

<https://doi.org/10.70517/ijhsa464357>

Research on the Review and Handling of Public Interest Litigation Evidence with the Assistance of Computer Vision

Yongjun Wang^{1,*}

¹ Law School, Henan University of Urban Construction, Pingdingshan, Henan, 467036, China

Corresponding authors: (e-mail: wyj30150603@163.com).

Abstract The wide application of artificial intelligence technology in the judicial field has brought profound changes to traditional legal practice. In this study, a public interest litigation evidence review and processing system based on Transformer and BERT model is constructed, and joint modeling of law recommendation and charge prediction is realized through a multi-task learning framework that fuses lawformer information. The methodology adopts Lawformer pre-training model for text encoding, combines the interactive attention mechanism to fuse the semantic information of the legal articles, and establishes the constraint relationship between the legal articles and the charges through the task-dependent constraint layer. The experimental results show that the MTL-LA-LJP model improves the accuracy of 0.130 in the law prediction task and 0.11 in the charge prediction task compared to CNN, and the performance advantage is more significant under the condition of small-sample data (1% training data), and the accuracy of the law prediction reaches 0.61. The study confirms the computer vision technology's effectiveness in the review of public interest litigation evidence, and provides an opportunity for the construction of intelligent justice. The study confirms the effectiveness of computer vision technology in the review of public interest litigation evidence, and provides technical support for the construction of intelligent justice.

Index Terms Computer vision technology, public interest litigation, evidence review, Transformer, BERT model, multi-task learning.

I. Introduction

Public interest litigation is a lawsuit filed in order to safeguard national interests and social public interests, and in this process, the review and handling of evidence is crucial [1], [2]. It is the key for the procuratorial authorities to identify the facts of the case and support the litigation request [3]. Without sufficient and effective evidence review and processing, public interest litigation is difficult to achieve good results [4]. Computer vision technology, as a technology that enables computers to obtain valuable information from images or videos, brings a new perspective and method to the review and processing of legal evidence [5], [6].

Computer vision technology has shown great potential in legal evidence collection [7]. In the face of massive image and video data, manual review is not only time-consuming and laborious, but also the accuracy is difficult to guarantee [8]. Computer vision technology can automatically identify important targets related to the case, such as people, vehicles, and objects, through image recognition and target detection algorithms, and screen them out, which greatly improves the efficiency and accuracy of evidence review [9]-[11]. Not only that, computer vision technology plays an irreplaceable role in evidence processing [12]. For fuzzy images or videos, computer vision can improve the clarity and quality of images through image enhancement and noise reduction techniques, so that the details that are originally difficult to recognize can be revealed [13], [14]. For example, in some surveillance videos, the facial features of suspects may not be clear due to insufficient light or long shooting distance [15], [16]. Through the technical processing of computer vision, the contrast excess and brightness of the image can be enhanced, highlighting the facial contours and features, and providing strong support for identification [17]-[18].

This study constructs a multi-task learning framework that integrates legal information, and realizes the automatic review and processing of public interest litigation evidence through deep learning technology. The study adopts a pre-trained language model as the infrastructure, combines the attention mechanism and multi-task learning strategy, and establishes a joint optimization model for the recommendation of legal articles and the prediction of charges. The model performance is verified through comparative experiments and interpretability analysis is conducted to provide theoretical basis and technical solutions for the construction of intelligent judicial system.

II. A model for reviewing and processing litigation evidence based on computer vision technology

II. A. Transformer model vs. BERT model

II. A. 1) Transformer model

Transformer model [19] is a bidirectional Transformer Encoded Representation Model (BERT) is a natural language processing model derived from the encoder of the Transformer model as a base model. The large amount of images and data in the field of computer vision provides rich training material for the Transformer model.

Since the Transformer model does not have the iterative operation of recurrent neural network, in order for the model to have the ability to recognize the word order information of the text, the positional information of each word in the text must be provided to the model, and the Transformer model solves this problem by the method of positional embedding [20]. There are two methods of positional embedding, one is to get the absolute position of each word in the text by calculating the encoding rules designed by themselves, and the second is to get the positional encoding by learning from neural network to get the relative position of each word in the text. Transformer model gets the positional encoding by the first calculation method.

