Hindawi
Advances in Multimedia
Volume 2023, Article ID 9836986, 1 page
https://doi.org/10.1155/2023/9836986

Retraction

Retracted: Evidence Prediction Method Based on Sentence Selection for Legal Documents

Advances in Multimedia

Received 15 August 2023; Accepted 15 August 2023; Published 16 August 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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] H. Bai, "Evidence Prediction Method Based on Sentence Selection for Legal Documents," *Advances in Multimedia*, vol. 2022, Article ID 1926347, 9 pages, 2022.

Hindawi
Advances in Multimedia
Volume 2022, Article ID 1926347, 9 pages
https://doi.org/10.1155/2022/1926347

Research Article

Evidence Prediction Method Based on Sentence Selection for Legal Documents

Huijie Bai

School of Animation, Shijiazhuang University of Applied Technology, Shijiazhuang, Heibei 050081, China

Correspondence should be addressed to Huijie Bai; baihuijie@sjzpt.edu.cn

Received 27 May 2022; Accepted 4 July 2022; Published 19 July 2022

Academic Editor: Qiangyi Li

In order to solve the problem that it is difficult to find evidence from a large number of legal document statements and the irrelevant statements in a large number of document sample data will cause a great interference to the prediction results and further improve the accuracy of evidence prediction, this paper puts forward an intelligent evidence criterion prediction method for legal documents based on the comprehensive consideration of legal problems, the nature of statements, and the characteristics of answers. The binary cross-entropy of different statements is used to obtain the interaction information between different statements. Through experiments, it is found that the score of Joint F1 proposed in this paper is 70.07%, which is more accurate than the mainstream model and also verifies the effectiveness of the scheme.

1. Introduction

As a special document type, legal documents are very strict in structure, which not only requires strict logic but also requires complete statements. In recent years, with the rapid growth of the number of legal documents and the rapid development of artificial intelligence technology, machine reading comprehension in the legal field has developed rapidly. With the help of machine learning and legal document reading, we can more clearly express the well-structured legal documents and further improve the efficiency of traditional manual work. However, in terms of practical application, the prediction of evidence needs to find the corresponding answers and relevant evidence from a large number of legal documents, so it is very difficult to achieve. Moreover, a large number of sample data of statements and documents will cause a lot of interference to the final prediction results of evidence. In order to further improve the accuracy of evidence prediction, this paper proposes an intelligent evidence criterion prediction method for legal documents and verifies the effectiveness and feasibility of the method through a series of experiments, as shown in Figure 1.

2. Literature Review

Sentence prediction task plays a great role in automatic sentencing in intelligent justice. At present, there are some researches at home and abroad. Among them, some scholars use their own defined tags as features to assist in sentencing prediction. It is found that the certainty of sentencing can be improved by reducing the ruling range of sentencing circumstances, and additional features play a great role in these studies. In machine learning, in most application scenarios, whether text, image, audio, or their corresponding machine learning methods, the types of data are diverse [1]. For text, features can be divided into multiple levels, such as sentence level features, word level features, and letter level features, and even structural data can be extracted from the problem. With the deepening of research, the scope of application field has become more extensive. It is difficult for a single feature or single model to perfectly complete various complex tasks and achieve ideal results [2]. Therefore, some scholars are labeling some words with vague meanings, accurately defining them, and combining a variety of information at the same time to improve the prediction accuracy. Not only text data but also other structured data are used for fusion, and

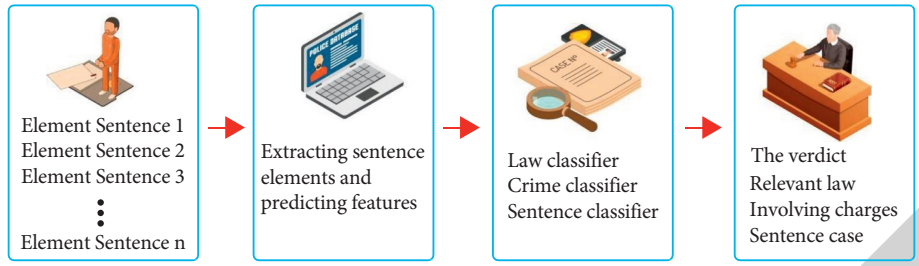

Figure 1: Legal instruments oriented.

then the effect of the model is improved through the attention mechanism. From the text, image, video, and other pieces of information to build a network for multimodal fusion to complete the task, multilevel feature fusion can better reidentify tasks. New progress has been made in unsupervised context discovery by trying heterogeneous feature fusion. It is necessary to balance the weight of features during feature fusion. In the case of multitask feature fusion, the feature balance is better [3].

Evidence prediction is to extract the sentences supporting the answers from the text. The HotpotQA data set was released in 2018, which provides evidence to support the answer. The difficulty of evidence prediction lies in that the problem of reading comprehension itself may not effectively provide clues to find evidence sentences. Some scholars regard the evidence prediction of interpretable multihop QA (question and answer) as a query-centered summary task and use the attention mechanism of RNN to the problem to predict the evidence. Imperfect tags are generated through remote monitoring, and they are used to train and predict evidence. Burris et al. designed a self-training method (STM), which generates evidence tags to supervise the evidence extractor during the iteration process to assist in answer prediction [4].

