

Yao Huang¹ and Yi Xin² 

¹Jiangxi Electric Vocational & Technical College, Nanchang, China

²Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, China

Correspondence should be addressed to Yi Xin; n830025116@mail.uic.edu.cn

Received 11 March 2022; Revised 13 April 2022; Accepted 17 May 2022; Published 14 July 2022

Academic Editor: Victor S. Sheng

Copyright © 2022 Yao Huang and Yi Xin. 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.

Neural machine translation (NMT) has been bringing exciting news in the field of machine translation since its emergence. However, because NMT only employs single neural networks to convert natural languages, it suffers from two drawbacks in terms of reducing translation time: NMT is more sensitive to sentence length than statistical machine translation and the end-to-end implementation process fails to make explicit use of linguistic knowledge to improve translation performance. The network model performance of various deep learning machine translation tasks was constructed and compared in English-Chinese bilingual direction, and the defects of each network were solved by using an attention mechanism. The problems of gradient disappearance and gradient explosion are easy to occur in the recurrent neural network in the long-distance sequence. The short and long-term memory networks cannot reflect the information weight problems in long-distance sequences. In this study, through the comparison of examples, it is concluded that the introduction of an attention mechanism can improve the attention of context information in the process of model generation of the target language sequence, thus translating restore degree and fluency higher. This study proposes a neural machine translation method based on the divide-and-conquer strategy. Based on the idea of divide-and-conquer, this method identifies and extracts the longest noun phrase in a sentence and retains special identifiers or core words to form a sentence frame with the rest of the sentence. This method of translating the longest noun phrase and sentence frame separately by the neural machine translation system, and then recombining the translation, alleviates the poor performance of neural machine translation in long sentences. Experimental results show that the BLEU score of translation obtained by the proposed method has improved by 0.89 compared with the baseline method.

1. Introduction

Machine translation has been touted as a way to break down language barriers since the beginning of computers. Its fundamental concept is to employ the computer's processing power and storage capacity to assist or replace translators in accomplishing difficult translation jobs, resulting in automatic language conversion [1]. People are accustomed to calling the language before translating the source language, and the translated language is therefore the target language.

Machine translation is one of the most beautiful ideas of scientists and even mankind since the birth of computers. Under the urgent needs of world such as economic development and cultural exchange, machine translation is

becoming more and more vibrant. Machine translation has experienced a tortuous development path, and numerous philosophers have devoted their efforts to it and made great progress. However, because of the current situation, there are still many problems to be solved in machine translation. The research of neural machine translation represents the combination of cognitive science and artificial intelligence. It uses the sequential to sequence model to transform the complex translation process into end-to-end, and at the same time, it also drives the research progress of other sequential to sequence tasks.

Machine translation has both theoretical and practical values and has experienced considerable development since its inception. At present, the application of deep learning methods makes neural machine translation mainstream. The

neural machine translation provides the following advantages over statistical machine translation:

- (1) End-to-end learning does not depend on too many prior assumptions. In the era of statistical machine translation, model design makes more or fewer assumptions about the process of translation. The phrase-based models, for example, assume that both the source and target languages are sliced into sequences of phrases, with some alignment between them. This hypothesis has both advantages and disadvantages. On the one hand, the hypothesis draws on the relevant concepts of linguistics and helps the model to integrate them into prior human knowledge. On the other hand, the more assumptions there are, the more constrained the model is. If the assumptions are correct, the model can describe the problem well. But if the assumptions are wrong, then the patterns can be biased. Deep learning does not rely on prior knowledge nor does it require the manual design of features. The model learns directly from the mapping of input and output (end-to-end learning), which also avoids possible deviations caused by assumptions to a certain extent.
- (2) The continuous space model of neural networks has a stronger representation ability. A basic problem in machine translation is how to represent a sentence. The statistical machine translation regards the process of sentence generation as the derivation of phrases or rules, which is essentially a symbol system in discrete space. Deep learning transforms traditional discrete-based representations into representations of continuous space. For example, a distributed representation of the real number space replaces the discrete word representation, and the entire sentence can be described as a vector of real numbers. Therefore, the translation problem can be described in continuous space, which greatly alleviates the dimension disaster of the traditional discrete space model. More importantly, the continuous space model can be optimized by gradient descent and other methods, which have good mathematical properties, and is easy to implement.

However, the effect of machine translation is far from reaching that of human translation, and neural machine translation still has a long way to go. At present, the research of neural machine translation still faces many challenges and needs to be improved. Therefore, the research on neural machine translation has a high academic significance.