Transformer model uses linear transformations of sine and cosine functions to represent the positional information of words in the text:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (1)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (2)$$

The $2i$ and $2i+1$ denote the component subscripts of the odd and even position encoding vectors, respectively. For example, for the 2nd word ($pos = 1$) in a text with a prescribed embedding dimension of 16 ($d_{model} = 16$), the positional embedding can be represented as:

$$PE_{(1)} = \left[\sin\left(\frac{1}{10000^{\frac{0}{16}}}\right), \cos\left(\frac{1}{10000^{\frac{1}{16}}}\right), \sin\left(\frac{1}{10000^{\frac{2}{16}}}\right), \dots, \cos\left(\frac{1}{10000^{\frac{15}{16}}}\right) \right] \quad (3)$$

The multi-head self-attention mechanism is derived from the self-attention mechanism. The detailed derivation process of Multihead Self-Attention Mechanism is as follows:

The multi-head self-attention mechanism is the most central structure of the Transformer model, which helps the model to accurately understand the meaning of sentences. Its main role is to learn the dependencies between the elements inside the sequence and extract the internal structural features of the text sequence. The multi-head self-attention mechanism directly compares the elements in the text sequence two by two, captures the global connection of the text sequence in one step, and solves the long-distance dependency problem of the recurrent neural network. In addition, this attention mechanism is highly efficient in parallel computation, whereas recurrent neural networks require step-by-step recursion and cannot perform parallel operations.

Transpose $MultiHead(Q, K, V)$ to be consistent with the $X_{embedding}$ dimension to get $X_{attention}$, compute:

$$X_{attention} = X_{embedding} + X_{attention} \quad (4)$$

which is the residual connection.

Layer regularization is required to normalize the hidden layers in the neural network to a standard normal distribution to speed up training and convergence:

$$Layer\ Norm(x) = \alpha \square \frac{x_{ij} - \mu_i}{\sqrt{\sigma_j^2 + \varepsilon}} + \beta \quad (5)$$

where x_{ij} is the element in $X_{attention}$, $\sigma_j^2 = \frac{1}{m} \sum_{i=1}^m (x_{ij} - \mu_j)^2$ is the $X_{embedding}$ is the variance computed in

behavioral units, $\mu_j = \frac{1}{m} \sum_{i=1}^m x_{ij}$, α, β is the trainable parameter to make up for the information lost in the

normalization process, ε is to prevent values that tend to zero in division by 0, and \square denotes elementwise multiplication. Layer regularization of $X_{attention}$ after residual concatenation yields:

$$X_{attention} = LayerNorm(X_{attention}) \quad (6)$$

II. A. 2) BERT model

BERT model [21] i.e. Bidirectional Transformer Encoded Representation Model, BERT model is actually the encoder part of the Transformer model.

BERT model is divided into two parts: pre-training task and downstream task. The pre-training task refers to the use of large datasets such as the Wikipedia corpus to train to get a good pre-training model, while the downstream task refers to the network structure transformation and parameter fine-tuning of the pre-training model according to different tasks, which is generally to transform the classifiers of the pre-training model.

The first step in the pre-training phase of the BERT model is to embed the input text data.

Since the BERT model needs to be flexibly migrated to various types of tasks, in order to deal with some tasks involving sentence pairs, some information should be provided to the model so that the model can distinguish between the upper and lower sentences in the sentence pairs.

The pre-training phase of the BERT model consists of two tasks, namely: the masked word modeling task and the next sentence judgment task.

Prediction of words with masking. Equation (7) is the expression of the softmax function and M here denotes the number of words in the text.

$$softmax(x) = \frac{e^{z_i}}{\sum_{j=1}^M e^{z_j}} \quad (7)$$

In the BERT model, by linearly transforming the final output of [CLS] and activating it with a sigmoid function, it is possible to determine whether two sentences are in context with each other. Equation (8) is the expression of the sigmoid function:

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

After pre-training, the BERT model already has the ability to capture the high-level abstract features of the text, so in the model fine-tuning stage, generally do not need to change the internal structure of the pre-trained model, only need to replace the specific classifier and fine-tune the parameters according to the needs of the downstream task, then it can be applied to text categorization and reading comprehension, named entity recognition, and other types of natural language processing tasks.

II. B. Data pre-processing

The first step of data preprocessing is to perform text segmentation processing on the case description part of the experimental data, i.e., to split the text composed of strings into text composed of words.

After the text has been completed by the word splitting process, the next step requires data cleaning of the text to remove invalid samples. Then all the punctuation marks and special characters, etc. are filtered, and in order to reduce the information dimension of the text, it is also necessary to delete the words that are too widely or frequently used.

In order to facilitate model training and model performance evaluation, the labels need to be vectorized. The paper uses unique thermal coding to binary transform the labels and realize the charge label vectorization.

The models studied in this paper all require the input data to be of the same length, and the simplest solution is to complement the text vector of the input data with 0, so that the length of all input data is consistent with the maximum input data length. However, the importance of the original semantics will be weakened due to too many

zeros supplementing the input data with shorter lengths. After comprehensive consideration, the thesis adopts different solutions according to the characteristics of the BERT model.