Many classical models of reading comprehension can be used for evidence prediction, such as BiDAF proposed by foreign scholars and R-Net proposed by Microsoft. These are language models based on learning word embedding, and there are many similar models. Since the BERT model was proposed, the best results have been achieved in tasks in multiple NLP fields, including machine reading comprehension [5].

3. Intelligent Decision Prediction and Recommendation Model for Semantic Matching of Legal Documents

3.1. Rule Recommendation Model of Semantic Matching Tandem Reselection Mechanism

3.1.1. General Framework. In this section, the word vector of the case description and the word vector of the i legal provision are, respectively, defined as formulas (1) and (2):

$$E = [E_1, E_2, E_3, \ldots, E_n, \ldots, E_N], \tag{1}$$

$$\hat{E}_l = [\hat{E}_{i1}, \hat{E}_{i2}, \hat{E}_{i3}, \ldots, \hat{E}_{in}, \ldots, \hat{E}_{iM}], \tag{2}$$

where N represents the number of words after the case description is cut and M represents the number of words after the i-th legal provision is cut. Here, the output of the semantic matching model is defined as R_i, and each R_l is the recommendation index described by the i-th relevant law for the case. At the same time, the input of the reselection mechanism is also defined on this basis; the sentence vector and the probability distribution of the recommendation index are, respectively, S as shown in formula (3):

$$R = [R_1, R_2, R_3, \ldots, R_I]. \tag{3}$$

The output is index P. Based on this definition, a rule recommendation model of semantic matching tandem reselection mechanism is proposed. The model includes a bidirectional transformer convolution network model and reselection mechanism.

The structure of the rule recommendation model of the semantic matching tandem reselection mechanism is the bidirectional transformer convolution network model and the reselection mechanism, which are connected in series. The bidirectional transformer convolution network model is composed of six layers: input layer, BERT layer, convolution layer, pooled activation layer, full connection layer, and output layer [6].

3.1.2. Bidirectional Transformer Convolution Network Model. The bidirectional transformer convolution network model (BCNN) is divided into the following parts.

Input layer: after a series of text preprocessing on the data, the corresponding word vector is obtained, and then according to the fixed format required by BERT, the word vector of the case description and the word vector of the i-th legal provision are spelled into a sentence pair vector matrix, as shown in the following formula:

$$[E; \hat{E}_l]. \tag{4}$$

As an input vector, it is input into the model through the interface of the BERT model.

BERT layer: the main function of the BERT layer is to extract the correlation between case description and answer and give greater weight to more relevant words. At the same time, the corresponding text semantic information can be obtained from the sparse long text vector [7], as shown in Figure 2.

BERT's word embedding method is different from other general word embedding methods. As shown in formula (5),

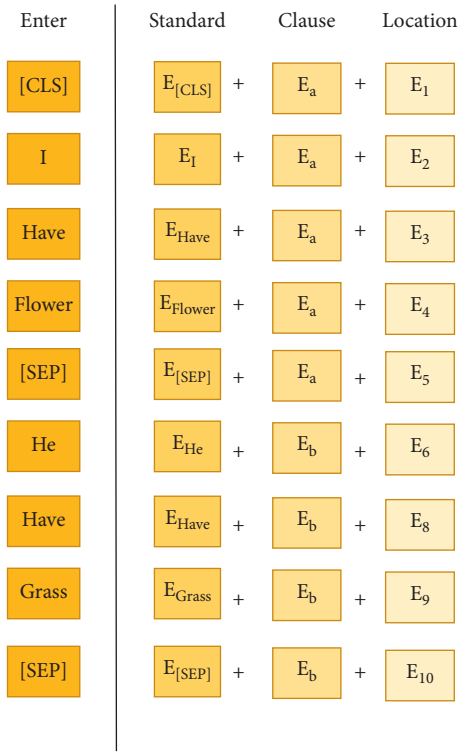

Figure 2: BERT word embedding method.

it is obtained by summing three types of word embedding representations.

$$E = E_{\text{word}} + E_{\text{seg}} + E_{\text{pos}}. \tag{5}$$

Convolution layer: the main function of the convolution layer is to focus on extracting local features in semantic representation. Since the BERT layer compresses the semantic relationship, word correlation, and other pieces of information in the long text sequence into the vector matrix W and sentence vector S, the convolution layer mainly extracts the most important semantic logic relationship from the semantic information contained in these high-dimensional vector representations (vector matrix W) as the extracted features [8].

In this paper, the convolution layer is used to receive the sequence vector matrix extracted by the BERT layer, which is a two-dimensional tensor. This convolution layer uses a user-defined convolution kernel to convolute the input tensor. In particular, it is generally a convolution kernel whose width is consistent with the length of the word embedding vector. The input tensor of the convolution check moves in parallel from top to bottom. After each translation, each parameter in the convolution kernel will be multiplied by the input of the corresponding position and added as the output. The specific process of using a convolution kernel is shown in Figure 3 [9]. For content a, b, c, d in the convolution window, the convolution kernel is w, x, y, z, so the primary convolution calculation is shown in the following formula:

$$[a, b, c, d] \cdot [w, x, y, z] = aw + bx + ey + fz. \tag{6}$$

Full connection layer: after various important representations with different granularity are extracted through the above method, these features need to be integrated. Because the full connection layer can provide richer nonlinear expression, it will not cause some unnecessary data loss when compressing data, so the full connection layer is used as a bridge between the activation layer and the output layer to provide the output layer with the representation after feature integration [10].