With the advent of personal computers and the shift of translation memory tools for translators, machine translation has been applied in practice. In recent years, the trend is to combine some of the traditional methods of statistical machine translation (SMT) with those of neural networks or purely based on neural networks. In particular, the neural network-based end-to-end machine translation, which abandons the complicated procedures of traditional

statistical machine translation, directly uses parallel corpus as input and output for end-to-end training, can better deal with the problem of long-distance dependence, and has become the mainstream of major companies and institutions. It has become a major research direction for decades to come.

Theoretically, machine translation involves many disciplines and is a typical interdisciplinary research topic. Research on machine translation is of great theoretical significance and can help promote the development of linguistics, computational linguistics, artificial intelligence, machine learning, and even cognitive linguistics. In addition, the research on machine translation can facilitate the development of other natural language processing tasks, such as named entity recognition, emotion analysis, automatic text generation, and so on. In terms of application, no matter the public, government enterprises, or national institutions, machine translation technology is urgently needed. Machine translation plays an important role in many fields.

Looking back on the history of translation, manual translation has been the mainstream way of translation since ancient times. However, with the growing maturity of computer technology and the rapid development of the Internet, machine translation technology has gradually entered the stage of history. Machine translation (MT), as the name implies, is a technology that uses the efficient computing power of computers to convert and express information between two languages [2]. Up to now, the development of MT technology has stepped from the stage of academic research to the stage of practical application. Machine translation is also the main force to promote the vigorous development of translation in the world.

However, since NMT only uses a single neural network to translate the natural language, it has two disadvantages in reducing translation time: NMT is more sensitive to sentence length than statistical machine translation and the end-to-end implementation process is not clear. Use language knowledge to improve translation performance. We constructed the network model performance of various deep learning machine translation tasks and compared them in English and Chinese bilingual directions. The attention mechanism was used to solve the problem of gradient disappearance and gradient explosion in the recurrent neural network of long-distance sequences, which alleviated the problem of neural machine with the problems of translations that do not perform well in long sentences.

2. Related Work

The history of machine translation dates all the way back to the late ninth century. Arabian cryptographers invented the systematic language translation techniques used in modern machine translation, such as frequency analysis, probability and statistical information, and cryptanalysis [3]. Rene Descartes proposed a universal language in the late 1620s, which gave rise to the concept of machine translation [4, 5]. In 1956, the first conference on machine translation was held, ushering in a new era of machine translation research.

Since then, scientists from all over the world have begun to investigate the machine translation technique. Although the Association for Machine Translation and Computational Linguistics (AMTCL) and the Advisory Committee on Automatic Language Processing (ALPAC) were founded concurrently in the United States, machine translation technology has made few advances in the decade since.

In 1972, the Defense Research and Engineering Agency announced the successful translation of an English military manual into Vietnamese using its Logos machine translation system, reestablishing the feasibility of machine translation. Following this, several researchers made significant advancements in machine translation technology. In the late 1980s, the advancement of computer hardware quality resulted in a decrease in computing costs, and the emergence of various machine translation methods signaled the rapid advancement of machine translation, as well as the beginning of various machine translation contests. Machine translation evolved gradually from academic research to practical applications. From its inception to its development, machine translation has grown in popularity due to its tremendous research potential and commercial value.

Hinton et al. proposed deep learning in 2006 [6], and since then, it has been quickly growing in the fields of image processing [7] and speech recognition [8]. In recent years, researchers have discovered that deep learning can help with statistical machine translation problems such as linear inseparability, a lack of proper semantic representation, difficulty in designing characteristics, making full use of the local context, data sparsity, and error propagation [9]. Thus, the machine translation research based on deep learning has become a hot topic of research.

At the moment, the combination of deep learning and machine translation mainly exists in two forms: first, take the advantages of deep learning and makeup statistics on the shortcomings of machine translation methods in key modules such as the language model [10], translation model [11], reordering model [12], and word alignment [13]; second, take the advantages of deep learning and makeup statistics on the shortcomings of machine translation methods in key modules such as the language model, translation model, reordering model, and word alignment. For example, the deep learning pioneer and the professor of the University of Montreal, Yoshua Bengio created a neural network-based language model in 2003, which effectively solves the problem of data sparsity that plagues the traditional methods. The neural machine translation (NMT) is totally based on the deep learning from start to finish. Machine translation professionals and scholars have given close attention to it because of its simple approach and unique methods, as well as its capacity to produce translations that are comparable to or even better than traditional methods. The end-to-end neural machine translation has come a long way in just a few years because of the combined efforts of academics from all over the world. It not only pioneered theories and methodologies but also outperformed competitors in

bilingual translation tasks like English-French, English-German, and Chinese-English [14].