II. C. Multi-task Learning Review and Processing Model Incorporating Legal Information

II. C. 1) Problem definition

The LJP task is an important research task in the smart justice system, and in real-life scenarios, there are dependencies between the subtasks of LJP. Given a crime fact, a judge in a civil law country first determines the relevant law articles of the case, and the content in the corresponding law articles reveals what crime the judge should convict the defendant of, e.g., whether it is robbery or theft. The research in this chapter is based on the case fact sequence, using a multi-task learning framework to jointly model the two tasks of law recommendation and crime prediction, and formalizing the two tasks as a multi-label classification problem to predict the law and crime involved in the defendant. Assume that the case fact sequence is $f = \{f_1, f_2, f_3, \dots, f_m\}$, where m is the length of the fact sequence; the set of sub-tasks $T = \{t_1, t_2\}$, t_1 is the law recommendation task, and t_2 is the offense prediction task; y_1 and y_2 are the corresponding prediction results of tasks 1 and 2 respectively. 2 corresponding to the prediction results.

II. C. 2) Overall model structure

The multitask learning judgment prediction model MTL-LA-LJP proposed in this chapter for fusing law and order information consists of three parts, namely, the text encoding layer, the semantic fusion layer, and the task-dependent constraint layer, and the overall structure is shown in Figure 1. The multi-task learning framework requires shared parameters, and the main part of the model proposed in this chapter adopts hard sharing of parameters.

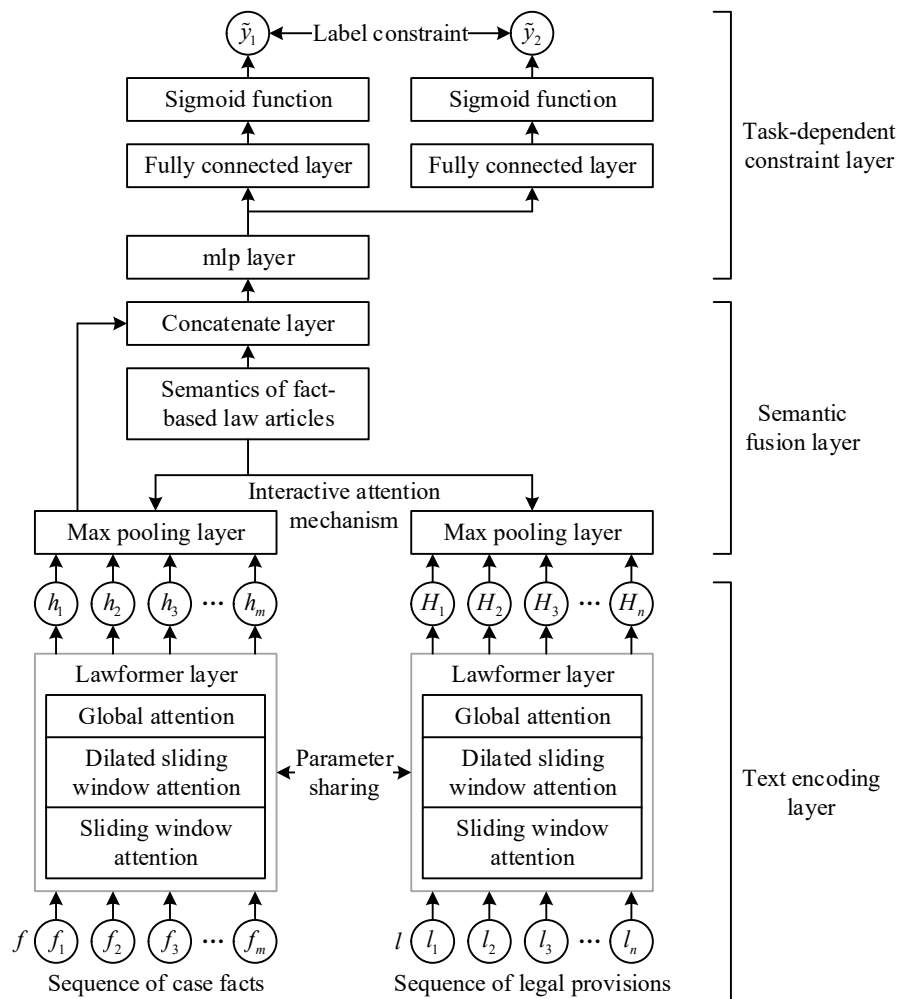


Figure 1: Structure diagram of the multi-task learning decision prediction model

II. C. 3) Text encoding layer

The text encoding layer generates dynamic word vector representations rich in contextual semantic information based on the sequence of case facts and legal content. Due to the strong linguistic representation capability of the pre-trained model, the pre-trained language model represented by BERT is chosen as the text encoder in this chapter. Considering that case facts and law sequences are usually long texts and there are a large number of proprietary terms in the legal texts, it is difficult to understand the deep law-related logical relationships using the pre-trained models in the general domain, which affects the performance of the models, Lawformer, a pre-trained model with prior knowledge of the law, is selected to encode the case facts and law texts with word vectors.