The activation layer generally appears at the same time as the pooling layer and receives the data output from the pooling layer. Because the neurons in the neural network are linear combinations of inputs, in order to make the neural network approach any function, it is necessary to introduce a nonlinear function as the excitation function to enrich the expression of the network. In this paper, the nonlinear function (ReLU) function is introduced into the active layer as the excitation function of the active layer, which is shown in the following formula:

$$\text{Re}lu(x) = \max(0, x). \tag{7}$$

3.2. Reselection Mechanism. XGBoost is an improved algorithm for the traditional GBDT. Its main improvement is that the complexity of the tree is also taken into account in the objective function, and the Taylor expansion of the objective function is used to solve the second-order approximate solution in the iterative optimization process, which can speed up the iterative process. The definition of the XGBoost objective function is shown in the following formula [11]:

$$L = \sum_{i=1}^{n} loss_{XG}(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k). \tag{8}$$

The first part of the above formula is used to measure the difference between the predicted score and the real score, and the second part is the regularization term of the tree complexity. Softmax is selected as the loss function in this paper. Further, equation (8) may be rewritten as follows:

$$\begin{aligned} L^{(t)} &= \sum_{i=1}^{n} loss_{XG}\left(y_i, \hat{y}_i^{t-1} + f_t(x_i)\right) + \Omega(f_t) \\ &\cong \sum_{i}^{n}\left[loss_{XG}\left(y_i, \hat{y}_i^{t-1}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right] + \Omega(f_t), \end{aligned} \tag{9}$$

where g is the first derivative and h is the second derivative, as shown in the following formulas:

$$g_i = \partial_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}_i^{t-1}\right), \tag{10}$$

$$h_i = \partial^2_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}_i^{t-1}\right). \tag{11}$$

3.3. Experimental Results and Analysis

3.3.1. Data and Preprocessing. In order to objectively describe the effectiveness of the article recommendation model of semantic matching tandem reselection mechanism designed in this section, this section will conduct

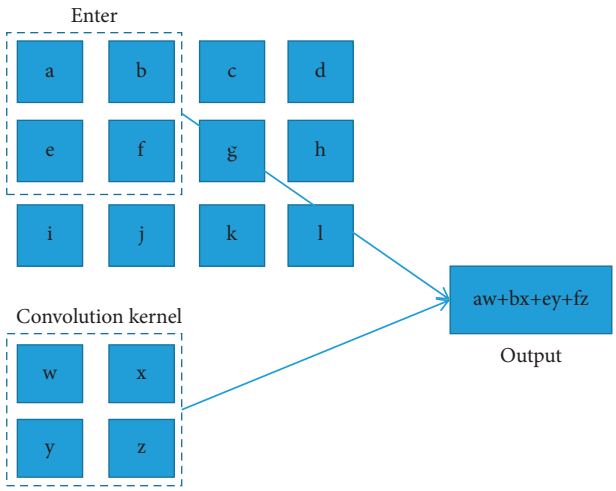

Figure 3: Convolution operation diagram.

experiments on CAIL2018, the largest open legal data set in China. The experimental data comes from the China FA Yan cup competition data set, and the cases in the judgment document network of the Supreme People's Court of China are used as data. There are 183 different relevant laws and regulations in the data set. This section extracts all single label samples from them, so there are only 163 relevant laws and regulations, and supplements the corresponding legal provisions for these relevant laws and regulations. In the experiment, the number of training sets is 114824, the number of verification sets is 14293, and the number of test sets is 23593 [12].

First, clean up the text and delete abnormal data, meaningless pause words, specific time, and other pieces of unimportant information [13]. Then use jieba word segmentation to divide a whole case description into many small segments into word units. When the recommendation index of a case description and each relevant law article is obtained, the probability distribution of the recommendation index can be obtained by combining them in order, and, at the same time, the ranking of the top five relevant articles with the largest recommendation index is constructed in order. Among them, P is the recommendation article of this case, and its index is the corresponding output of the reselection mechanism [14].

3.3.2. Experimental Setup and Evaluation Index. First, set the number of words in the case description and relevant provisions to 270 and 30, respectively, and the total number of words in the two text splicing is 300. In the experiment, the Word2Vec word vector used is a 300-dimensional word vector trained by the corpus provided by Baidu Encyclopedia, Chinese Wikipedia, people's daily, and so on. [15]. For all the experiments, this section uses jieba word segmentation tool to preprocess the text, such as stopping this filtering and the corresponding word segmentation.