Baidu claimed to have released the world's first neural network translation system as early as 2015, and with the real development of neural network machine translation taking only one or two years, neural network machine translation quickly swept academic and industrial circles. In 2017, Brown et al. [15], a Google brain team researcher, proposed the transformer model, an entirely attention-based model. This is a novel architecture that is entirely based on the attentional mechanism and does not make use of recurrent neural networks. This has the advantage of allowing for direct extraction of global connections, as the attentional mechanism assumes that the distance between each input and output word is equal to one, whereas the recurrent neural network requires gradual recursion [16]. The transformer not only outperformed other existing algorithms in machine translation tasks but also significantly outperformed them in other language comprehension tasks. This model, on the other hand, performs poorly on smaller, more structured language comprehension tasks or on simple algorithmic tasks. As a result, Brown et al., a Google researcher, focused on the shortcomings of the transformer and extended it into a general computing model with a new and efficient time-parallel loop that outperformed the standard transformer in a wide variety of algorithm tasks as well as in a large number of large-scale language understanding tasks [17, 18].

3. Basic Principles of Deep Learning Translation

The language model is the basis of natural language processing. Its function is to express natural language into a mathematical form that can be processed by the computer. In 2003, Koehn et al. [19] first used a neural network to build a language model, which became the beginning of the application of neural networks in the field of natural language processing. Let V be the thesaurus, and for sentence $Y = \{y_1, y_2, \dots, y_T\}$, the language model is approximately equal to a series of parameters θ such that the following formula is satisfied.

$$P(Y; \theta) = P(y_1, y_2, \dots, y_T; \theta). \quad (1)$$

The autoregressive language model is the essence of neural machine translation (NMT). In regressive language generation, the output words are decoded one by one, and the previous decoding results are used as input for decoding the current words. The mathematical model thus constructed is

$$P(Y|X; \theta) = \prod_{t=1}^{T+1} P(y_t|y_{0-t-1}, X; \theta). \quad (2)$$

In general, the maximum likelihood estimation method is used to train the autoregressive machine translation model, with the cross-entropy loss as the loss function. The following maximizing formula is used in the neural machine translation training process, where N is the number of parallel corpus pairs.

$$L^{ML}(\theta) = \frac{1}{N} \sum_{n=1}^N \sum_{t=1}^{T+1} \log p(y_t^n | y_{0 \sim t-1}^n; \theta). \quad (3)$$

Greedy search means that when the model is decoded word by word, each word is selected with the maximum probability. If the output sequence of the decoder is $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_T\}$, the process of yt output by greedy search of the decoder at time T is shown in the following formula, where V represents the target language word list.

$$\hat{y}_t = \arg \max_{y \in V} \log p(y | \hat{y}_{0 \sim t-1}, x_{1 \sim T}; \theta). \quad (4)$$

The beam search can alleviate the above problems to a certain extent and make the translation results closer to the global optimal. Different from greedy search, which selects the word with the largest generation probability at every moment, cluster search first determines a beamwidth, saves the results of the number of cluster widths in each decoding process, and selects them in order of probability. In essence, cluster search may be similar to greedy search, but cluster search expands the search space by caching candidate sequences of cluster width and selecting the one with the highest comprehensive probability as the output, thereby making the translation results more diverse and approaching the global optimal. Combined with the following formulas, the process of cluster search decoding is briefly explained here:

$$\begin{aligned} C_{t-1} &= \{\tilde{y}_{0 \sim t-1}^{(1)}, \tilde{y}_{0 \sim t-1}^{(2)}, \dots, \tilde{y}_{0 \sim t-1}^{(K)}\}, \\ C_{t-1} &= \{\tilde{y}_{0 \sim t-1}^{(1)}, \tilde{y}_{0 \sim t-1}^{(2)}, \dots, \tilde{y}_{0 \sim t-1}^{(K)}\} \\ &= \text{argsort} \sum_{t=0}^t \log p(y_t | y_{0 \sim t-1}, x_{1 \sim T}; \theta), \quad (5) \\ \hat{Y} &= \arg \max_{Y_1, \dots, Y_K} \left(\frac{1}{Y}\right)^\alpha \log p(Y|X; \theta), \end{aligned}$$

where C_{t-1} is the input sentence, and \hat{Y} is the output sequence of the decoder.

