Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume **7**, Number **1**, February **2022**

Chinese Text Implication Recognition Method based on ERNIE-Gram and CNN

Jingdong Wang

School of computer science
Northeast Electric Power University
No.169 Changchun Road, Jilin City, Jilin Province, 132012, China
wangjingdong@neepu.edu.cn

Huimin Li*

School of computer science
Northeast Electric Power University
No.169 Changchun Road, Jilin City, Jilin Province, 132012, China
*Corresponding Author: 1363022737@qq.com

Fanqi Meng*

School of computer science
Northeast Electric Power University
No.169 Changchun Road, Jilin City, Jilin Province, 132012, China
School of information engineering
Guangdong Atv Academy For Performing Arts
Huijing Road, Dongguan City, Guangdong Province, 523710, China
*Corresponding Author: mfq81@163.com

Peifang Wang

School of computer science
Northeast Electric Power University
No.169 Changchun Road, Jilin City, Jilin Province, 132012, China
w13578519072@163.com

Yujie Zheng

School of computer science
Northeast Electric Power University
No.169 Changchun Road, Jilin City, Jilin Province, 132012, China
278764886@qq.com

Xiaolong Yang

School of Economics and Management
Northeast Electric Power University
No.169 Changchun Road, Jilin City, Jilin Province, 132012, China
yangxiaolong@neepu.edu.cn

Jieping Han

School of Economics and Management
Northeast Electric Power University
No.169 Changchun Road, Jilin City, Jilin Province, 132012, China
hanjieping@126.com

Received September 2021; revised November 2021

Abstract. *Aiming at the problems of low accuracy in recognition of Chinese text implication and inability to better support machine reading comprehension, a Chinese text implication recognition method based on ERNIE-Gram and CNN is proposed. First, we use BERT word vector coding to encode the sentences to be input into the ERNIE-Gram model in three stages to improve the generalisation ability of the model. Then, the ERNIE-Gram model is used to obtain and integrate the semantic information of the text word level and sentence level, and different Warmup is introduced to optimize the learning rate in the ERNIE-Gram model training stage, and the ERNIE-Gram model parameters are continuously optimized. Finally, the fused semantic text is further extracted from the CNN for deeper semantic information, and the dimensionality reduction of the fully connected layer is used to identify the implied relationship of the text. The innovations are: (1) By introducing different Warmup in the ERNIE-Gram model to optimize the learning rate, it overcomes the shortcomings of insufficient semantic extraction due to model overfitting in traditional methods. (2) By adding a convolutional neural network to overcome the shortcomings of traditional methods of extracting semantic noise. Experimental results show that the accuracy of the method in identifying implication relations reaches 86 %, which lays a foundation for further efficient recognition of text implication.*
Keywords: Textual implication, ERNIE-Gram, Semantic information, CNN

1. **Introduction.** Textual implication is also called the reasoning relationship between texts, which is to judge whether there is an implication relationship between two texts [1]. At present, Chinese text entailment methods are mainly divided into two types: recognition methods based on component alignment and recognition methods based on machine learning. The recognition method based on component alignment is mainly the alignment of upper and lower words and some synonyms, this method is not suitable for the judgment of complex semantic relations. However, the method based on machine learning is difficult to completely extract the deep features of the sentence and the training time is long. This severely restricts the development of tasks such as information retrieval, machine reading comprehension, and machine translation [2].

Based on the above analysis, for Chinese text, in order to allow the computer to better understand its semantics, a Chinese text implication recognition method based on ERNIE-Gram and CNN is proposed. This method innovatively adds different Warmup in the model training stage to optimize the learning rate, and adds CNN to extract the semantic information twice, which effectively improves the accuracy of Chinese text implication recognition. It not only proposes to introduce different Warmup methods in the training stage of the ERNIE-Gram model to optimize the learning rate while optimizing the parameters of the model, thereby solving the problem that the model learning speed is too fast and the semantic information cannot be fully absorbed; but also through CNN, the fused text information is extracted in a deeper level, which solves the problem of insufficient semantic mining depth and difficulty in capturing deep semantic information. Then, after the probability is normalized by the fully connected layer and Softmax, the implication relationship between the texts is recognized.

In order to verify the effectiveness and stability of the proposed method, we conducted effectiveness experiments and stability experiments on Liu Huanyong Chinese textual implication data set *(https://github.com/liuhuanyong/ChineseTextualInference)*. Validity experiments show that the accuracy of our method reaches 86%, which is significantly higher than other methods. Stability experiments show that the accuracy of our method for recognizing text implication fluctuates within 2%.

The main contributions of this article:

(1) Migrate the ERNIE-Gram model to the field of Chinese text implication, extract and merge the semantic features of Chinese text at the word level and sentence level, and solve the problem of incomplete Chinese text semantic extraction.

(2) Innovatively add different warmups to the ERNIE-Gram model to optimize the learning rate, control the step length of model learning, and let the model use different learning rates at different stages to learn semantic features, so that the model can fully understand the semantics of Chinese text .

(3) Convolutional neural network is used to extract the semantics of the text twice, to capture the local semantic features of the sentence, and at the same time solve the difficulty of extracting deep-level semantic information from Chinese text, and provide a reference method for extracting deep-level semantic information from Chinese text.

The remaining parts are organized as follows: Section 2 reviews related work; Section 3 describes the Chinese textual implication recognition method based on ERNIE-Gram and CNN, including the three-stage text vector representation, the construction of a semantic recognition network fused with Warmup, and the deep-level semantic extraction and implication recognition based on CNN; Section 4 verifies the validity and stability of the model through experiments; Section 5 summarizes the work of this article, and tells the shortcomings and future research directions.

2. **Related Work.** The recognition of textual implication relations is a complex and important task, involving techniques such as extracting lexical features, semantic reasoning, and semantic mapping. At present, the recognition of textual implication relations mainly includes two methods: the recognition method based on component alignment and the recognition method based on machine learning.

Recognition methods based on component alignment can be subdivided into two categories, based on linguistic rules and adding manual annotation. Bentvogli et al. [3] analyzed the linguistic rules proposed in rte-5 data set, proposed the implication types such as vocabulary, vocabulary syntax, syntax, discourse and reasoning, incorporated the reasoning logic and common sense of quantity and space into the text implication recognition model, and summarized a set of model system suitable for text implication. Wang and Jiang [4] proposed the Mlstm model, which focused on the semantic matching of the aligned parts in the premise and hypothetical sentences. Chambers et al. [5] used the "alignment-filter" method to add parameter information to named entities, which were then merged into the text, and trained a maximum entropy classifier using manually annotated data to perform textual entailment recognition. Zhou et al. [6] marked some words such as "so" in the text to optimize the vector representation of premise sentences and hypothetical sentences, and then mapped the alignment components of premise sentences and hypothetical sentences to improve the accuracy of implication recognition. Ranjan et al. [7] used a variety of attention mechanisms to train the premise sentence and the hypothetical sentence separately, and recognize the relationship between the two sentences by the alignment components of the premise sentence and the hypothesis sentence. Tsuchida et al. [8] scored based on the degree of implication of vocabulary alignment and added a filtering mechanism of deep learning to obtain information from the different granularities of sentences to determine whether two sentences are implied. The above models are all based on keyword information for mapping and matching, which can only extract shallow semantic information, and cannot achieve the purpose of extracting deep semantic information of text.