For the semantic matching algorithm, the convolution kernel widths of the QACNN model are 2, 3, 4, 5, 7, and 9, respectively, the node dimensions of the first layer of the full connection layer are set to 1024, the adaptive learning rate adjustment algorithm (AdaDelta) is used to update the model training parameters, the learning rate is set to $1e-5$, the decay coefficient of the learning rate is set to 0.95, the constant ε is set to 610, and the sigmoid classifier is used to calculate the recommendation score. In XGBoost, the parameter gamma to control whether to prune is set to 0.1, the max_depth to control the depth of the tree is set to 8, the L2 regularization coefficient is set to 10, the minimum leaf node sample weight and min_child_weight are set to 1, and multi_softmax is used as the loss function.

The experimental environment of this paper is configured as follows: Intel (R) Xeon (R) CPU e5-2650 V4 @ 2.20 GHz; 128 G DDR4 memory; Titan XP model GPU; CUDA version 10.1. The experimental code is implemented by Python of version 3.6, Keras framework, and multiple third-party machine learning libraries and tested and run in Anaconda3 environment [16].

3.3.3. Result Analysis. In order to demonstrate the help of adding legal provisions to the model, this paper compares the case description without legal provisions with that with legal provisions and makes a visual attention test. This shows the feasibility and effectiveness of the problem transformation in this section. In the BERT model, all the contents in the case description are highly dependent on the word "human property," but this word is obviously not very helpful for the semantic matching task and the theft of the legal provisions corresponding to the case description [17]. The content of the case description is also highly dependent on the words "illegal occupation" and "pickpocketing," which are very helpful for the semantic matching task and the theft of the legal provisions corresponding to the case description. This can verify the feasibility of problem transformation in this task and the effectiveness of adding legal provisions [18].

When comparing the following traditional semantic matching algorithms QACNN, Seq2Seq, and BERT models with the semantic matching model proposed in this section, this paper uses the accuracy rate as the evaluation index and tests with Top1, Top5, and Top10. This shows that various semantic matching models have achieved good results for this task, but the lack of a reasonable selection mechanism within a certain range has led to a decline in accuracy. The experimental results are shown in Table 1. At the same time, the reselection mechanism proposed in this section is concatenated after each semantic matching model to demonstrate whether the reselection mechanism is effective. The experimental results are shown in Table 2 [19].

As can be seen from Table 2, the reselection mechanism proposed in this paper has significantly improved the algorithm of the semantic matching system. It is 0.267 higher than QACNN on the data set used in this section. For Seq2Seq, it is 0.298 higher. For the BERT model, it is 0.301 higher. For our semantic matching model, it is 0.303 higher. This is because the reselection mechanism implemented by XGBoost in this paper can reselect the recommendation index. After reselection of the original inaccurate prediction,

Table 1: Experimental results of semantic matching model effect comparison (accuracy).

Model	QACNN	Seq2Seq	BERT	BCNN (ours)
Top1	0.314	0.522	0.522	0.628
Top5	0.633	0.948	0.826	0.956
Top10	0.768	0.975	0.929	0.978

Table 2: Experimental results (accuracy) of effect comparison of whether there is a reselection mechanism.

Type	QACNN	Seq2Seq	BERT	BCNN (ours)
No reselection mechanism	0.315	0.511	0.552	0.628
With reselection mechanism	0.582	0.809	0.853	0.915

correct relevant legal provisions are recommended for each case description, which significantly improves the prediction accuracy [20].

In order to demonstrate the effectiveness of feature fusion, this paper mainly compares the traditional text classification algorithms CNN, TextCNN, LSTM, and GRU with the causal TextCNN proposed in this section after feature fusion and uses score and RMSE as evaluation indicators. The experimental results are shown in Tables 3 and 4 [21].

In order to know which feature is more effective in improving the sentence prediction model based on causality in this section, this paper mainly compares the traditional text classification algorithms CNN, TextCNN, LSTM, and GRU sentence model based on causality with the probability distribution of charges and the recommendation index distribution of legal provisions as features and uses score and RMSE as evaluation indicators. The experimental results are shown in Tables 5 and 6.

It can be seen from Table 6 that, in the case of using feature fusion, using the probability distribution in Section 2 as the feature alone can better improve the sentence prediction model based on causality than using the recommended index distribution in Section 3 as the feature alone [22]. Take score as the evaluation index, using probability distribution as the feature is 0.047, 0.027, 0.019, 0.017, and 0.047 higher than CNN, LSTM, GRU, TextCNN, and the sentence prediction model based on causality on the data set we tested. Take RMSE as the evaluation index, using probability distribution as the feature is 1.12, 2.03, 1.89, and 1.05 higher than CNN, LSTM, GRU, TextCNN, and the sentence prediction model based on causality on the data set we tested 210 [23].

In order to verify the rule recommendation model of semantic matching tandem reselection mechanism proposed in this paper, this section compares the following traditional semantic matching algorithms QACNN, Seq2Seq, and BERT models and classification algorithms CNN, TextCNN, LSTM, and GRU with the rule recommendation model of semantic matching tandem reselection mechanism. In this section, the accuracy is used as the evaluation index, and the experimental results are shown in Figure 4.

Table 3: Score of model before and after feature fusion.