Figure 1 shows the machine translation process of a sequential model consisting of a two-layer cyclic network. It can be seen that words or sentences are parsed into vectors in a continuous space in the whole process. In this way, semantic connections between the words can be grasped, reflecting the advantages of neural networks.

The group in embedding is to divide each word in a sentence into groups according to certain or multiple grouping methods, after which each word in the sentence has its corresponding one or more groups.

Set definition language based on minimum unit WordUnit = {word|char|letter|subword}, in which the smallest unit of English can be a word, letter, or subword (subword part of words can have an independent meaning, such as reproducing the "re"). The smallest unit of Chinese is a word or character. The smallest units of processing covered in this article are words. If each word in a sentence is divided into groups in a certain way, then all possible corresponding

groups form a set GroupUnit = {group}. If not limited to the classification principle of one group, all possible corresponding groups can form multiple group units, namely, GroupUnit1 and GroupUnit2.

Suppose that the sentence S after the participle can be expressed as an ordered set, namely:

$$S = \{w_1, w_2, \dots, w_t\}. \quad (6)$$

The $w_i \in \text{WordUnit}$, ($i = 1, 2, \dots, T$), for each corresponding word in the sentence.

If the division mode of pair group is f , then

$$f(S) = \{g_1, g_2, \dots, g_t\}. \quad (7)$$

$G_i \in \text{GroupUnit}$, ($i = 1, 2, \dots, T$), for each word in the sentence. Its subscripts correspond to the words in the original sentence S .

The so-called embedding is the distributed representation of one_hot into a multidimensional continuous vector. The set of embeddings corresponding to different words represented by one_hot in WordUnit is

$$E = \{e_1, e_2, \dots, e_t\}, \quad (8)$$

where $E_i \in E$ is a vector with dimension m . Each word W_i has a corresponding m -dimensional continuous vector representation E_i , which is commonly known as word embedding.

The only difference between the deep learning model using group embedding and the common deep learning model is that the group embedding information and the word embedding of the original corpus are used as the input of the model. The input of the ordinary model without group embedding g is the word embedding of each word in the sentence, namely:

$$\text{Input} = \{e_1, e_2, \dots, e_t\}. \quad (9)$$

The input of the model using group embedding is

$$G_input = \{C(e_1, g_{-e_1}), C(e_2, g_{-e_2}), \dots, C(e_t, g_{-e_t})\}. \quad (10)$$

Functions such as Sigmoid have many advantages and are very suitable to be used as activation functions. Therefore, under the condition of not changing the activation function, the mean square error function can be replaced and cross-entropy can be used as the loss function. If cross-entropy is used as the loss function, the loss function for a certain word translation can be

$$C_i(\theta) = - \sum_{x_i} [y(x_i) \ln a + (1 - y(x_i)) \ln (1 - a)]. \quad (11)$$

In comparison to the mean square error function, cross-entropy training has a number of advantageous properties and has been widely used to train a variety of models.

After preprocessing, each sentence with translation will have an end-of-sentence character. The translation is complete when the end-of-sentence character is encountered. The true length of a sentence is the distance between the first and last characters. Then, the result of the position decoding unit will

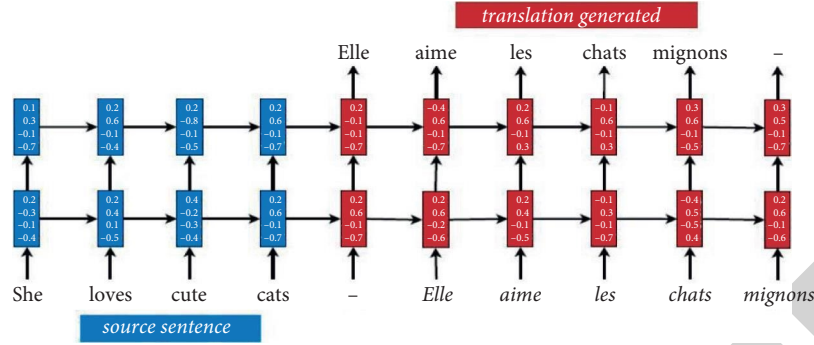


FIGURE 1: Sequence model of the two-layer cyclic network.

TABLE 1: Parameter settings.

The parameter types	Parameter
Vocabulary size	8000
Word vector dimension	512
Batch size	1024
Network layer	6
Dropout	0.2
Vector	0.1
The optimizer	Adam

TABLE 2: BLEU-4 score table of the translation model.