The recognition method based on machine learning can be subdivided into the method based on the pre-training model fine-turn and the method based on migration. The method based on migration refers to improving the machine learning framework with

good results in other fields and migrating to the field of text implication recognition. The pre-trained network model shows good results in the recognition of Chinese textual implication relations, and has good reliability and portability. With the emergence of BERT [9], many scholars have migrated it to the field of textual implication recognition in recent years. ELMo [10] and BERT [11] were trained on the textual implication recognition data set when the model was proposed. He et al. [11] migrated the multi-task model to the field of textual implication recognition. The model recognizes the relationship between two sentences by training multiple tasks in parallel. It used to have the highest accuracy on the SNLI data set. Blunsom et al. [12] migrated the feedback adjustment mechanism to the field of textual implication, added artificial interpretations of implication relations to the SNLI data, and added these artificial interpretations to the model training process to form a feedback to continuously adjust the parameters of the model, is the first model to add interpretation to the recognition of implication. Migrating the interactive network framework to the related research on the task of processing textual implication, such as ESIM [13], BiMPM [14] and DIIN [15], etc, firstly, the premise sentence and hypothetical sentence are semantically represented by neural network coding, then the similarity between the word sequences of the two sentences is calculated through some complex attention mechanisms, the interaction matrix of the semantic information of the two sentences is constructed, and finally the interaction information is integrated. Nowadays, the most State-of-the-art method in the RTE field is the neural network method, which has a high recognition accuracy rate, but the neural network model runs slowly, and optimization requires certain skills. It is prone to problems such as overfitting or non-convergence due to improper parameter settings.

Different from the above methods, a Chinese text implication recognition method based on ERNIE-Gram [16] and CNN is proposed, which can extract semantics with multiple granularity, deep level and easy operation. Since the ERNIE-Gram model can extract and integrate the word-level semantic information and sentence-level semantic information of the text, and fully extract the text information at multiple granularities, the ERNIE-Gram model is selected as the basis, and the optimizer is added on this basis. Through Warmup, different optimization methods are used for different stages, and each group of parameters is saved with its own learning rate to update the parameters, which solves the problem of over-fitting or non-convergence of the traditional model due to improper parameter settings. Then the semantics are input to CNN for further extraction, which achieves the purpose of deep-level extraction of semantics. Convolutional neural network parallel processing solves the problem of slow running speed of traditional models. After the output of the fully connected layer and the probability of Softmax are normalized, a higher recognition accuracy of the relationship between sentences can be achieved.

3. **Chinese text implication recognition method based on ERNIE-Gram and CNN.**

The overall framework of Chinese text implication recognition method based on ERNIE-Gram and CNN is shown in Figure 1. It is generally divided into four steps: text vector representation based on three stages, semantic recognition network construction integrated with Warmup, deep semantic extraction based on CNN and implication relationship recognition.

Step 1: Based on three-stage text vector representation: in the encoding input stage, first encode the premise and hypothesis sentences and enter them into the ERNIE-Gram model. The input after the encoding of the two sentences is a 768-dimensional vector.

Step 2: Semantic recognition network construction based on Warmup: the innovative addition of ERNIE-Gram model is more suitable for Warmup optimization of learning rate contained in Chinese text, so as to prevent the fast learning speed in the beginning and

the model cannot absorb the semantic information well, and update the model parameters while optimizing the learning rate. The updated parameters of the last sentence pair need to be passed to the initial parameters of the next sentence pair (◯ in Figure 1 represents parameters), and then the extracted word-level semantics and sentence-level semantics are fused and sent to CNN for further semantic extraction and fusion.

Step 3: Deep-level semantic extraction based on CNN: the output of the ERNIE-Gram model is a 768-dimensional vector. The convolutional layer further extracts the deep-level semantics of the sentence, and then reduces the dimensionality into a 3-dimensional vector through pooling and full connection.

Step 4: Recognition of implication: by normalizing the probability of the 3-dimensional vector output by the fully connected layer, the relationship between sentences is recognized.

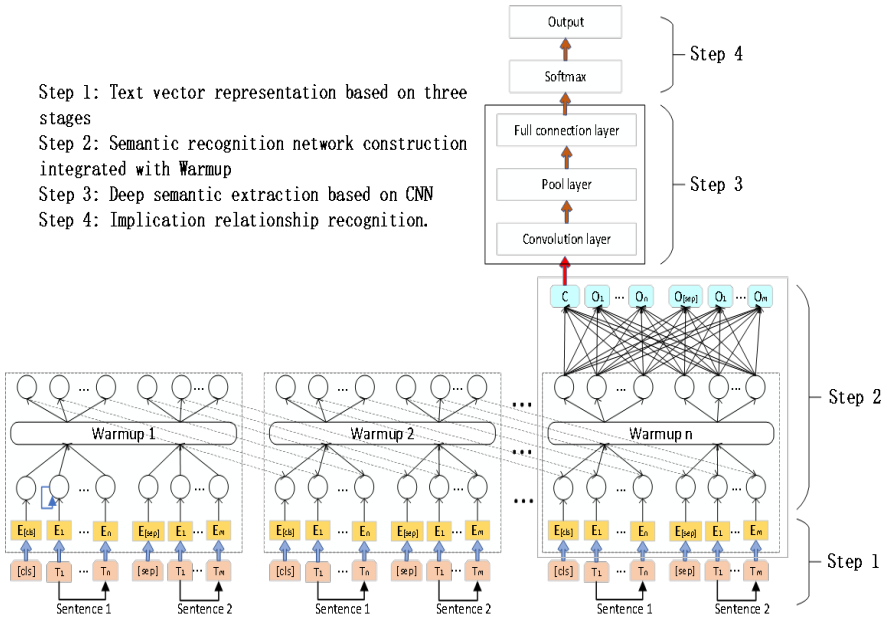


FIGURE 1. Overall framework of text implication recognition method

3.1. **Three-stage text vector representation.** In order to make full use of the characterization information of the sentence and improve the generalization ability, the BERT word vector coding is adopted. The coding input rules are shown in Figure 2, which are divided into three kinds of coding: Token Embeddings, Segment Embeddings, and Postition Embeddings.

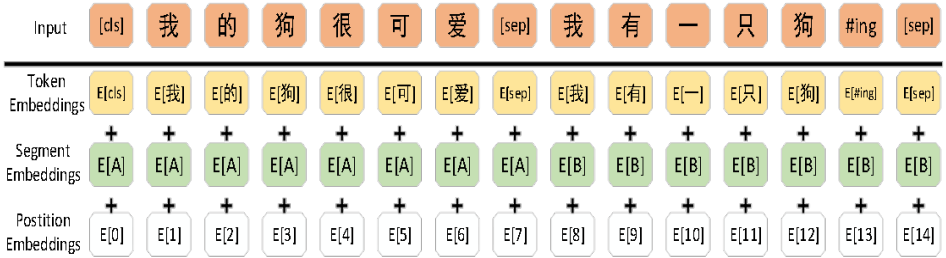


FIGURE 2. Text vector representation diagram based on three stages

(1) Token Embeddings is a 768-dimensional initialization vector, and the number of the corresponding word in the sentence is the corresponding Token Embeddings. For

example: the number of the word 'I' in the dictionary is 419, and the Token Embeddings of 'I' is a 768-dimensional vector on line 419.

(2) The Segment Embeddings is due to the Embeddings of two sentences in a model. To distinguish the two sentences, the symbols E[A] and E[B] are added to each sentence so that the model could distinguish the two sentences.

(3) The Postition Embeddings is set from 0. because the distance between the tokens could not be learned. The three codes simply add up to input Embeddings.

[*cls*] is a symbol indicating that it is the beginning of two sentences, [*sep*] represents the separator of two sentences, the sentence before [*sep*] is the predicate sentence, the sentence after [*sep*] is the hypothetical sentence, and finally the ending is indicated by *ing*.

The tokenizer is the same and the Postition Embeddings maximum value is 512. If you want to embed a partial function, you'll want to embed a set of parameters that are fixed when you're doing the plaintext translation, this will facilitate the processing of the data. For example, if sentence A and B are of length N and m respectively, then the input is a $768 * (m + n)$ matrix. After entering the ERNIE-Gram model, this matrix is multiplied by initializing different matrices K, Q and V respectively, and the two matrices are multiplied separately. Finally, the matrix is the relationship between each word of these two sentences, which can preserve the integrity of text information to the greatest extent.

3.2. **Construction of a semantic recognition network fused with Warmup.**

3.2.1. *ERNIE-Gram model construction based on relationship enhancement.* In order to better extract the semantic relationship of the text and make the model make better use of the text features, an enhanced n-gram relationship modeling mechanism is introduced. First, the original n-gram is masked with the likelihood n-gram identifier sampled from the generator model, and then the pairwise mapping relationship between the likelihood and the original n-gram is used to fuse them into a new n-gram. And added the target of the replacement tokens detection to distinguish between the original and the likelihood n-grams, which enhances the interaction between explicit n-grams and fine-grained context tokens.