	CNN	LSTM	GRU	TextCNN	Causal textCNN
Before fusion	0.382	0.371	0.238	0.397	0.339
After fusion	0.325	0.315	0.311	0.327	0.424

Table 4: RMSE scores of models before and after feature fusion.

	CNN	LSTM	GRU	TextCNN	Causal textCNN
Before fusion	30.23	26.03	40.81	23.87	28.90
After fusion	25.23	33.03	32.75	29.87	18.72

As can be seen from Figure 4, the accuracy of the method proposed in this paper is much higher than that of the traditional classification algorithm and the reordered semantic matching models CNN, TextCNN, GRU, and LSTM by 0.073, 0.064, 0.060, and 0.090, respectively. The rule recommendation model of semantic matching tandem reselection mechanism proposed in this paper is to perform fine-tune on the data set by using BERT. First, the ability of the model itself ignores the distance between words and is good at understanding long text sequences. In addition, because the BERT model itself has been trained through a large number of corpora and can be better used with this data set through the role of fine-tune, its model is more robust [24].

4. Evidence Prediction Method Based on Sentence Selection for Legal Documents

4.1. Method Introduction. This paper uses the encoder stack based on BERT as the base model, as shown in Figure 5. The basic model is used in three modules: sentence selection, answer prediction, and evidence prediction [25].

4.2. Tightly Connected Encoder Stack. This paper uses the closely connected encoder stack based on BERT as the basic model, which learns the deep semantic information and surface semantic information of the model, greatly reducing the loss of features learned by the model at the beginning. As shown in the DencseEncoder Block in the lower part of Figure 5, different coding layers of BERT have learned different representations of the language. Legal documents are composed of the detailed contents of the case. The rigorous structure shows that the information characteristics of each layer of the model may be useful. Therefore, in the sentence selection module, answer prediction module, and evidence prediction module, this paper uses this basic model to improve the accuracy of evidence prediction.

4.3. Multihead Self-Attention Layer. In fact, there is a certain relevance between the evidence and the questions and answers, including in the legal documents, and there is also a certain relevance between different sentences. Exploring the relevance between different sentences can promote the downstream prediction evidence. In order to consider these correlations more comprehensively, a multihead self-

Table 5: Score of model before and after feature fusion.

	CNN	LSTM	GRU	TextCNN	Causal textCNN
Probability distribution of accusation	0.389	0.357	0.293	0.372	0.405
Recommended index distribution of relevant articles	0.342	0.330	0.274	0.355	0.358

Table 6: RMSE scores of model before and after feature fusion.

	CNN	LSTM	GRU	TextCNN	Causal textCNN
Probability distribution of accusation	27.56	27.31	33.98	26.91	25.22
Recommended index distribution of relevant articles	28.68	29.34	35.87	27.96	23.12

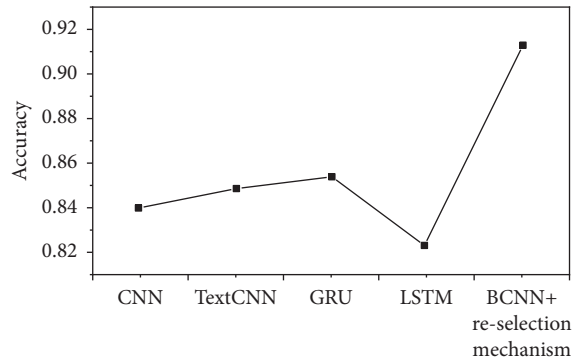

Figure 4: Experimental results (accuracy) of comparison between methods and classification algorithms in this section.

attention layer is added to the interaction between attention statements. The formula is as follows:

$$\text{Attention}(Q,K,V) = \text{softmac}\left(\frac{QK^T}{\sqrt{d_k}V}\right), \quad (12)$$

$$\text{head}_i = \text{Attention}\left(QW_I^Q, KW_i^K, VW_i^V\right), \quad (13)$$

$$\text{MultiHead} = \text{Concat}\left(\text{head}_i, \ldots, \text{head}_n\right)W^0, \quad (14)$$

where Q, K, V is the linear projection from the labels of different statements [CLS], representing the attention query, key, and value, respectively. The multihead self-attention layer pays attention to the [CLS] tags of different sentences in order to pay attention to the interaction between sentences, let the model learn the relevance between them, and then promote the work of evidence prediction.

4.4. Binary Cross-Entropy Loss Function. In the statement selection module, this paper uses the idea of similar threshold to rank different statements C in the data set and sets the score S for each statement. Set the statement i score $S(C_i)$ according to the ranking. The higher the ranking, the higher the score. Set the statement score containing the answer to positive infinity and the lowest score to 0. In order to reduce the amount of calculation, this paper adopts a method similar to calculating the binary cross-entropy loss. First, define the labels of each pair of statements i and j as shown in the following formula:

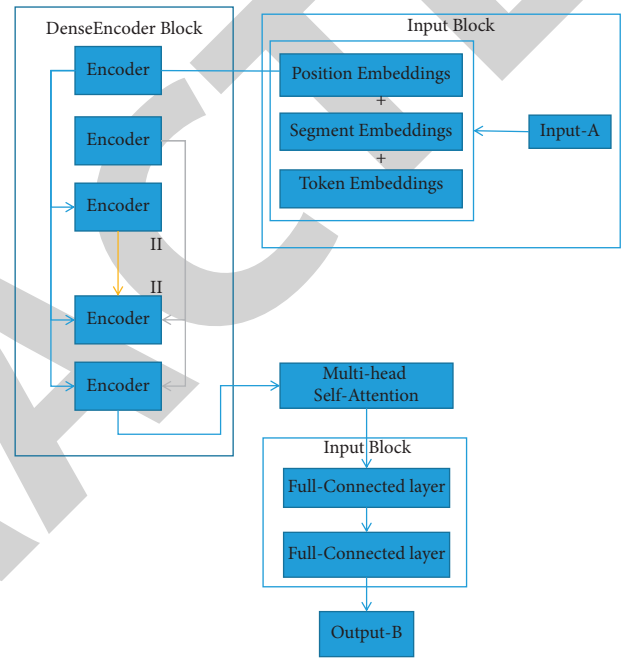

Figure 5: Basic model, used for sentence selection, evidence prediction, and answer prediction.

$$l_{i,j} = \begin{cases} 0, & if S(C_i) > S(C_j), \\ 1, & \text{others}. \end{cases} \quad (15)$$

In this way, it can ensure that the statements with higher relevance to the questions and answers get higher scores, the statements containing answers get higher scores than other statements, and the control score is between 0 and 1. The binary cross-entropy is calculated as follows:

$$L = -\sum_{i=0}^{i}\sum_{(j=0,j^1i)}^{i} l_{i,j}\log P\left(C_i, C_j\right) + \left(1 - l_{i,j}\right)\log\left(1 - P\left(C_i, C_j\right)\right). \quad (16)$$

Among them, $LP(C_i, C_j)$ is the probability that the model predicts that the statement C_i is more relevant than the statement C_j. In this paper, the first 10 statements are selected as documents filtered by the statement selection module, which can be better used for evidence speculation.

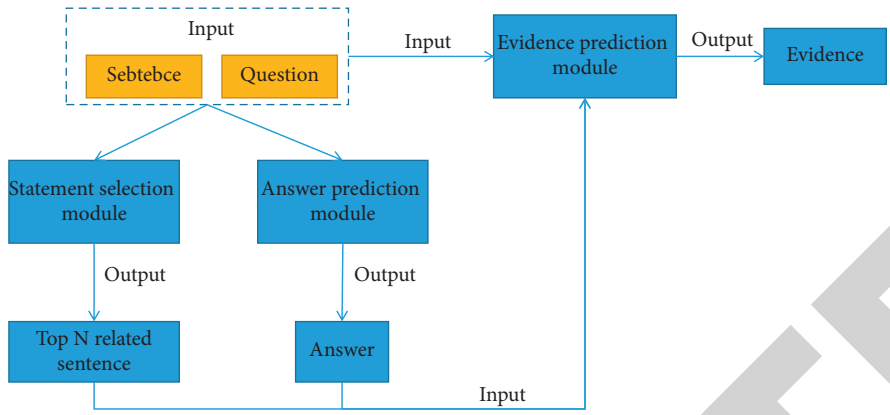

Figure 6: Model flowchart during test.

4.5. Model Training and Testing

4.5.1. Model Training. During training, the sentence selection module, answer prediction module, and evidence prediction module share a basic model. See 4.1–4.4 for details of the basic model. The three modules are trained separately. Next, we will focus on the training of the statement selection module. Here, we will introduce the input and output of the module in detail. The training of answer prediction module and evidence prediction module is similar. Statement selection module: the function of statement selection module is to filter statements, prevent irrelevant statements from distracting attention, reduce training time, increase performance, and minimize irrelevant information transmitted to subsequent tasks. This module is very important for the later prediction support statements.

Answer prediction module: similar to the statement selection module, input In-put-A becomes [CLS] + question + [SEP] + document statement + [SEP], and output-b becomes the predicted answer. Although the use of the sentence selection module for evidence prediction has achieved good results, there is still room for improvement. This paper considers adding another factor, that is, the predicted answer, to assist in evidence prediction.

Evidence prediction module: the evidence prediction module is also similar to the statement selection module. The input-a becomes [CLS] + question + [SEP] + document statement + [SEP] + answer + [SEP], which is used for input. The question directly comes from the data set, the document comes from the statement selection module, and the answer is the answer predicted by the answer prediction module. Output-b becomes the predicted evidence.

After using the statement selection module, a large number of invalid statements are eliminated. This paper believes that we can not only deduce the evidence from the question like CogQA but also add new factors to deduce through the answer. Different from the joint training of answer prediction and evidence prediction, the evidence prediction module does not help answer prediction but uses answer prediction to assist in deriving evidence. This is because the accuracy of answer prediction is much higher than that of evidence prediction, and joint training will have a negative impact on the answer prediction task.

4.5.2. Model Test. As shown in Figure 6, the model testing process can be seen as a combination of the above three modules. After the test data passes through the statement selection module and the answer prediction module, the filtered statements and answers are obtained. They are tested together with the questions as the input of the evidence prediction module and predict the evidence.