Lawformer internally combines three different attention mechanisms.

The first attention mechanism is the sliding window attention mechanism, which computes attention by focusing only on the character information within a fixed window around it. Assuming that the size of the sliding window is w , computing the attention value for a character at a certain position will only focus on information about characters within $1/2w$ of the left and right.

The second attention mechanism is the null sliding window attention mechanism, which can focus on a larger range of character information when computing attention. This attention mechanism is similar to the null convolutional model in that the sliding window range can be further expanded. However, each window is not consecutive, and there is an interval d when attending to character information.

The third type is the global attention mechanism, which computes attention to all characters in the sequence. There are some specific tasks that require certain characters to be able to attend to all the information of the entire sequence.

Given a sequence of case facts as $f = \{f_1, f_2, f_3, \dots, f_m\}$ and the sequence of laws $l = \{l_1, l_2, l_3, \dots, l_n\}$, where m and n are the fact sequence length and the normal sequence length, respectively. The fact sequence is fed into the pre-trained model Lawformer to obtain the dynamic word vector representation:

$$h_1, h_2, \dots, h_m = \text{Lawformer}(f_1, f_2, \dots, f_m) \quad (9)$$

$h \in R^{m \times d}$ is the hidden vector representation of the sequence of facts, and d is the dimension of the Lawformer hidden layer. The sequence of facts is input to the Lawformer with shared parameters for vector representation:

$$H_1, H_2, \dots, H_n = \text{Lawformer}(l_1, l_2, \dots, l_n) \quad (10)$$

$H \in R^{n \times d}$ is the hidden vector representation of the French bar sequence. Then the hidden vector representations of the two sequences are used to retain the important features using a maximum pooling strategy to obtain the contextually relevant semantic interrogatives:

$$\bar{h} = \max_pooling(h) \quad (11)$$

$$\bar{H} = \max_pooling(H) \quad (12)$$

$\bar{h} \in R^d, \bar{H} \in R^d$ of them.

II. C. 4) Semantic Fusion Layer

Given a case fact and a law library, context-sensitive semantic vectors are obtained after a text encoding layer, where the vector of case facts is denoted as $\bar{h} \in R^d$, and the vector of a particular law in the law library is denoted as $\bar{H}_i \in R^d$, and the range of values of i is $[1, p]$, where p is the number of legalbars in the legalbars database. Then the semantic information of each legal article in the legal article database is fused using the interactive attention mechanism, and the attention weight is calculated as:

$$\beta_i = \bar{h}^T W \bar{H}_i, i = [1, 2, \dots, p] \quad (13)$$

where $W \in R^{d \times d}$ is a learnable matrix. The meaning of this formula is to calculate the semantic similarity between the semantic vectors of the case facts and the candidate laws in the law library, and the weight values with high similarity will be high. Then the weight values are normalized using the softmax function to turn the weight values into a numerical distribution that sums to one:

$$a_i = \frac{\exp(\beta_i)}{\sum_{j=1} \exp(\beta_j)} \quad (14)$$

According to the normalized weight values to fuse the semantic information of all the legal articles in the law library, the more relevant to the facts of the case the fusion of the legal articles weight will be greater, the vector that fuses the semantic information of all the legal articles is:

$$L = \sum a_i \bar{H}_i \quad (15)$$

where $L \in R^d$ is the case-related legal semantic vector. Then the case fact semantic vector \bar{h} and L are spliced for feature fusion:

$$FL = CAT(\bar{h}, L) \quad (16)$$

where $FL \in R^{2d}$, CAT is the splice function.

II. C. 5) Task dependency constraint layer

After obtaining the hybrid feature FL through the semantic fusion layer of the French article, the feature dimensionality reduction is first performed by a multilayer perceptron:

$$Q = mlp(FL) \quad (17)$$

where $Q \in R^d$, the parameters are still shared between the two tasks at this point. Then the feature dimension is downgraded through a fully connected layer, the parameters are not shared between the two tasks in the fully connected layer, and the left side will be downgraded to the number of categories of the normal bar through the fully connected layer feature dimension:

$$Q_1 = FC_l(Q) \quad (18)$$

where $Q_1 \in R^{N_l}$, and N_l is the number of categories of the normal bar; the right-hand side passes through the fully-connected layer the feature dimensions are reduced to the number of categories of the charge:

$$Q_2 = FC_c(Q) \quad (19)$$

where $Q_2 \in R^{N_o}$ and N_o is the number of categories of the offense. Q_1 and Q_2 are the feature representations of the law and offense predictions.

