International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 8188-8200

https://doi.org/10.70517/ijhsa464701

Optimization Strategies for Labor Dispute Resolution Mechanisms in Artificial Intelligence Scenarios

Meng Mei[1,*]
[1] School of Public Administration, Hunan Labor and Human Resources Vocational College, Changsha, Hunan, 410100, China
Corresponding authors: (e-mail: gracemei12@163.com).

Abstract In the integration of artificial intelligence technology with the judicial field, due to issues such as the complexity of legal texts and the large number of disputed issues, the labor dispute resolution mechanism still needs to be optimized and improved. In the retrieval of labor dispute cases, this paper proposes a similar case matching model consisting of four modules: an embedding layer, an attention layer interaction, a pooling layer, and a prediction layer. This model enhances the interactivity between text representations by adding an attention mechanism to the expression model. Additionally, it introduces a pairwise comparison (PairWise) task to promote the model's learning of relative ranking position information, designing a multi-task training method that combines sentence pair ranking. In identifying the focal points of labor disputes, this paper employs a classification network module based on LegalBert for category prediction, utilizes the Skip-Gram model to achieve vector representation of text, and uses a BiLSTM-Attention representation module to operate on sentence matrices. Through the prediction output layer, the focal points of disputes are identified, thereby establishing a dispute focal point identification model based on multi-feature fusion. Subsequently, a statistical analysis module for labor dispute-related legal data was constructed, which, together with the proposed case matching model and dispute focus identification model, constitutes the labor dispute resolution system mechanism. After 25 rounds of training, the proposed labor dispute resolution system achieved an accuracy rate of 90.69% with a loss of only 21.40%.

Index Terms labor dispute resolution, dispute focus identification, case matching, artificial intelligence

I. Introduction

In recent years, changes in China's population structure, technological research and development, and economic conditions have weakened the country's labor cost advantage, threatening its status as the “world's factory.” The development of artificial intelligence (AI) technology represents a critical opportunity and key driver for China's transition from “Made in China” to “Intelligent Manufacturing in China” [1], [2]. Artificial intelligence, also known as machine intelligence, refers to the use of ordinary computers to simulate certain human cognitive processes and intelligent behaviors, such as learning, reasoning, thinking, and planning, enabling computers to achieve higher-level applications [3]-[5]. China's manufacturing sector began applying artificial intelligence technology as early as 2012, and it has since spread to multiple industries including services, finance, and retail [6]. While enhancing production efficiency, AI also exhibits certain characteristics of “replacing humans with machines.” Therefore, while developing AI, it is essential to address the potential issues it may bring, such as unemployment, changes in the labor force structure, and income inequality [7]-[9]. The outcomes of previous industrial revolutions have demonstrated that the benefits of technological progress have not been equitably shared by society. While infringing upon the rights of workers, such advancements may also increase the burden on enterprises, disrupt harmonious labor-management relations, and undermine social stability, leading to frequent collective labor disputes. The impact of AI on the labor sector may be even more severe, prolonged, and widespread [10]-[13].

The mechanism for resolving labor disputes under the trend of AI development is a new issue that has emerged since the rise of AI technology. From existing research literature, most studies have been conducted from either a legal or an economic perspective [14]. For example, literature [15] analyzes the impact of AI on labor relations, explores the conflicts arising from its application, and proposes solutions to balance corporate efficiency and worker rights protection through social negotiation and transparent policies. Literature [16] found that the number of job vacancies related to AI is increasing rapidly, driven by institutions with matching task structures, rather than a reduction in hiring for non-AI-related positions, and that skill requirements are changing, though the overall impact is too small. Literature [17] explores the potential impacts of AI and automation on the labor market, identifies the challenges in understanding these impacts, and proposes improvements to better predict labor trends and enhance adaptability to technological changes. Literature [18] proposes the use of artificial intelligence technology to develop a labor dispute resolution method (LDMLSV), which utilizes SHAP and a soft voting strategy to predict the key paths in labor dispute resolution, achieving an accuracy rate of up to 0.90, and demonstrates its effectiveness in improving the efficiency of dispute resolution. Literature [19] explores the application of artificial intelligence in dispute resolution, focusing on the development of a user-centered intelligent online dispute resolution system to provide effective support for labor dispute resolution.

This paper first discusses the framework composition of the case matching model for the ranking task of collaborative sentences, with a particular focus on the formal mathematical representation and computational operations of the text attention interaction mechanism in the expression case matching model. It then integrates PairWise and contrastive learning ideas into the text retrieval task and collaborates with PairWise multi-task training. Second, based on the definition of dispute focal points in labor arbitration, a dispute focal point identification model based on multi-feature fusion SKIPGRAM is constructed, with the four modules of the model described in sequence, along with their respective functions and implementation steps. Third, the performance of the proposed case matching model is evaluated through rule matching experiments, comparative experiments, and module ablation tests. Using the LcCaRD test set as experimental data, the identification performance of the proposed dispute focal point identification model is compared with that of similar models. Finally, a legal data statistical analysis module is established, and a labor dispute resolution system mechanism is constructed to train and analyze the system model's performance.