The experiment is carried out on a Linux server, which is composed of four E5 processors and four TITANX GPUs. Due to the change in the official baseline, the prediction model of this study is RoBERTa-wwm-ext, a Chinese pretraining model based on Whole WordMasking published by PyTorch. The overall structure of the model is exactly the same as the RoBERTa base. Due to the limitation of conditions, this paper sets the batch size to 2, the maximum SEQ length to 512, the step length of the sliding window of the channel to 128, the maximum question length to 64, and the maximum answer length to 55. It trains for 8 hours on the four TITANX GPUs with an initial learning rate of $1e-6$.

In order to accurately evaluate the effect of the model, F1 and EM and Joint F1 and Joint EM are used for the answer prediction and evidence prediction used in the evidence prediction in this paper. It should be noted that the official baseline model is compiled based on Jinshan Spider Net.

In this paper, experiments are conducted in cjrc 2020 data set. The experimental results are shown in Table 7. The results in the table are from the competition list and the experiments conducted in this study, both of which adopt the results of nonintegrated single model. The model in this paper has achieved good results. The baseline model is provided by the official French research cup and is written based on Jinshan Spider Net. It should be noted that spider net has now topped the HotPotQA list.

Compared with the official baseline model, the model in this paper improves the SupF1 index by 6.53%, which proves that the work done in the part of evidence prediction in this paper is effective. The improvement of AnsF1 is attributed to the work of the answer prediction module, and Joint F1 is the result of the two. Experiments show that, compared with other models, this model can predict evidence more accurately and achieve better results. In the experiment, the use of graph neural network for reasoning is not significantly

Table 7: Experimental results based on CJRC 2020 data set (bold is the optimal result).

Model	Ans-F1	Sup-F1	Joint F1
Spider Net (baseline)	70.33	74.61	57.38
BERT-GCN	06.67	76.47	64.23
BERT + CapsNet	77.68	77.42	65.30
BERT + Focal loss	71.2	75.20	58.53
BERT + ResNet2d	76.39	74.52	63.53
BERT + Bi-GRU	69.35	78.25	56.51
Our method	82.56	81.14	70.07

better than the use of CapsNet or ResNet2d for classification. After analysis, it is found that the performance of graph neural network is significantly lower than that of the model in the paragraphs where the questions or answers do not contain entities. Because the model in this paper adds a statement selection module, compared with other methods, it reduces the interference of irrelevant statements to the model. In this paper, the evidence prediction module uses the answers to help find evidence, which also improves the performance of this model.

5. Conclusion

For legal documents with clear structure and rigorous expression, it is helpful to improve human work efficiency to let machines understand and read legal documents. The purpose of reading comprehension in the legal field is to train the machine model through legal documents so that it can answer various questions according to the given case description. An excellent reading and understanding system in the legal field can assist judges, lawyers, and other professionals in their work and also make it easy for people to understand the basic situation of each case. It has a wide range of application prospects, such as crime prediction, evidence prediction, legal provisions recommendation, and intelligent court trial. This paper mainly studies the evidence prediction in the legal field. Taking the prediction of reading and understanding evidence in the legal field as the research task, this paper puts forward a prediction method of evidence based on sentence selection for legal documents. A sentence selection module is designed to remove irrelevant sentences, and questions and answers are used to infer evidence, which has achieved good results. Through experiments, it is found that the score of Joint F1 proposed in this paper is 70.07%, which is more accurate than the mainstream model.

In the following research work, we can continue to explore whether other better models have better effects on sentence selection and evidence prediction tasks. This model uses a non-end-to-end multimodule design method, which has some drawbacks. During the first stage of sentence selection, the results will affect the next step, thus affecting the results of the whole training. In the follow-up, when facing the text segment and multihop reading comprehension task with more entities, we start with the graph neural network to improve the accuracy of each stage by exploring the relevance between sentences and the relationship between different entities.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by the School of Animation, Shijiazhuang University of Applied Technology.

References

[1] S. T. Cheng, K. L. Chan, R. Lau et al., "Correction to: a multicomponent intervention for the management of chronic pain in older adults: study protocol for a randomized controlled trial," *Trials*, vol. 22, no. 1, pp. 842–851, 2021.

[2] A. E. Cust, C. Badcock, J. Smith, N. E. Thomas, L. E. Haydu, and B. K. Armstrong, "A risk prediction model for the development of subsequent primary melanoma in a population-based cohort," *British Journal of Dermatology*, vol. 182, no. 5, 2020.

[3] V. L. Girko, "Victoria transform, respect and reform methods for the proof of theg-permanent pencil law underg-lindeberg condition for some random matrices fromg-elliptic ensemble," *Random Operators and Stochastic Equations*, vol. 29, no. 2, pp. 111–148, 2021.

[4] S. Burris, A. R. Ghorashi, L. F. Cloud, R. Rebouché, P. Skuster, and A. Lavelanet, "Identifying data for the empirical assessment of law (ideal): a realist approach to research gaps on the health effects of abortion law," *BMJ Global Health*, vol. 6, no. 6, 2021.