Model	Val	Test 1	Test 2
The mavericks translation	22.34	22.02	20.76
RNN	22.94	22.22	20.88
LSTM	23.56	22.7	22.57
BiRNN	22.37	22.25	22.05
RNN + attention	25.44	25.24	24.56
LSTM + attention	26.32	25.84	25.69
BiRNN + attention	25.6	25.43	25.52
Transformer	28.38	27.92	27.2

undergo masking, and its loss function value will be excluded from the final sentence's loss function value. The loss function for the entire sentence that needs to be translated is as follows:

$$C = \frac{1}{t} \sum_{i=1}^t C_i(\theta). \quad (12)$$

In the process of neural network training, multiple sentences are translated at a time, that is, one batch of sentences. Then, the loss function of a translation process can be expressed as

$$\text{Loss} = \frac{1}{n} \sum_n C. \quad (13)$$

As we know from the previous introduction, the loss function of the sentence to be translated is

$$C = \frac{1}{t} \sum_{i=1}^t w_i C_i(\theta). \quad (14)$$

To set the different weight of loss functions for words at different positions during training, a constant greater than 1 is introduced as a weight attenuation factor. Assuming that the sentence target translation length is T , the loss weight of the i^{th} word is

$$\text{Weight} = \frac{1}{\text{factor} + i - 1}. \quad (15)$$

4. Experimental Results and Analysis

Based on the principle of comparative experiments on different network models, the software and hardware environment and other experimental parameters were set as invariable

TABLE 3: Word list size comparison experiment.

Vocabulary size	Val	Test 1	Test 2
8000	28.38	27.92	27.2
16000	20.49	20.05	29.27
32000	26.58	26.22	25.56

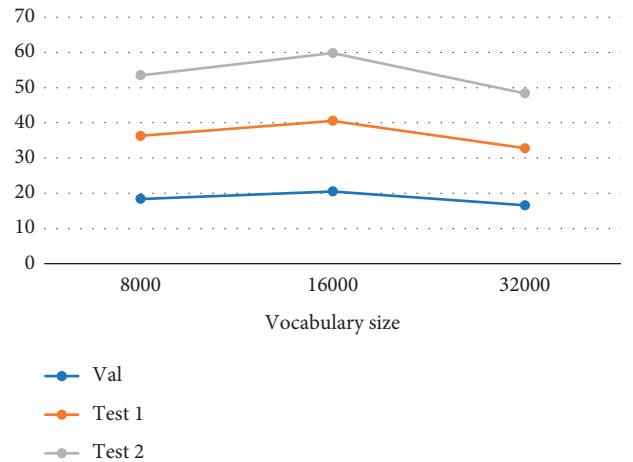


FIGURE 2: BLEU-4 line chart.

TABLE 4: Experimental results of machine translation based on the attenuation weight loss function model.

	Baseline	Factor = 2	Factor = 3	Factor = 4
BLEU score	27.24	27.42	27.42	27.33
Improve the ratio	—	2.63%	2.63%	2.22%

TABLE 5: Experimental results of Chinese sentiment analysis.

Training set size	1000		500		300	
Methods	The part of speech is embedded	Baseline	The part of speech is embedded	Baseline	The part of speech is embedded	Baseline
Train the number of iterations to the optimal	5	12	8	11	13	19
The identification accuracy	79.97%	94.09%	91.53%	90.11%	92.1%	79.97%

conditions in addition to the network model as the changing condition. This experiment uses Python3.7 in Ubuntu and is developed based on deep learning frameworks TensorFlow and Tensor2Tensor. The hardware is a single Nvidia TITAN GPU. Table 1 provides some parameter settings.

It was decided to use the statistical machine translation model as the baseline system and compare it to the neural network model, both of which were developed in the same environment and with the same parameter settings as the baseline system. The attentional mechanism algorithm and the transformer model were added for comparison, and the recurrent neural network model (RNN), long and short-term memory network model (LSTM), and bidirectional recurrent neural network model (BiRNN) were built for training using the above parameters. The translation quality measuring index BLEU-4 was chosen as the basis for this study. This experiment yielded the results given in Table 2.

The table provides the BLEU-4 score of this experiment. This experiment demonstrates the effectiveness of the machine translation model based on the deep learning neural network and obtains the performance of several networks in machine translation tasks according to the comparative experiment. The results obtained from the validation set and the two test sets show that the BLEU score of several neural network-based translation models exceeds the baseline system, and the transformer has the highest BLEU-4 score.

The LTNN translation model outperformed all other models without an attention mechanism, followed by bidirectional and unidirectional recurrent neural network models. From the control group of the four networks incorporating an attention mechanism, each translation model's performance improved significantly after incorporating an attention mechanism, and the addition of the attention mechanism significantly narrowed the gap with the baseline.