At the same time, it predicts n-grams in a fine-grained and single-marker [M] coarse-grained manner, which helps to extract comprehensive n-gram semantics. Its loss function is as formula 1:

$$-\log p_\theta\left(y_M, z_M \mid \bar{z} \backslash M\right)=-\sum_{y \in y_M} \log p_\theta\left(y|_z^- \backslash M\right)-\sum_{z \in z_M} \sum_{x \in z} \log p_\theta\left(x|_z^- \backslash M\right) \quad (1)$$

$x = \{x_1, x_2, ..., x_{|x|}\}$ represents the input sequence;$y = \{y_1, y_2, ..., y_{|b|-1}\}$ representing the n-grams set; $b = \{b_1, b_2, ..., b_{|b|}\}$ representing the starting boundary set of n-grams; $z = \{z_1, z_2, ..., z_{|b-1|}\}$ representing the n-grams sequence set; M represents the randomly selected start boundary set, z_M represents the n-grams sequence set corresponding to M; $\bar{z} \backslash M$ Represents the sequence after z_M is masked;y_m represents the randomly selected n-grams set;

In order to predict all tokens contained in an n-gram from a single $[M]$, rather than a continuous $[M]$ sequence, a unique mask symbol $[M_i]$, $i = 1, 2, ...$ is used to de-aggregate the context representation to predict the tag in n-gram.

In order to clearly learn the semantic relationship between n-grams, a small generator model $\theta^{'}$ and an explicit n-gram MLM objective function are jointly trained to sample the n-gram identification. The model architecture diagram is shown in Figure 3:

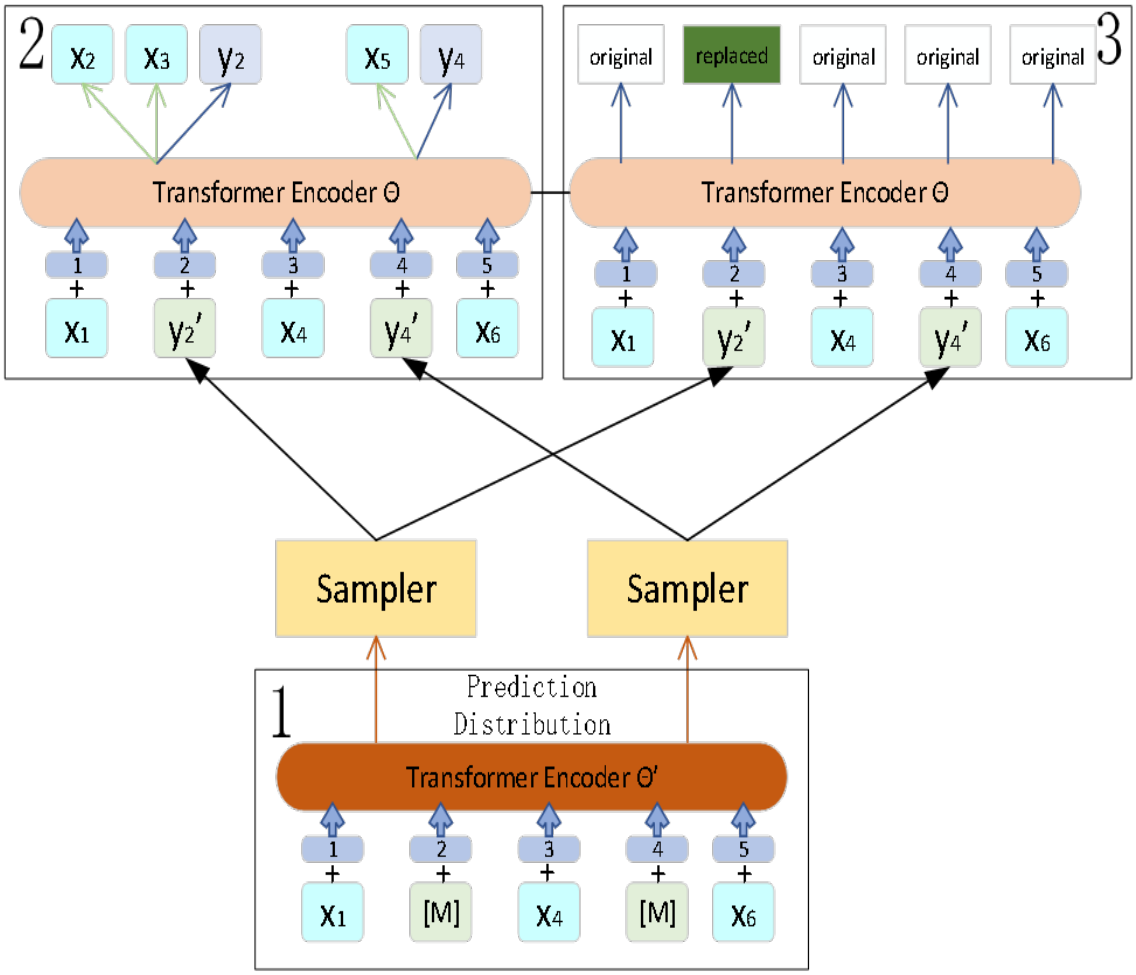


FIGURE 3. Detailed architecture diagram of relationship modeling

Use the generated identification to mask, train the standard model θ to predict the original n-gram in a coarse-grained and fine-grained manner, this model can effectively model the paired relationship between similar n-grams, and guarantee the accurate recognition of the implication relationship between similar sentence pairs.

In Figure 3, the Transformer Encoder $\theta^{'}$ in 1 is an explicit n-gram MLM. The loss function is shown in formula 2:

$$loss_1 = -\log_{p_{\theta'}} \left(y_M | \bar{z}_{\backslash M} \right) \tag{2}$$

In formula (2): $y_M = \{y_2, y_4\}$, $\bar{z} \backslash M = \{x_1, [M], x_4, [M], x_6\}$, moreover, $y'_M = \{y'_2, y'_4\}$ is the predicted result .

Transformer Encoder θ in 2, is a special comprehensive n-gram MLM. The loss function is shown in formula 3:

$$loss_2 = -\log_{p_\theta} \left(y_M, z_M \mid z' \backslash M \right) \tag{3}$$

In formula (3): $z' \backslash M$ is the original sequence with y'_M as the mask.

The Transformer Encoder θ in 3, completes the replaced token detection objective loss function as shown in formula 4:

$$-\log p_\theta \left(¶ \left(\bar{z}'_{\backslash M} = \hat{z} \backslash M \right) | \bar{z}_{\backslash M} \right) = -\sum_{t=1}^{|\bar{z} \backslash M|} \log p_\theta \left(¶ \left(\bar{z}' \backslash M, t = z \backslash M, t \right) | \bar{z}'_{\backslash M, t} \right) \tag{4}$$

In formula 4: $z \backslash M = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, if $¶ \left(\bar{z}'_{\backslash M} = \hat{z} \backslash M \right)$ is true, ¶ is 1, otherwise, ¶ is 0.

In general, in Figure 3, 1 uses contextual information to predict n-grams, and captures the connection between n-grams and context; 2 uses the sequence predicted in 1 to mask the sequence after the mask to predict the n-gram, and capture the connection between n-gram and n-gram. 3 has completed 2 more explicitly, it also uses n-gram information to predict the token of the context, which strengthens the connection.

3.2.2. *Warmup learning rate embedding based on dynamic adjustment.* Considering that the over-fitting phenomenon is prone to appear in the initial stage of the ERNIE-Gram model, the Warmup learning rate is introduced, and different learning rates are set for different stages. Optimizing parameters has always been a hot spot in deep learning [17, 18].The learning rate is an important parameter of the model. How to adjust the learning rate is one of the key elements of training a good model. When solving the minimum value of the problem through SGD, the gradient cannot be too large or too small. Too large is prone to overshoot, that is, continuous divergence or violent oscillation at both ends of the extreme point, and the loss does not decrease as the number of iterations increases; too small will result in the inability to quickly find a good drop direction, and the loss will remain basically unchanged as the number of iterations increases. The smaller the learning rate, the slower the loss gradient drops, and the longer the convergence time.