II. A case matching model for collaborative sentence pairing tasks

The case matching model proposed in this chapter, which combines expressions and interactive structures, is shown in Figure 1. The main components of the model include the embedding layer, attention interaction layer, pooling layer, and prediction layer.

The embedding layer is the foundation layer of the model and adopts a twin network structure, containing two parallel pre-trained language models, which are used to encode the two input texts, respectively. The two models share parameters, with each model independently processing its respective text to convert the original text into high-dimensional semantic representation vectors.

The attention interaction layer (Attention) is located above the twin network layer, with its primary function being to facilitate information exchange between the two text representations output by the twin network. Through the attention mechanism, the model can learn how to adjust its focus on one text based on the content of the other text, thereby capturing the relevance and semantic connections between the two texts.

The pooling layer aggregates the text representations processed by the interaction layer into a fixed-length vector. This process can employ various strategies, such as max pooling, average pooling, or custom pooling strategies to capture the most important features.

The output layer converts the output of the pooling layer into a final similarity score, which directly reflects the semantic similarity between the two texts.

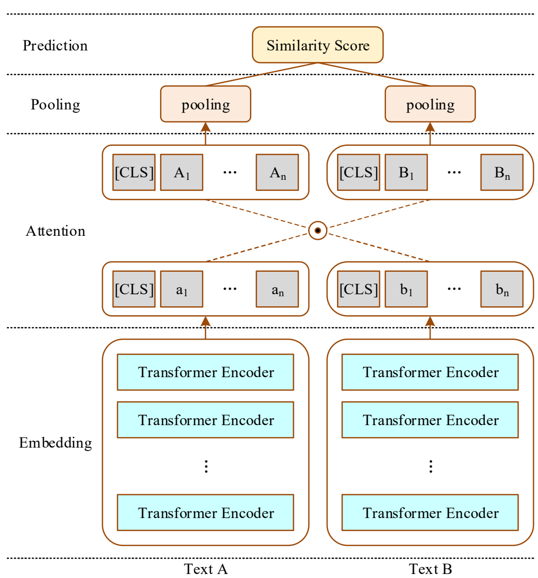


Figure 1: Model structure

II. A.Expression case matching model based on attention mechanism

In the legal domain, the text describing the case facts to be queried is defined as Q, and the candidate case fact description texts are defined as D. This paper uses the BERT pre-trained language model as a text representation encoder, mapping text into dense vectors in a low-dimensional space. For Q and D, after language model encoding, we obtain $\hat{q}=\{e_{q_1},e_{q_2},\cdots,e_{q_n}\}$ and $\hat{d}=\{e_{d_1},e_{d_2},\cdots,e_{d_n}\}$, where e_{q_i} and e_{d_i} represent the embedding vectors of the i th word in the two texts, respectively, and the dimension is determined by the parameters of the pre-trained language model. The text attention interaction mechanism is formally described as follows:

For the query case representation $\hat{q}$ and $\hat{d}$, through the attention mechanism in the interaction layer, the interaction of text representations can obtain the new representation of case Q and case D after interaction, as shown in equations (1)-(2):

$$\hat{q}_{interact}=Attention\left(Q=\hat{d},K=\hat{q},V=\hat{q}\right) \tag{1}$$

$$\hat{d}_{interact}=Attention\left(Q=\hat{q},K=\hat{d},V=\hat{d}\right) \tag{2}$$

In the pooling layer, the outputs $\hat{q}_{interact}$ and $\hat{d}_{interact}$ from the interaction layer are aggregated with the original text representations $\hat{q}$ and $\hat{d}$ using an average pooling operation, as shown in Equations (3)-(4):

$$p_q=Pool_{avg}\left(\hat{q},\hat{q}_{interact}\right) \tag{3}$$

$$p_d=Pool_{avg}\left(\hat{d},\hat{d}_{interact}\right) \tag{4}$$

Finally, the p_q and p_d outputs from the pooling layer are used to calculate similarity. This paper uses the cosine similarity function to evaluate the similarity between two texts. Cosine similarity is a commonly used similarity measure that calculates the similarity between two vectors in terms of direction. The calculation is performed as shown in Equation (5):

$$\mathrm{Cos}(q,d)=\frac{p_q\cdot p_d}{\left\|p_q\right\|\times\left\|p_d\right\|} \tag{5}$$

In this context, $\cdot$ denotes the dot product operation of vectors, $\left\|p_q\right\|$ and $\left\|p_d\right\|$ represent the vector lengths of p_q and p_d, respectively, and $\mathrm{Cos}(q,d)$ represents the similarity between p_q and p_d, with values ranging from -1 to 1.