[5] K. Sluka, L. F. Law, C. Coffey et al., "Conceptual design and protocol for the acute to chronic pain signatures program (a2cps)," *The Journal of Pain*, vol. 22, no. 5, p. 586, 2021.

[6] I. N. Bindeman, J. F. Wotzlaw, R. Stern, M. Chiaradia, M. Guillong, and D. Colón, "Geochronology and geochemistry data for the elbrus, tyrnyauz, and chegem magmatic centers, greater caucasus, Russia," *Data in Brief*, vol. 35, no. 5, Article ID 106896, 2021.

[7] M. Dahlan, "Envisioning foundations for the law of the belt and road initiative: rule of law and dispute resolution challenges," *Harvard International Law Journal*, vol. 62, 2020.

[8] J. K. Ratzloff, N. M. Law, H. T. Corbett, O. Fors, and D. S. Ser, "Robotilter: an automated lens/ccd alignment system for the evryscope," *Journal of Astronomical Telescopes, Instruments, and Systems*, vol. 6, no. 01, p. 1, 2020.

[9] E. S. Aydil, W. M. M. Kessels, S. Law, M. Sankaran, and J. M. O. Zide, "Preface for the avs peter mark award 40th anniversary collection," *Journal of Vacuum Science and Technology A*, vol. 39, no. 3, Article ID 031601, 2021.

[10] B. Liu, M. Totten, S. Nematollahi et al., "Development and evaluation of a fully automated molecular assay targeting the mitochondrial small subunit rRNA gene for the detection of pneumocystis jirovecii in bronchoalveolar lavage fluid specimens," *Journal of Molecular Diagnostics*, vol. 22, no. 12, pp. 1482–1493, 2020.

[11] G. H. Elta and R. L. Law, "What are the needed criteria for the adoption of new technology aimed at preventing duodenoscope-transmitted infections?" *Gastrointestinal Endoscopy*, vol. 92, no. 1, pp. 209-210, 2020.

[12] T. K. Katz, "Framed by the law: experimental evidence for the effects of the salience of the law on preferences," *Theoretical Inquiries in Law*, vol. 22, no. 2, pp. 21–34, 2021.
[13] S. Han, J. Liu, B. Zhao, and J. G. García, "Some partial results on the cancellation law for the tensor product of complete lattices," *Algebra Universalis*, vol. 83, no. 2, p. 15, 2022.
[14] K. Czudek, T. Szarek, and H. W. Ściążko, "The law of the iterated logarithm for random interval homeomorphisms," *Israel Journal of Mathematics*, vol. 246, no. 1, pp. 47–53, 2021.
[15] L. Huigang, H. A. Cui, Z. Xiaoli, and Y. Zhiming, "Significance of and outlook for the biosecurity law of the people's Republic of China," *Journal of Biosafety and Biosecurity*, vol. 3, no. 1, pp. 46–50, 2021.
[16] A. A. Rihan, T. A. Mohamed, and W. H. Alawee, "Physical simulation for the flow in straight and rectangular loop manifolds," *University of Baghdad Engineering Journal*, vol. 27, no. 3, pp. 15–32, 2021.
[17] A. F. Mora, "Trade mark functions in business practice: mapping the law through the search for economic content," *IIC - International Review of Intellectual Property and Competition Law*, vol. 52, no. 10, pp. 1370–1404, 2021.
[18] M. Stadlbauer and X. Zhang, "On the law of the iterated logarithm for continued fractions with sequentially restricted partial quotients," *Nonlinearity*, vol. 34, no. 3, pp. 1389–1407, 2021.
[19] F. Caponigri, L. Casini, and S. Cassese, "Securing cultural heritage? understanding the law for our monuments, artworks, and archives today," *International Journal of Constitutional Law*, vol. 19, no. 5, pp. 1683–1689, 2022.
[20] M. N. Magomedov, "On the deviation from the vegard's law for the solid solutions," *Solid State Communications*, vol. 322, no. 6, Article ID 114060, 2020.
[21] A. Sharma, R. Kumar, M. W. A. Talib, S. Srivastava, and R. Iqbal, "Network modelling and computation of quickest path for service-level agreements using bi-objective optimization," *International Journal of Distributed Sensor Networks*, vol. 15, no. 10, Article ID 155014771988111, 2019.
[22] J. Jayakumar, B. Nagaraj, S. Chacko, and P. Ajay, "Conceptual implementation of artificial intelligent based E-mobility controller in smart city environment," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1–8, Article ID 5325116, 2021.
[23] X. Liu, C. Ma, and C. Yang, "Power station flue gas desulfurization system based on automatic online monitoring platform," *Journal of Digital Information Management*, vol. 13, no. 06, pp. 480–488, 2015.
[24] R. Huang, S. Zhang, W. Zhang, and X. Yang, "Progress of zinc oxide-based nanocomposites in the textile industry," *IET Collaborative Intelligent Manufacturing*, vol. 3, no. 44, pp. 281–289, 2021.
[25] Q. Zhang, "Relay vibration protection simulation experimental platform based on signal reconstruction of MATLAB software," *Nonlinear Engineering*, vol. 10, no. 1, pp. 461–468, 2021.