Since the depth and breadth of what the neural network learns at the beginning of training is very unstable, the initial learning rate should be set relatively low, which is to prevent the model from not converging. However, if the learning rate is too small, the training process will become very slow. Therefore, the initial stage of network training is realized by gradually increasing the learning rate from a lower learning rate to a higher learning rate. This process is called Warmup stage. But if the loss of network training is made small, then a higher learning rate cannot be used all the time, because it will make the gradient of the weights oscillate back and forth, the training results of the model will diverge, and it is difficult to make the training loss value reach the global minimum. Therefore, it is necessary to gradually reduce the learning rate after some steps.

It can be considered that the model's knowledge of the data is zero at the beginning, or that it evenly recognizes the data to be trained; In the first round of training, each data point is new to the model, and the model will quickly correct the data distribution. If the learning rate is very high at this time, it is likely to lead to "over fitting" of the data at the beginning, after many times of training, the model will have a new understanding and correct its original learning understanding, so as to pull the model back on the right track, training waste a lot of time and reduce the efficiency of the model. After training for a period of time, such as two or three rounds, the model is familiar with each data and model parameters, or has some correct priors for the current batch, a large learning rate is not so easy to bias the model. At this time, the university learning rate can be adjusted appropriately.

The principle of Consine decay is shown in formula 5:

Reduced learning rate:

$$\eta_t = \eta_{\min}^i + \frac{1}{2}\left(\eta_{\max}^i - \eta_{\min}^i\right)\left(1 + \cos\left(\frac{T_{cur}}{T_i}\pi\right)\right) \tag{5}$$

Meaning of characters in formula 5:

i is the index value; $\eta_{\max}^i$ and $\eta_{\min}^i$ represent the maximum and minimum of the learning rate, respectively, and specify the range of the learning rate. Keep $\eta_{\max}^i$ and $\eta_{\min}^i$ unchanged after each restart. T_{cur} indicates how many epochs are currently executed, but T_{cur} will be updated after each batch is run. At this time, an epoch has not been executed, so the value of T_{cur} can be a decimal. For example, if the total sample is 80

and the size of each batch is 16, the batch will be read five times in a cycle in an epoch. After the first batch is executed in the first epoch, the value of T_{cur} will be updated to $1/5 = 0.2$, and so on. T_i represents the total number of epochs in the $i-th$ run. T_i relatively small a will be initialized at the beginning, after each restart, T_i will increase by multiplying by a T_{mult}, that is, fix T_i as the number of epochs of the training model.

The control of the initial stage of Chinese textual entailment recognition based on ERNIE-Gram and CNN on the learning rate is shown in Figure 4.

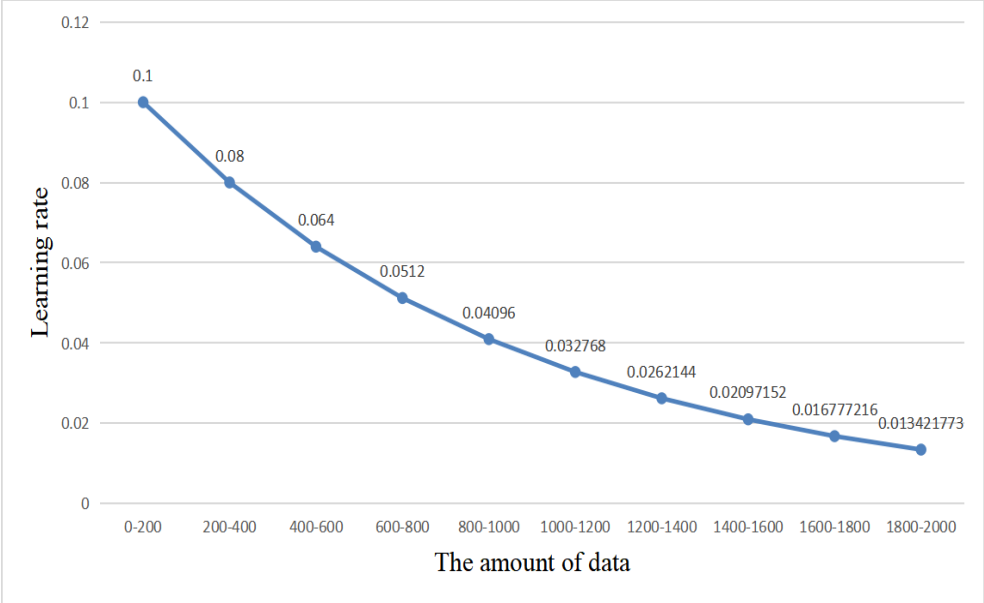


FIGURE 4. Control chart of the initial stage of learning rate

In Figure 4, the horizontal axis represents the training progress, and the vertical axis represents the learning rate. The initial learning rate is 0.1. Every 200 pieces of data are learned, the learning rate drops to 80% of the original. The first 2000 pieces of data are trained in this way, and the remaining learning rate is set to 0.1. This solves both the problem of over-fitting the model in the initial stages of training and the problem of slow training.

3.3. **Deep-level semantic extraction and implication recognition based on CNN.** In order to extract the deep semantic information of the text and identify the sentence relationship, firstly, the convolution layer of convolution neural network is used to convolute the text information and integrate different levels of semantic information, then the sentences are semantically matched through the pool layer and full connection layer, and finally the relationship between sentences is recognized through Softmax.

(1) CNN structure for extracting semantic information:

Combining previous work [19–22] and the characteristics of the model in this article, in order to further extract the text semantics, the CNN structural framework design is proposed as shown in Figure 5, combined with the model characteristics.

Input: In 'My dog is cute', each word is represented by a row vector with a shape of 1*5, and then these 7 words are stacked vertically into a two-dimensional matrix. The shape of the two-dimensional matrix is $Count(word) \times 5$.

Convolution kernel: after the input is determined, the following layer shows three convolution kernels of different sizes. It can be seen that one dimension of the convolution kernel is determined, which is equal to the dimension of the word vector. Then the convolution here is no longer a two-dimensional convolution in the image, but a one-dimensional convolution, and the convolution kernel only translates in the height dimension.

Convolution operation: after the convolution kernel is determined, each convolution kernel performs convolution operation with the input to obtain the output of a characteristic graph. This step is convolution.

Pooling operation: it can be seen from the figure that after convolution, the maximum pooling is carried out, and then the results of maximizing the pooling of each feature are stacked vertically.

Full connection layer: after the pooling operation, you can calculate the full connection in order to reduce it to a three-dimensional vector.