The features are transformed into predicted probability values for each category using the activation function sigmoid:

$$\tilde{y}_1 = \text{sigmoid}(Q_1) \quad (20)$$

$$\tilde{y}_2 = \text{sigmoid}(Q_2) \quad (21)$$

\tilde{y}_1 is the predicted probability value for each category of the statute, and \tilde{y}_2 is the predicted probability for each category of the offense.

For the two tasks of statute recommendation and offense prediction, the binary cross-entropy loss function is used to compute the respective losses:

$$loss_l = - \sum_{i=1}^{N_l} (y_i \log \tilde{y}_i + (1 - y_i) \log(1 - \tilde{y}_i)) \quad (22)$$

$$loss_c = - \sum_{i=1}^{N_c} (y_i \log \tilde{y}_i + (1 - y_i) \log(1 - \tilde{y}_i)) \quad (23)$$

$loss_l$ is the loss of the statute recommendation task, $loss_c$ is the loss of the charge prediction task, y_i is the true label, and \tilde{y}_i is the predicted probability value.

By observing the relationship between the label of the statute and the label of the charge, it is found that there is a constraint relationship between the statute and the charge. Changes were made to the loss function for the charge prediction task to motivate the model to learn the constraint relationship between the labels, and the charge prediction loss function was modified as follows:

$$loss_c = \begin{cases} loss_c, & \text{if } \tilde{y}_1 \neq y_1 \\ -\sum_{i=1}^{N_c} mask_i (y_i \log \tilde{y}_i + (1 - y_i) \log(1 - \tilde{y}_i)), & \text{if } \tilde{y}_1 = y_1 \end{cases} \quad (24)$$

When the law recommendation prediction result is inconsistent with the true label, the loss function of the charge prediction task is the ordinary binary cross entropy loss function; when the law recommendation prediction result is consistent with the true value, each law related to the facts has its own permitted charges, and these permitted charges form a set, and the mask value of the charges in the set takes the value of 1 when calculating the loss, and when calculating the loss for the charges not in the set mask takes the value 0, which is equivalent to not calculating the loss value. The loss function for the entire model is as follows:

$$L = \lambda_l loss_l + \lambda_c loss_c \quad (25)$$

III. Experimental results and analysis of the review and processing model

III. A. Experimental setup

For the coding layer the official Chinese BERT model is used as the fact encoder. The factual descriptions are processed as character sequences, and the maximum input sequence length is set to 256. The Adam optimizer is used in the experiments, and the learning rate is set to 2.2×10^{-5} , epoch is set to 4, and batch_size is set to 18. The BERT is also fine-tuned by the characteristics of the judicial decision prediction dataset to improve the downstream task.

At the causal clustering composition layer, different training ratios (1%, 10%, 40%, 60%, and 100%) are selected in this paper to investigate how the performance gap varies with different training data available. Also, since different training ratios result in different number of keyword occurrences. Therefore, when the training set ratio is 1%, 30 keywords are selected and these keywords are clustered into 20 cause elements; when the training set ratio is 10% and 40%, 40 keywords are selected and these keywords are clustered into 30 cause elements; when the training set ratio is 60% and 100%, 50 keywords are selected and these keywords are clustered into 35 cause elements. In addition, all other hyperparameters are chosen empirically and remain constant for different models in the same dataset and the same training set ratio.

III. B. Comparative tests of different models

In this paper, experiments are conducted on two judicial decision prediction subtasks, namely, law prediction and charge prediction, and the experimental results are shown in Tables 1 and 2, respectively. By comparing the experiments of MTL-LA-LJP on the two subtasks with the effective baseline models CNN and RNN on the judicial verdict prediction task, both show that the performance of this paper's model MTL-LA-LJP is generally better than the performance of inference using neural networks alone. Which when trained on 100% of the data, respectively, improves by 0.130 on the forensic prediction task compared to the CNN, and 0.11 on the charge prediction task. Meanwhile, after adding causal ideas to the CNN, and RNN, respectively, the model's effectiveness is improved, and in the forensic prediction of 100% of the training data, the LSTM+Casual is 0.5 times more accurate than the use of the LSTM alone to perform the judgment inference accuracy is 0.18 higher. This is because neural network models such as CNN, RNN are difficult to directly identify distinguishable elemental words in confusing judgment results, although they have the advantage of learning elemental words from a large amount of unstructured data. And the causal graph constructed by the model in this paper can carefully judge the confusable elemental words again on the basis of the neural network judgment, and can obtain better reasoning ability.

At the same time, this paper's model helps to judge the condensed key information, especially in the small sample data the better this paper's model will perform. From the analysis of the experimental results, on the one hand, the causal-based method performs more stably than the neural network method in small sample data; on the other hand, the neural network tends to be mismatched under the training of a small amount of data. However, the performance gap becomes smaller as the training data increases. MTL-LA-LJP is 0.27 more accurate than CNN when 1% of data is used as training data in the law prediction task, which is much higher than the experimental results with 100% training data.