From the aforementioned results, we can conclude that the transformer model is the best. To provide a reference for the following experimental parameters according to the data scale in this study, a parameter comparison experiment is designed based on the transformer model. This study mainly explores the influence of the thesaurus size setting on model performance. The experimental design is as follows: under the condition that other parameters remain unchanged, the thesaurus size is modified to 16000 and 32000, respectively, and two models are obtained based on transformer framework retraining and tested on two test sets, respectively, for comparison. The experimental results are given in Table 3.

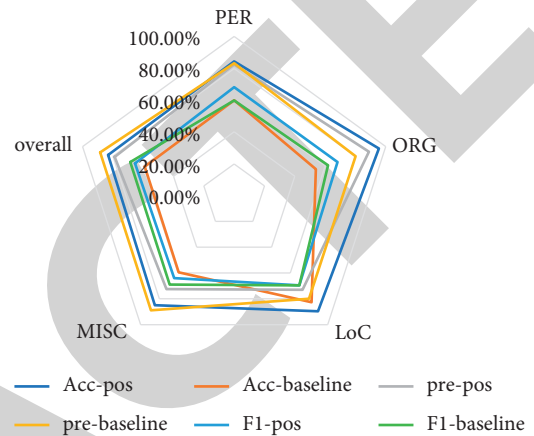


FIGURE 3: Experimental results of English NER.

It can be seen from table that the model with a vocabulary size of 16000 has an average improvement of 2.3 points in BLEU-4 compared with the model with a vocabulary size of 8000, while the model with a vocabulary size of 32000 has a decrease of 1.67 points compared with the model with a vocabulary size of 8000. Figure 2 shows a line chart, through which we can more intuitively see the influence of thesaurus size settings on model performance.

The following analysis can be obtained from the experimental results. The size of the word list has a certain effect on the performance of the neural machine translation model, but it does not increase linearly with the size of the word list. When the size of the word list exceeds a certain size, the model performance is reduced. The reason may be that the BPE algorithm for rare word processing has some limitations. For example, some illegal characters may be added to the word table when the set of the word table is too large, thus affecting the translation performance. Therefore, it can be concluded that adjusting the size of thesaurus appropriately in a certain range according to the size of the dataset is helpful to improve the performance of the model.

As can be seen from Table 4, after the loss weight of words at the beginning of a sequence is improved by using the loss function based on attenuation weight, the effect of translation can be improved to some extent by setting different attenuation factors' values of weight. Compared with the traditional model using unweighted loss function, the BLEU score of the translation improved up to 1.63% (factor 2 or 3). The reason for the improvement is that the new loss function can make the model more inclined to translate the earlier words so that the information used in the later words is more accurate in the process of translation.

TABLE 6: Experimental results of English named entity recognition.

	Acc		Pre		F1	
	POS (%)	Baseline (%)	POS (%)	Baseline (%)	POS (%)	Baseline (%)
PER	84.35	59.95	82.42	93.2	88.18	59.85
ORG	95.49	93.9	98.92	90.12	98.19	91.95
LoC	89.58	82.58	82.91	89.92	89.48	89.49
MISC	84.98	99.48	82.35	98.83	83.94	98.95
Overall	83.13	98.98	88.83	98.38	85.4	98.53

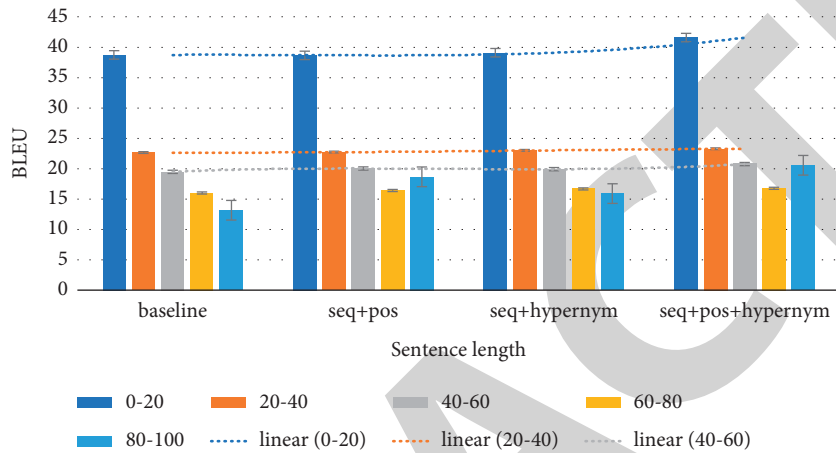


FIGURE 4: Sentence length sensitivity of multisequence encoding and baseline method.