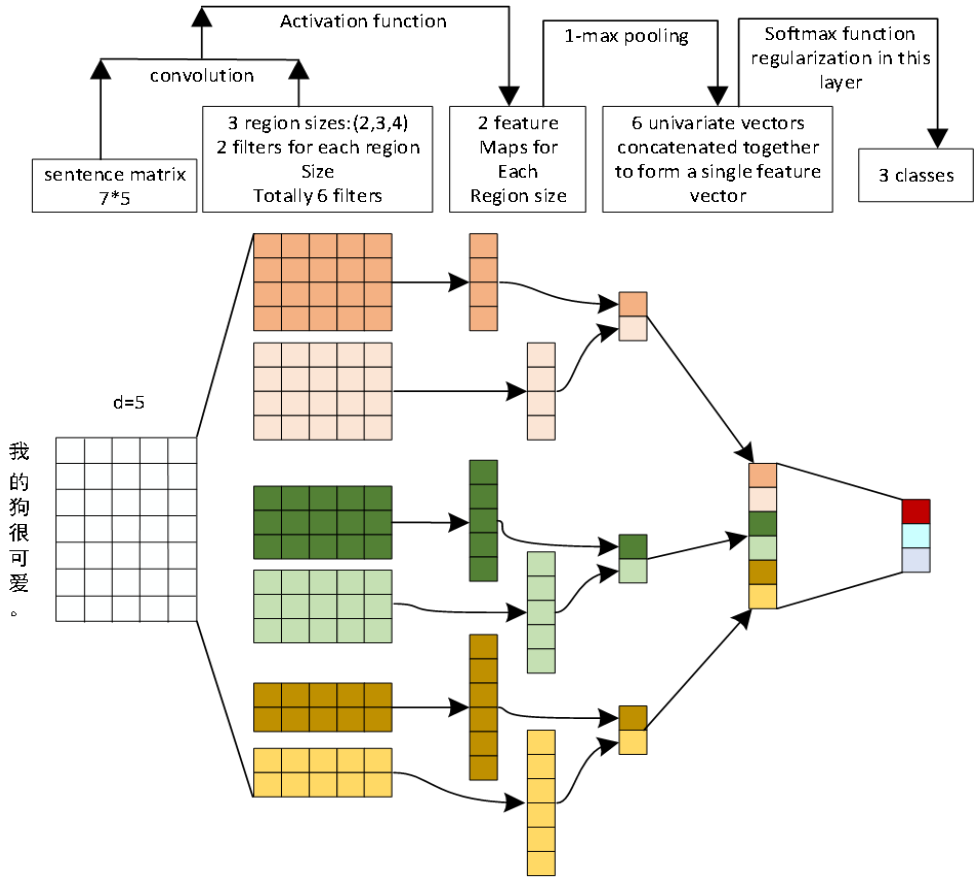


FIGURE 5. Control chart of the initial stage of learning rate

(2) Implication recognition

After the CNN full connection layer, a three-dimensional vector is obtained, and then normalized by Softmax to judge the relationship between the two sentences. Sigmoid function can map an input real number to the interval 0-1, then any X_1 can get a γ_1 on [0,1], that is, all values can be compressed to the interval 0-1. For an input score X of each category, the score can be mapped to the interval [0,1], that is, the score value can be converted into the corresponding probability value. Then the relationship between the two sentences can be obtained by normalizing the probability value.

4. **Experiment and result analysis.** In order to verify the effectiveness and stability of the proposed method, 12000 pieces of data from Liu Huanyong Chinese textual implication data set1, the most widely used data set at present, were randomly selected for two groups of experiments. The data set was comprehensively constructed by means of manual translation, machine translation and manual sorting based on the English text containing data sets SNLI and MultiNLI. Each piece of data contains a premise sentence, a hypothetical sentence and a label. The label is used to explain the relationship between

the two sentences. There are three categories of labels: Entailment, Neutral and Contradiction. In order to ensure the reliability, fairness and reproducibility of the experiment, the main parameters of the experiment are set as follows:

The initial learning rate of model training is 2e-5; Batch is 32; The maximum fixed length of a statement pair is 115; The maximum value of position mark is 512; The word vector dimension is 768; The label vector dimension of semantic role coding is 120. In order to control the complexity of the network and prevent over fitting, a dropout layer is added to the model. The initial values of dropout ratio between hidden layers and in attention layer are set to 0.1, CNN_Set size to 3, CNN_Num is set to 200.

The experiment takes accuracy as the evaluation index, and its specific definition is shown in formula 6:

$$Accuracy = \frac{N_C}{N_p} \times 100 \tag{6}$$

Where N_c is the number of sentence pairs with correct relationship prediction; N_p is the total number of predicted sentence pairs.

4.1. **Effectiveness experiment.** In order to evaluate and select the model, the model is tested by 50% discount. At the same time, in order to avoid the limitations and particularity of the data set, 12000 data are randomly selected twice on the data set, and the training set and test set are divided according to the ratio of 5:1.

Table 1. Statistics of two experiments

Relationship classification	First experiment data	Second experiment data
Entailment	4090	4103
Contradiction	3923	3978
Neutral	3987	3919
Total	12000	12000

It can be seen from table 1 that the three implication relationships of the data sets used in the experiment are evenly distributed, which avoids the deviation of the experiment and enhances the reliability of the experimental results.

In order to verify the reliability of ERNIE-Gram + CNN method in text implication recognition, the mainstream methods Bi-LSTM, Siamese + Bilstm, ESIM and ABCNN are used as baseline methods for comparison, Table 4.1 shows the implication recognition accuracy of different methods on the selected data set:

Table 2. Implication recognition accuracy of different methods

Groups	Data	Bi-LSTM	Siamese+Bilstm	ESIM	ABCNN	ERNIE-Gram+CNN
	Training 1	56.72%	33.92%	37.66%	34.93%	89.89%
	Test 1	58.68%	35.89%	32.33%	36.88%	86.22%
	Training 2	56.82%	34.87%	37.58%	36.12%	88.22%
first	Test 2	58.99%	36.34%	32.53%	36.97%	86.92%
experiment	Training 3	57.23%	34.95%	37.96%	35.73%	89.39%
	Test 3	56.44%	35.72%	34.82%	36.95%	86.43%
	Training 4	59.82%	34.05%	37.69%	35.79%	89.32%
	Test 4	58.66%	36.41%	32.41%	36.88%	89.03%
	Training 5	56.89%	34.12%	34.51 %	35.85%	89.41%
	Test 5	58.68%	35.77%	39.32%	37.32%	87.20%

Groups	Data	Bi-LSTM	Siamese+Bilstm	ESIM	ABCNN	ERNIE-Gram+CNN
	Training 6	56.92%	33.72%	37.66%	35.17%	89.17%
	Test 6	57.62%	35.69%	35.33%	37.38%	88.22%
	Training 7	57.43%	33.98%	37.96%	34.93%	89.01%
second	Test 7	58.48%	36.23%	39.43%	36.98%	86.32%
experiment	Training 8	56.32%	34.66%	38.55%	35.93%	89.91%
	Test 8	58.88%	35.89%	37.31%	38.28%	88.23%
	Training 9	56.97%	34.14%	38.66%	36.93%	89.39%
	Test 9	58.89%	35.99%	35.33%	37.48%	86.02%
	Training 10	56.22%	33.79%	39.32 %	35.53%	89.37%
	Test 10	58.17%	35.93%	35.21%	37.88%	87.49%

The experimental results in Table 2 show that compared with other methods, ERNIE-Gram + CNN method has the best performance. Finally, the accuracy on the training set is more than 89%, and the accuracy on the test set is more than 86%, which is significantly higher than other mainstream models.

Add the accuracy of the training set and the test set to calculate the average accuracy of the two sets of data. The results are shown in Figure 6 and Figure 7.