Table 1: Macro F1 value of different models on the method prediction task

Model	Law prediction				
	1%	10%	40%	60%	100%
LSTM	0.33	0.40	0.55	0.65	0.70
CNN	0.34	0.45	0.58	0.73	0.84
LSTM+Casual	0.50	0.56	0.65	0.82	0.88
CNN+Casual	0.57	0.59	0.74	0.84	0.89
MTL-LA-LJP	0.61	0.68	0.79	0.90	0.97

Table 2: Macro F1 value of different models in crime prediction task

Model	Law prediction				
	1%	10%	40%	60%	100%
LSTM	0.30	0.33	0.42	0.53	0.65
CNN	0.32	0.38	0.49	0.64	0.78
LSTM+Casual	0.48	0.53	0.58	0.67	0.74
CNN+Casual	0.49	0.57	0.693	0.73	0.85
MTL-LA-LJP	0.52	0.59	0.69	0.79	0.89

III. C. Descriptive statistical analysis of data

Table 3 shows the statistical information related to the text length of each part of the dataset. The length of the litigation request part is the smallest, the average length is only 78 tokens, and the fluctuation of the data text length distribution is relatively small. The text length of the factual description part is 280 tokens on average, and the distribution of text length is relatively large, the longest text even reaches 11,578 tokens. The text of the courtroom record part is generally longer, the average length of each courtroom record reaches 1,780 tokens, which is much higher than the litigation request and the factual description, and the length of more than half of the data is above 1,595 tokens. More than half of the data is longer than 1595 tokens.

Such text length far exceeds the maximum input length limit of 256 tokens in the base model such as BERT. When the input text is too long, the model has to adopt processing methods such as segmenting by sentence, intercepting by fixed length or applying the sliding window strategy, etc., and the output differences and effects caused by different processing methods need to be further explored. In view of this, in this paper, we set up two kinds of experiments, namely, segmentation by sentence and fixed-length interception, to analyze the impact of different text-processing strategies on the performance of the model, and to provide empirical evidence for further optimization of the model.

Table 3: Data set text length statistics

Data name	Mean length	Maximum length	Minimum length	25%	50%	75%
Litigation request	78	940	0	52	76	103
Court record	1780	11364	32	1162	1595	2149
Factual description	280	11578	6	125	201	336

III. D. Interpretability analysis

The pre-training model encodes the text with both word granularity and word granularity, and the gradient information of word granularity, although it can be interpreted to a certain extent, is more fragmented, which weakens the interpretability of the experimental results to a certain extent, as shown in Figure 2 for a sample of the dataset. Considering the experimental arithmetic resource limitation and the problem of model characteristics, it is more difficult to modify it to word granularity in the pre-training step. To make up for this defect, this paper obtains the interpretive score and then performs the word-splitting operation on the text, and the score of the word is the sum of the scores of its corresponding words, so as to obtain the importance of the word granularity information. The specific experimental process is as follows:

Firstly, the Jieba word segmentation tool was used to segment the text, and the score was calculated. For example, if the score of "month" is 0.15 and the score of "borrowing" is 0.25 in the original attention mechanism, the score of the word "monthly interest rate" is 0.40. In the same way, you can score specific points for other words in the text after the word. Finally, the content of the attention mechanism is visualized according to the score after word

segmentation. Figure 3 shows the visualization results of the word granularity attention mechanism score, compared with the word granularity shown in Figure 2, the explanation of the contribution of each word after word segmentation is more intuitive, and it can be clearly seen that the key information such as "remittance", "loan" and "term" in the text are highlighted, making the results more explanatory and providing better evidence that the model predicts the outcome of litigation claims.

Then, combined with specific cases and the elemental trial hierarchy concept tree of private lending disputes, the interpretability of the model is analyzed based on the attention mechanism.

Example of the plaintiff's claim: "Claim: 1. Order the defendant to repay the loan of 200,000 yuan and interest (the interest is calculated at a monthly interest rate of 3% from September 28, 2015 to the date of actual repayment); 2. The litigation costs of this case shall be borne by the defendant. "

The model pays attention to the keywords such as "plaintiff", "bank remittance" and "remittance", which indicate that the two parties agree to borrow in this case, and the value of "whether the lending relationship is established" in the second-level concept tree is 1. At the same time, the case can be successfully accepted, so the value of "whether the statute of limitations has expired" is 1, and the judgment result of the "contract effect" dimension in the first-level concept tree is deduced: the loan has contractual effect. Similarly, the model pays attention to keywords such as "monthly interest rate", "3%", "calculation" and "expiration of the loan period", indicating that the borrower and the borrower have agreed on the interest rate and loan term. In addition, the key words such as "defendant", "so far", "unpaid" and "repaid" indicate that the lender has not yet repaid, so it can be seen that the values of "whether to agree on the loan term", "whether to agree on the default clause", "whether to agree on the interest rate" and "whether there is repayment behavior" in the second layer of the hierarchical concept tree are 1, 1, 1, and 0 in order, and the judgment result of the "repayment rules" dimension in the upper layer of the hierarchical concept tree is as follows: according to the trial rules of private lending, the principle of "there is an agreement and an agreement to follow" should be followed, and it should be implemented in accordance with the agreement. The amount of principal payable remains unchanged.