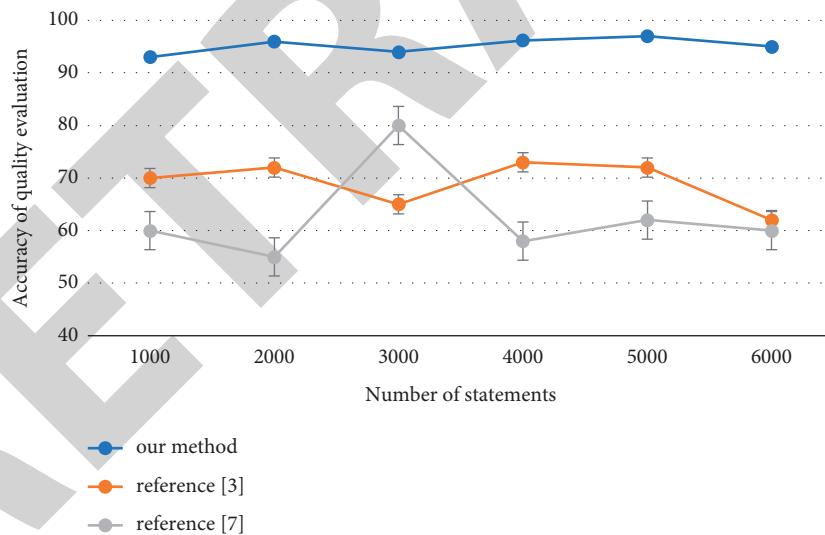


FIGURE 5: Quality evaluation efficiency of the three methods under different numbers of statements.

It can be seen from the results in Table 5 that in the training set scale of 1000, 500, and 300 samples, the model using group embedding has a faster convergence speed. At the same time, in these three training sets, the fewer the training samples, the higher the accuracy of the model using group embedding compared with the ordinary model. The experimental results of English named entity recognition are represented by the radar diagram in Figure 3. Through the radar chart, we can more intuitively see the influence of thesaurus size settings on model performance.

In addition, as given in Table 6, the effect of named entity recognition is significantly enhanced when parts of speech information is embedded in the recognition process. Convenient improvement is most noticeable in the recall rate, which is particularly high. Because of the addition of group embedding information, the model is able to correctly identify a large number of previously unrecognized entities. Due to the overlap between the information required by a named entity recognition task and the results of parts of speech tagging, this is the case in the majority of cases. As a

result, when used in conjunction with parts of speech information, group embedding can significantly improve the recognition effect in tasks such as named entity recognition.

It is investigated in this study whether the multisequence coding approach and the traditional method are sensitive to sentence length. This is done by testing the source language sentences in the test dataset according to the distribution of sentence length and evaluating the translation separately. Figure 4 shows the BLEU score of translation in the baseline and multisequence coding systems for different phrase length distributions of test data in the baseline and multisequence coding systems.

In Figure 4, it can be seen that when a sentence's length exceeds 20, the translation quality of both the baseline and multisequence coding methods shows a significant decline in quality, as can be seen in the graph. The SEQ + POS method, SEQ + HEAD method, and SEQ + POS + HEAD method all resulted in a reduction of 16.07, 15.91, 16.07, and 18.30 points in the BLEU score of translation when compared to the baseline technique. In general, as the length of a phrase increases, the multisequence coding method outperforms the baseline method in terms of translation performance.

This time, the quality evaluation efficiency of the three models was compared in order to better analyze their application value. The results of this comparison are shown in Figure 5, which shows the results of the comparison.

In Figure 5, it can be seen that no matter how many sentences are used, whether it is 1000 or 2000 or 3000 or 4000 or 6000, the quality assessment efficiency of the proposed model is significantly higher than that of the other two methods, while still maintaining the precision of the quality assessment. The highest efficiency in quality assessment is greater than 95% in some cases. It can, to a certain extent, demonstrate the viability of the model under consideration.