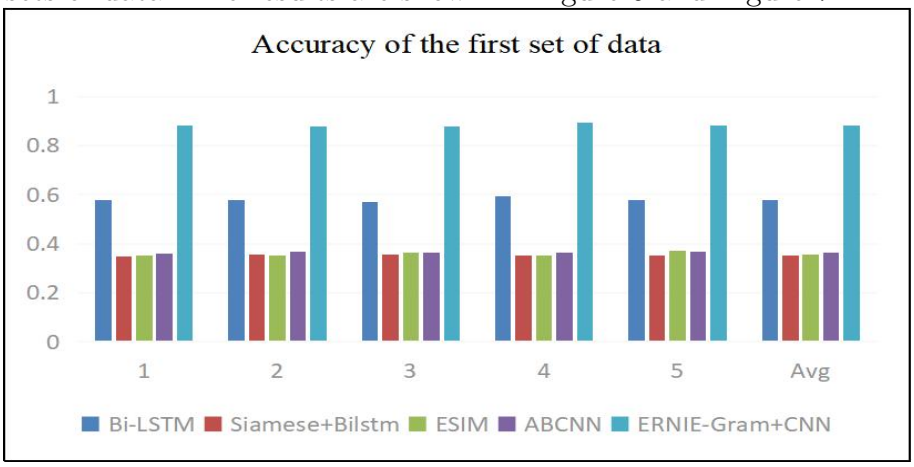


FIGURE 6. The average accuracy of the first experiment

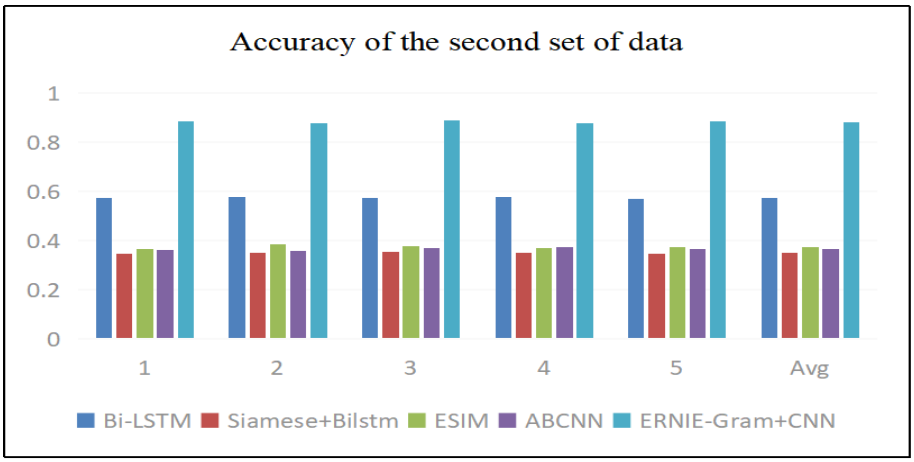


FIGURE 7. The average accuracy of the second experiment

Compared with traditional methods, the Chinese text meaning recognition method based on ERNIE-Gram and CNN effectively improves the accuracy of Chinese text meaning recognition, and the results of many experiments are relatively stable.

The average confusion matrices of the ERNIE-Gram+CNN model obtained from the two experiments are shown in Figure 8 and Figure 9:

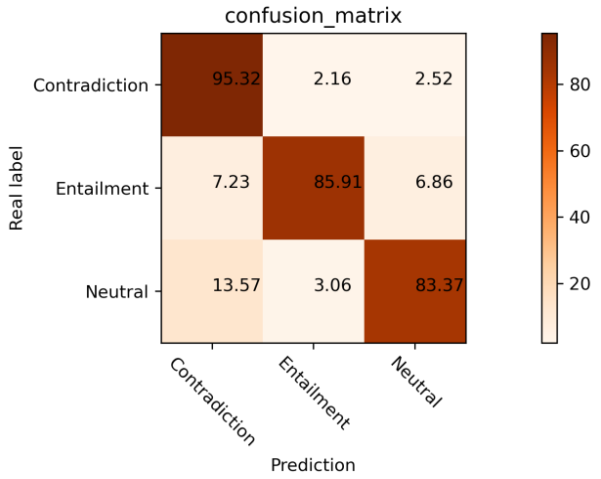


FIGURE 8. Confusion matrix of the first experiment

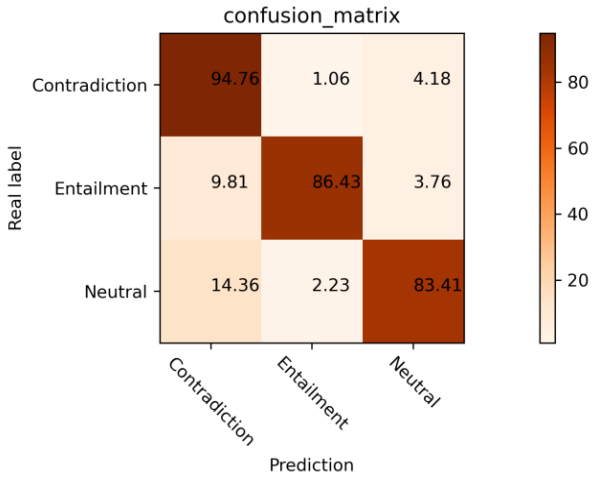


FIGURE 9. Confusion matrix of the second experiment

It can be seen from Figure 8 and Figure 9 that there are more predictions of implication relations as contradictions, and the correct rate of contradictions is the highest, followed by implication relations, and finally neutral relations.

4.2. **Stability experiment.** In order to verify the stability of the method, the two experimental data of the first group of experiments were divided into training set and test set again by 8:2 and 7:3, and then four experiments were done. The recognition accuracy rate is shown in Table 3:

TABLE 3. Implication recognition accuracy of different methods

Groups	Proportion	Bi-LSTM	Siamese+Bilstm	ESIM	ABCNN	ERNIE-Gram+CNN
1	8:2	56.72%	35.92%	40.66%	39.06%	90.81%
	7:3	58.68%	37.89%	35.33%	40.55%	89.82%
2	8:2	56.92%	36.72%	39.60%	37.97%	88.47%
	7:3	57.62%	39.99%	38.83%	41.18%	88.22%

It can be seen from Table 3 that the five models are relatively stable. ERNIE-Gram+CNN has the highest accuracy, followed by the Bi-LSTM model. The recognition accuracy of Siamese+Bilstm, ESIM and ABCNN is not much different.

The confusion matrices of the four experiments ERNIE-Gram+CNN are shown in Figure 10, Figure 11, Figure 12, and Figure 13.