In summary, the plaintiff's claim 1 in this case was predicted to be "supportive" for claim 1 to "order the defendant to repay the loan of 200,000 yuan and interest (the interest was calculated at a monthly interest rate of 3% per annum from September 28, 2015 to the date of actual repayment)", and according to the provision that the litigation costs of private lending shall be borne by the losing party, the establishment of claim 1 made claim 2 "the litigation costs in this case shall be borne by the defendant" was also predicted to be "support". The prediction result of the model is the same as the actual judgment of the case. This paper confirms the reliability of the model review and processing results.

The defendant repaid the loan for \$200,000 and interest,
The interest is from September 28,
2015 to the day of actual repayment,
According to the monthly rate of 2%,
The legal fee of the case is borne by the defendant.
In the case of the trial,
the defendant borrowed 200,000 yuan from the plaintiff.
And issue the loan according to one portion, the agreed repayment date,
The monthly rate is calculated by 3%,
The plaintiff sent the defendant \$200,000 through bank remittance.
After the term of the loan expires,The defendant has not been repaid.

Figure 2: Data set sample display

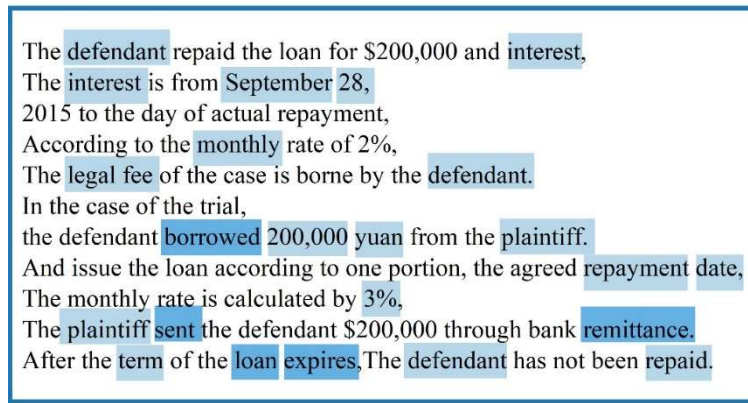


Figure 3: Word size attention mechanism scores

Figure 4 shows the word size attention mechanism score. The model pays attention to the key sentences related to the information about the qualification of the subject of the lawsuit, the civil litigation process specification and also the main information about the qualification of the subject of the lawsuit, the civil litigation process specification and so on.

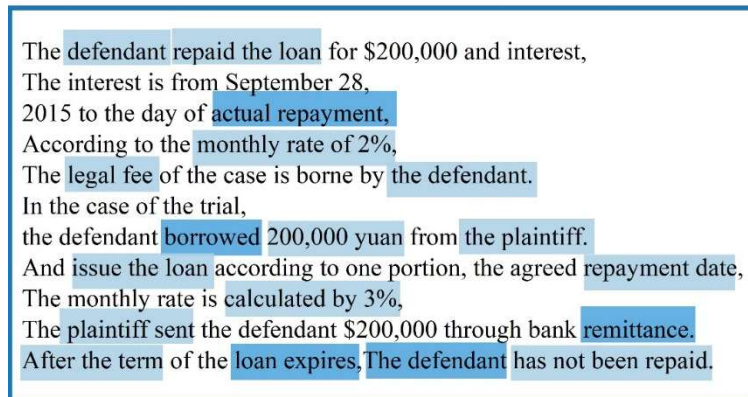


Figure 4: The word size attention mechanism scores

IV. Conclusion

In this study, we constructed a public interest litigation evidence review and processing model based on computer vision technology, and verified the effectiveness of the deep learning method in the judicial verdict prediction task. The MTL-LA-LJP model shows excellent performance under different training data ratios, especially in the case of data scarcity the advantage is obvious, and it achieves an accuracy rate of 0.61 in the forensic article prediction task when using 1% training data, which It significantly outperforms the traditional neural network approach. The multi-task learning architecture of the model effectively utilizes the intrinsic correlation between laws and charges, and improves the prediction accuracy through the task-dependent constraint mechanism. Experiments show that the improved neural network method based on the causal idea is more stable in small sample scenarios, and the accuracy of LSTM+Casual is 0.18 higher than that of LSTM alone in the task of statutory prediction. The statistical analysis of text length reveals the differences in the characteristics of different types of evidence, which provides data support for the optimization of the model. The interpretability analysis of the attention mechanism confirms that the model is able to accurately identify key evidence elements, providing a credible basis for decision-making in judicial practice and promoting the in-depth application of artificial intelligence technology in the field of public interest litigation.