5. Conclusion

Artificial intelligence technology has exploded in popularity in recent years. It has made significant strides in the image, voice, and video fields. Although natural language processing is dubbed the crown jewel of artificial intelligence, it remains difficult to comprehend and process the natural language, and machine translation's representative application is still far from ideal. The advent of deep learning is unquestionably a watershed moment for machine translation. While traditional neural machine translation models treat each word in the target language sentence equally, the current word's decoding depends on the result of the previous word's decoding. As a result, this study proposes an attenuating weight loss function, which assigns different weights to each word in the training set. The earlier a word is translated, the greater is its weight, and training tends to allow for accurate translation of the first word in order to ensure the accuracy of previous information when translating the subsequent word. While the coding model based on pos embedding improves the efficiency of translation model convergence, the translation effect is not significantly improved. However, the proposed coding model based on parts of speech embedding can significantly improve the

performance of the two tasks of emotion analysis and named entity recognition. This result may be because translation requires a sizable corpus and linguistic knowledge of the parts of speech, so the information can be learned automatically during model training from a sizable corpus.

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 conflicts of interest.

References

- [1] M. A. Ranzato, C. Poultney, S. Chopra, and Y. Cun, "Efficient learning of sparse representations with an energy-based model," *Advances in Neural Information Processing Systems*, pp. 1137–1144, The MIT Press, Cambridge, MA, USA, 2006.
- [2] F. Seide, G. Li, and D. Yu, "Conversational speech transcription using context-dependent deep neural networks," in *Proceedings of the International Conference on Machine Learning*, pp. 1–2, Edinburgh, Scotland, July 2012.
- [3] N. Kalchbrenner and P. Blunsom, "Recurrent continuous translation models," in *Proceedings of the EMNLP*, pp. 1700–1709, Seattle, WA, USA, October 2013.
- [4] S. Hochreiter and J. Schmidhuber, *Supervised Sequence Labelling with Recurrent Neural Networks*, Springer, Berlin, Germany, 2012.
- [5] J. Chung, C. Gulcehre, K. H. Cho, and Y. Bengio, *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*, USA, 2014.
- [6] F. A. Gers and J. Schmidhuber, "Recurrent nets that time and count," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, p. 3189, Como, Italy, July 2000.
- [7] K. Cho, B. V. Merriënboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," pp. 1724–1734, 2014, <https://arxiv.org/abs/1406.1078>.
- [8] M. Wang, L. Gong, W. Zhu, J. Xie, and C. Bian, "Tencent neural machine translation systems for Wmt18," in *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*, pp. 522–527, Brussels, Belgium, October 2018.
- [9] Y. Xia, X. Tan, F. Tian et al., "Microsoft research Asia's systems for Wmt19," in *Proceedings of the Fourth Conference on Machine Translation*, vol. 2, pp. 424–433, Florence, Italy, August 2019.
- [10] F. Meng, J. Yan, Y. Liu et al., "Wechat neural machine translation systems for Wmt20," in *Proceedings of the Fifth Conference on Machine Translation*, pp. 239–247, USA, 2020.
- [11] J. Hao, X. Wang, S. Shi, J. Zhang, and Z. Tu, "Towards better modeling hierarchical structure for self-attention with ordered neurons," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 1336–1341, Hong Kong, China, November 2019.
- [12] T. Ge and F. W. M. Zhou, "Fluency boost learning and inference for neural grammatical error correction," in *Proceedings of the 56th Annual Meeting of the Association for*

- Computational Linguistics*, vol. 1, pp. 1055–1065, Melbourne, Australia, July 2018.
- [13] S. Jean, K. Cho, R. Memisevic, and Y. Bengio, “On using very large target vocabulary for neural machine translation,” pp. 1–10, 2015, <https://arxiv.org/abs/1412.2007>.
- [14] M. Nagao, *Artificial and Human Intelligence*, Elsevier, Amsterdam, Netherlands, 1984.
- [15] P. Brown, J. Cocke, S. D. Pietra et al., “A statistical approach to language translation,” in *Proceedings of the Conference on Computational Linguistics. Association for Computational Linguistics*, pp. 71–76, Budapest, Hungary, August 1988.
- [16] P. F. Brown, J. Cocke, S. A. D. Pietra et al., “A statistical approach to machine translation,” *Computational Linguistics*, vol. 16, no. 2, pp. 79–85, 1990.
- [17] P. F. Brown, V. J. D. Pietra, S. A. D. Pietra, and R. L. Mercer, “The mathematics of statistical machine translation: parameter estimation,” *Computational Linguistics*, vol. 19, no. 2, pp. 263–311, 1993.
- [18] F. J. Och and H. Ney, “A systematic comparison of various statistical alignment models,” *Computational Linguistics*, vol. 29, no. 1, pp. 19–51, 2003.
- [19] P. Koehn, F. J. Och, and D. Marcu, “Statistical phrase-based translation,” in *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pp. 48–54, Association for Computational Linguistics, Rochester, NY, USA, 2003.