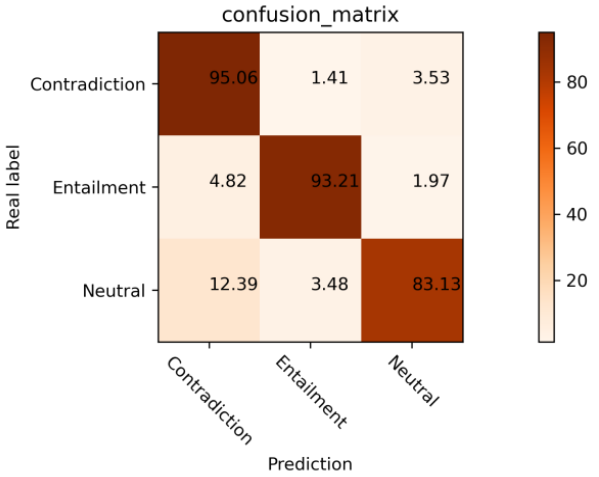


FIGURE 10. Confusion matrix of the first experiment

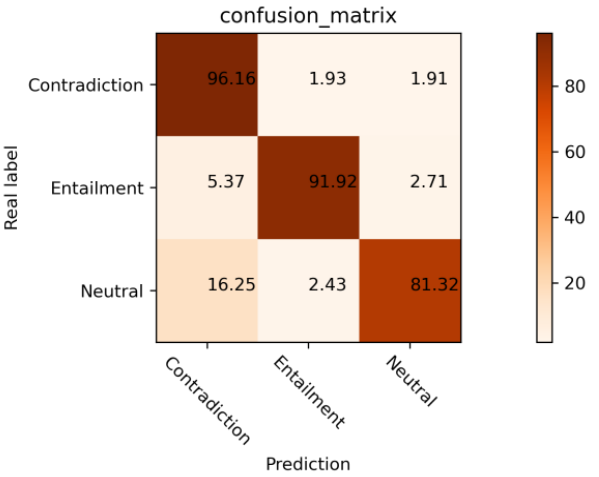


FIGURE 11. Confusion matrix of the second experiment

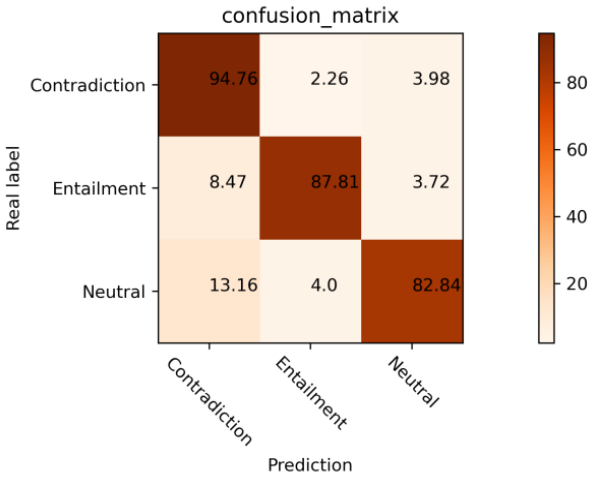


FIGURE 12. Confusion matrix of the thrid experiment

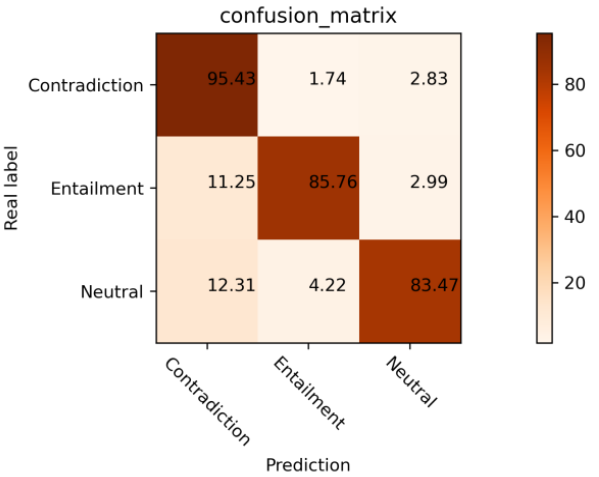
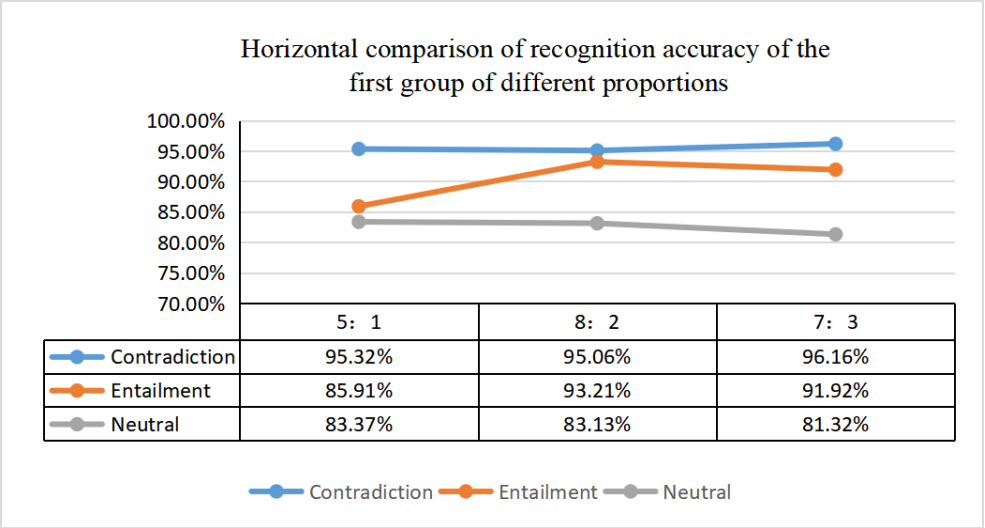


FIGURE 13. Confusion matrix of the fourth experiment

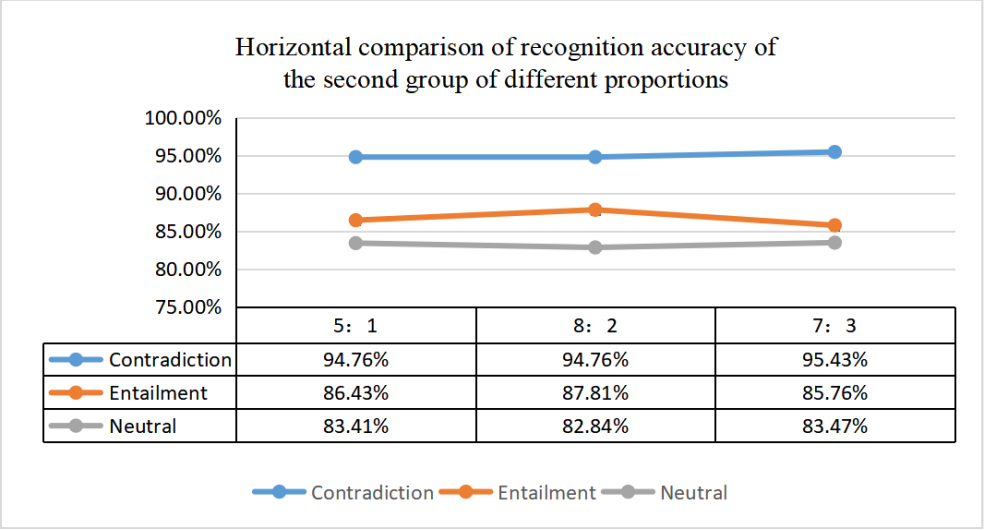
It can be seen from Figure 10, Figure 11, Figure 12, and Figure 13 that the Chinese text implication recognition method based on ERNIE-Gram and CNN has better stability. However, this method is easy to tilt the recognition result to the contradiction. The prediction accuracy of the neutral relationship is low, and the highest recognition error rate is to recognize the neutral relationship of the sentence pair as a contradiction relationship. In order to compare the stability of the proposed method more clearly, the two sets of data divided into different proportions are horizontally compared with the recognition results. As shown in Figure 14, Figure 15:

4.3. **Clustering methods.** A distinctive feature of LSGDM different from traditional GDM is that there are a large number of DMs, ranging from tens to hundreds or thousands. Therefore, it is very important to reduce the dimension of DMs for LSGDM problems. Clustering can divide DMs into different subgroups, which is an effective means to reduce the scale. In this paper, the following two categories are classified according to whether the clustering method needs to specify the number of clusters in advance:



	5: 1	8: 2	7: 3
Contradiction	95.32%	95.06%	96.16%
Entailment	85.91%	93.21%	91.92%
Neutral	83.37%	83.13%	81.32%

FIGURE 14. Horizontal comparison of recognition accuracy of the first group of different proportions



	5: 1	8: 2	7: 3
Contradiction	94.76%	94.76%	95.43%
Entailment	86.43%	87.81%	85.76%
Neutral	83.41%	82.84%	83.47%

FIGURE 15. Horizontal comparison of recognition accuracy of the second group of different proportions

It can be seen from 14 and Figure 15 that the ERNIE-Gram+CNN method has the highest and most stable recognition accuracy of contradictory relations, with an accuracy rate of about 95% and a stability of ±1%. The recognition accuracy of implication relations is the most unstable, and the recognition accuracy of implication relations is generally lower than that of contradictory relations, but generally higher than that of neutral relations. The accuracy rate of the neutral relationship recognition of this method is also relatively stable, but the accuracy rate is relatively low.

4.4. **Discussion.** 200 examples of prediction errors were randomly sampled from the data set for analysis. The reasons for the prediction errors are roughly divided into 7 categories, and the reasons for the errors are analyzed below:

(1) Ambiguity: due to the complexity and flexibility of Chinese language, segmentation of words in different positions of sentences will cause changes in sentence semantics. Even if the sentence segmentation is correct, different people's understanding will be different and there will be some ambiguity. For example, the premise sentence is: the girl is a middle school teacher. The hypothetical sentence is: The girl is teaching people to draw. It is unclear whether the girl who is teaching people to draw is a middle school teacher. The model identifies neutrality as an implicit relationship.

(2) Negative words: if there are negative words such as "no" and "no" in the sentence, the model is more inclined to judge it as a contradictory relationship, especially when there is only one negative word in the premise sentence and hypothetical sentence. However, double negation means that advanced grammatical models such as affirmation cannot be distinguished yet. For example, the premise sentence is: We do not consider whether this matter is right or wrong. The hypothetical sentence is: We don't care if this thing is done right or not. The model recognizes implication as a contradictory relationship.

(3) Calculation: involving the semantics of mathematical calculations, the model is difficult to identify. For example, the premise sentence is: There are two apples on the table. The hypothetical sentence is: The boy takes an apple from the table, and there is an apple left on the table. The model recognizes implication as a contradiction relationship.