References

- [1] You, W., Liang, S., Feng, L., & Cai, Z. (2023). Types of environmental public interest litigation in China and exploration of new Frontiers. *International Journal of Environmental Research and Public Health*, 20(4), 3273.
- [2] Cummings, S. L. (2020). Public interest litigation in comparative perspective. *Australian Journal of Human Rights*, 26(2), 184-194.
- [3] Helmers, C., & Love, B. J. (2023). Patent validity and litigation: Evidence from us inter partes review. *The Journal of Law and Economics*, 66(1), 53-81.

- [4] Mohammad, S., & Karim, T. (2019). Role of NGOs in Developing Public Interest Litigation: An Analytical Study. *Environmental Policy and Law*, 49(2-3), 145-152.
- [5] Aronson, J. D. (2018). Computer vision and machine learning for human rights video analysis: Case studies, possibilities, concerns, and limitations. *Law & Social Inquiry*, 43(4), 1188-1209.
- [6] Radeva, E. (2021). The potential for computer vision to advance accountability in the Syrian crisis. *Journal of International Criminal Justice*, 19(1), 131-146.
- [7] Gless, S. (2019). AI in the Courtroom: a comparative analysis of machine evidence in criminal trials. *Geo. J. Int'l L.*, 51, 195.
- [8] Shaligar, S., Arefnia, T., & Amiri, M. M. (2024). Applications of Artificial Intelligence in the Production and Use of Digital Documents and Electronic Evidence as Proof in Civil and Criminal Litigation. *Legal Studies in Digital Age*, 3(2), 10-30.
- [9] Fraser, H., Simcock, R., & Snoswell, A. J. (2022, June). Ai opacity and explainability in tort litigation. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 185-196).
- [10] von Lewinski, K., Beurskens, M., & Scherzinger, S. (2024). Data modelling as a means of power: At the legal and computer science crossroads. *Computer Law & Security Review*, 52, 105865.
- [11] Beckedorf, J., Hartung, D., & Sittig, P. (2020). Analyzing high volumes of German court decisions in an interdisciplinary class of law and computer science students. In *Computational Legal Studies* (pp. 328-344). Edward Elgar Publishing.
- [12] Thakkar, P., Patel, D., Hirpara, I., Jagani, J., Patel, S., Shah, M., & Kshirsagar, A. (2023). A comprehensive review on computer vision and fuzzy logic in forensic science application. *Annals of Data Science*, 10(3), 761-785.
- [13] Idrees, H., Shah, M., & Surette, R. (2018). Enhancing camera surveillance using computer vision: a research note. *Policing: An International Journal*, 41(2), 292-307.
- [14] Gupta, A. (2019). Current research opportunities for image processing and computer vision. *Computer Science*, 20, 387-410.
- [15] Noiret, S., Lumetzberger, J., & Kampel, M. (2021, December). Bias and fairness in computer vision applications of the criminal justice system. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1-8). IEEE.
- [16] Grimm, P. W., Grossman, M. R., & Cormack, G. V. (2021). Artificial intelligence as evidence. *Nw. J. Tech. & Intell. Prop.*, 19, 9.
- [17] Ahmed, S., Khan, M. F., Singh, B., Singh, N., & Sharma, B. (2025). Enhancing Crime Scene Analysis: The Impact of AI Technologies on Evidence Processing. In *Forensic Intelligence and Deep Learning Solutions in Crime Investigation* (pp. 63-84). IGI Global Scientific Publishing.
- [18] Klymchuk, M., Marko, S., Priakhin, Y., Stetsyk, B., & Khytra, A. (2021). Evaluation of forensic computer and technical expertise in criminal proceedings. *Amazonia Investiga*, 10(38), 204-211.
- [19] Feng Yang, Xi Liu, Botong Zhou, Xuehua Guan, Anyong Qin, Tiecheng Song... & Chenqiang Gao. (2025). Aerial video classification with Window Semantic Enhanced Video Transformers. *Expert Systems With Applications*, 285, 127883-127883.
- [20] Sicong Zang & Zhijun Fang. (2025). Equipping sketch patches with context-aware positional encoding for graphic sketch representation. *Computer Vision and Image Understanding*, 258, 104385-104385.
- [21] Wenxuan Xing, Jie Zhang, Chen Li & Gaifang Dong. (2025). iAMP-EmGCN: A new design for identifying antimicrobial peptides based on BERT and Graph Convolutional Network. *Expert Systems With Applications*, 283, 127811-127811.