(4) Reasoning: the judgment of text implication relationship requires certain knowledge reasoning, otherwise it is difficult to make a correct judgment on the relationship between two sentences. For example, the premise sentence is: the roller is going uphill. The hypothetical sentence is: the road roller is struggling. The model recognizes implication as a neutral relationship.

(5) Corresponding words: if the premise sentence and hypothetical sentence contain corresponding words, the model prefers to judge them as implication relations, which is the local alignment semantics of the sentence, but not the semantic expression of the whole sentence. For example, the premise sentence is: a little boy stands on the road and looks into the distance. The hypothetical sentence is: A little boy looks at the road in the distance. The model will identify the contradictory relationship as an implicit relationship.

(6) Synonymous heteromorphism: for the expression of synonymous heteromorphism, the model is difficult to identify. For example, the premise sentence is: about 1/3 of the students in the class failed the exam. The hypothetical sentence is: the passing rate of the exam is about 72%. The model will identify the implication relationship as a neutral relationship.

(7) Other: Some instances have no obvious source of error. For example, the premise sentence is: a woman hiding from a dog at the beach. The hypothetical sentence is: the woman is afraid of the dog. The model identifies contradictions as implicit relations.

From the above error analysis, the current Chinese text contains many areas that need to be improved.

5. **Conclusions.** Recognizing textual implication is a very important task. At present, the main difficulty in recognizing Chinese textual implication lies in considering the incomplete semantics, and it is difficult to make full use of the semantic information of the text. In order to improve the accuracy of Chinese text implication recognition, a Chinese text implication recognition method based on ERNIE-Gram and CNN is proposed, this method recognizes the relationship between sentences by embedding Warmup's ERNIE-Gram model and convolutional neural network to extract the semantics of sentences in two layers. We select 10,000 pieces of data as training data and 2,000 pieces of data as test data on the public data set, after many experiments, the results show that the accuracy

of this method in identifying Chinese texts is higher than that of the current mainstream methods. In order to test the stability of the method, the data set was divided into 7:3 and 8:2 to re-experiment, and the recognition accuracy rates are 90%, 89%, 88%, and 88%, respectively. Based on these results, the effectiveness of the method in the recognition of Chinese text implication is proved. The significance of this method is that the Warmup optimized learning rate that conforms to the Chinese text is added to the pre-training model, so that the pre-training model can extract semantics more fully. At the same time, after the pre-training model extracts the semantics once, the convolutional neural network is added to extract the semantics in a deeper level, which solves the problem that the traditional methods are difficult to fully capture the semantic information due to the large and insufficient semantic extraction noise. However, this method has room for improvement. In future research, the double tower model can be used to extract semantics, combined with Chinese grammatical characteristics to refine the semantic fusion method, and further strengthen the model's ability to extract semantic features. In addition, there are currently few Chinese text data sets, and there are almost no Chinese text data sets specific to a certain field, making the current research not comprehensive enough, so the establishment of a Chinese text data set in a specific field is also a meaningful work.

6. **Acknowledgment.** This work is supported by the National Key&D Program of China, Ministry of Science and Technology of the People's Republic of China (No 2020YFB1707804), and the 2020 Jilin City Science and Technology Development Plan Project "Jilin City Tourism Online Comment Text Emotion Classification Research" (No.20200104108).

REFERENCES.

[1] L. Zhang, D. Moldovan, Multi-Task Learning for Semantic Relatedness and Textual Entailment, *Journal of Software Engineering and Applications*, vol. 12, no. 6, pp. 199-214, 2019.

[2] M.Q. Pham,M.L. Nguyen,A. Shimazu, Learning to Recognize Textual Entailment in Japanese Texts with the Utilization of Machine Translation, *ACM Transactions on Asian Language Information Processing*,vol. 11, no. 4, pp. 1-23, 2012.

[3] L. Bentivogli, E. Cabrio, I. Dagan, D. Giampiccolo, B. Magnini, Building Textual Entailment Specialized Data Sets: a Methodology for Isolating Linguistic Phenomena Relevant to Inference, *Proceedings of the International Conference on Language Resources and Evaluation,Valletta, Malta*, pp. 3542-3549, 2010.

[4] S. Wang, J. Jiang, Learning Natural Language Inference with LSTM, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 2015*, 2015, https://arxiv.org/pdf/1512.08849.pdf

[5] N. Chambers, D. Cer, T. Grenager, Learning alignments and leveraging natural logic, *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pp. 165-170, 2007.

[6] X. Zhou, C. Yao, H. Wen, EAST: An Efficient and Accurate Scene Text Detector, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, https://arxiv.org/pdf/1704.03155v2.pdf

[7] V. Ranjan, H. Le, M. Hoai, Iterative crowd counting, *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 270-285, 2018.

[8] Tsuchida, Masaaki, K. Ishikawa, IKOMA at TAC2011: A Method for Recognizing Textual Entailment using Lexical-level and Sentence Structure-level features, *TAC*, 2011, https://tac.nist.gov/publications/2011/participant.papers/IKOMA.proceedings.pdf

[9] J. Devlin, M.W. Chang , K. Lee, Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint*, 2018, https://arxiv.org/pdf/1810.04805.pdf

[10] B. Yang, X.Y. Du, Natural scene text location algorithm based on improved East, *Computer engineering and application*, vol. 55, no. 18, pp. 161-165, 2019.

[11] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.

[12] P. Blunson, OM. Camburu, T. Lukasiewicz, e-SNLI: Natural language inference with natural language explanations, *Proceddings of the 32nd Annual Conference on Neural Information Processing Systems*, pp.9560-9572, 2018.

[13] Q. Chen, X. Zhu, Z. Ling, Enhanced LSTM for Natural Language Inference, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 1657-1668, 2016.

[14] Z. Wang, W. Hamza, R. Florian, Bilateral multi-perspective matching for natural language sentences, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 4144-4150, 2017.

[15] Y. Gong, H. Luo, J. Zhang, Natural language inference over interaction space, *International Conference on Learning Representations*, 2018, https://arxiv.org/pdf/1709.04348.pdf

[16] D. Xiao, Y.K. Li, H. Zhang, ERNIE-Gram: Pre-Training with Explicitly N-Gram Masked Language Modeling for Natural Language Understanding, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2021, https://arxiv.org/pdf/2010.12148.pdf

[17] Q. Yang, S.C. Chu, J.S. Pan, C.M. Chen, Sine Cosine Algorithm with Multigroup and Multistrategy for Solving CVRP, *Mathematical Problems in Engineering 2020*, pp. 1-10, 2020.

[18] L.L. Kang, R.S. Chen, N.X. Xiong, Y.C. Chen, Y.X. Hu, C.M. Chen, Selecting Hyper-Parameters of Gaussian Process Regression Based on Non- Inertial Particle Swarm Optimization in Internet of Things, *IEEE Access*, vol. 7, pp. 59504-59513, 2019.

[19] E.K. Wang, X. Zhang, F. Wang, T.Y. Wu, C.M. Chen, Multilayer Dense Attention Model for Image Caption, *IEEE Access*, vol. 7, pp. 66358-66368, 2019.

[20] J. M.-T. Wu, M.-H. Tsai, Y.-Z. Huang, SK H. Islam, M. M. Hassan, A. Alelaiwie, G. Fortino, Applying an ensemble convolutional neural network with Savitzky-olay filter to construct a phonocardiogram prediction model, *Applied Soft Computing*, vol. 78, pp. 29-40, 2019.

[21] J. M.-T. Wu, M.-H. Tsai, S.-H. Xiao, Y.-P. Liaw, A deep neural network electrocardiogram analysis framework for left ventricular hypertrophy prediction, *Journal of Ambient Intelligence and Humanized Computing*, 2020, https://doi.org/10.1007/s12652-020-01826-1

[22] J. M.-T. Wu, Z. Li, G. Srivastava, M.-H. Tasi, J. C.-W. Lin, A graph based convolutional neural network stock price prediction with leading indicators, *Software: Practice and Experience*, vol. 51, no. 3, pp. 628-644, 2021.