II. B.Multi-task training with PairWise collaboration

In text retrieval tasks, commonly used loss functions for fine-tuning models in downstream tasks include TripletLoss, ContrastiveLoss, and InfoNCELoss. The core objective is to maximize the similarity scores of positive sample pairs while minimizing the scores of negative sample pairs. While these loss functions are effective in training models across various scenarios, they overlook the information related to text pair ranking. As a result, they may fail to adequately capture the subtle semantic differences between text pairs, leading to issues with the accuracy of ranking results in retrieval tasks. Therefore, to address this issue, this section introduces a pairwise similarity loss based on the contrastive learning loss training method, inspired by the pairwise comparison (PairWise) approach. The Pairwise method can be viewed as a special case of contrastive learning. Compared to conventional loss functions such as InfoNCE, pairwise similarity loss is more aligned with the objectives of retrieval tasks, which involve ranking based on the relevance between texts. It not only focuses on the representation of individual texts but also on capturing the contrasting features between text pairs. By combining these two tasks, the model can better identify and distinguish between similar and dissimilar sentence pairs, thereby enhancing the extraction of textual semantic information and improving the accuracy and efficiency of text retrieval.

In this subsection, to better understand the application of different tasks in the model fine-tuning process, the data is uniformly defined as follows:

For query text A, there is an annotated dataset $G=\{B_4,C_3,D_1,E_0,\cdots\}$ where 4, 3, 1, and 0 are labels representing their similarity to text A. Label 4 indicates the highest similarity, while label 0 indicates the lowest similarity. When the label value is greater than 2, it is considered a positive example A^+ of the query text A, otherwise it is considered a negative example A^- of the text A.

(1) Training method based on contrastive learning

In the training method based on contrastive learning, this section uses InfoNCELoss as the contrastive learning loss function. The model fine-tuning process is shown in Figure 2. where E represents the model's representation layer. For the text vector representations output by E, scores are calculated using the cosine similarity function, with the representation of the query text A as the anchor. Texts B and C from the labeled dataset G that have high similarity with A are selected as positive samples A^{+}, while texts D and E with low similarity are selected as negative samples A^{-}. The model's optimization objective is to maximize the similarity between the query text A and its positive sample A^{+} in the semantic space, while simultaneously maximizing the difference from the negative sample A^{-}, as shown in Equation (6):

$$L_{InfoNCE} = -\log \frac{e^{\cos\left(A, A^{+}\right)}}{e^{\cos\left(A, A^{+}\right)} + \sum_{A^{-} \in G^{-}} e^{\cos\left(A, A^{-}\right)}} \tag{6}$$

Among them, $\cos(u,v)$ is the cosine similarity function between u and v, A^{+} is the positive sample, and G^{-} is the negative sample set.

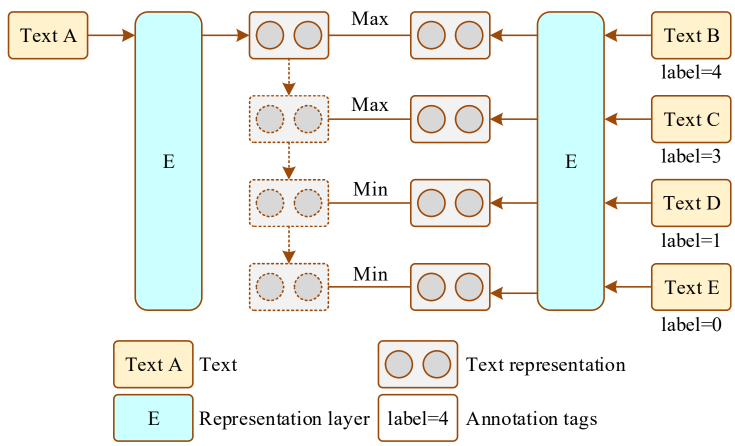


Figure 2: Model fine-tuning process

(2) PairWise-based training method

In the PairWise training method, this section uses Pair Wise Ranking Loss as the contrastive learning loss function, with the task objective shown in Figure 3, transforming the model's optimization objective into: for the same text A, the similarity score of text pair $P_{ab} = \langle A, B \rangle$ should be much higher than that of $P_{ac} = \langle A, C \rangle$, similarly, the similarity score of P_{ac} should be much higher than that of P_{ad}, and so on. Therefore, the loss does not include specific positive or negative labels, but rather compares the similarity scores of "$P_{aa^{+}} = \langle A, A^{+} \rangle$ and $\mathrm{Cos}\left(P_{aa^{+}}\right)$" with those of "$P_{aa^{-}} = \langle A, A^{-} \rangle$ and $\mathrm{Cos}(P_{aa^{-}})$," as shown in Equation (7):

$$Loss_{PairWise} = \log\left(1 + \sum_{sim\left(A, A^{+}\right) > sim\left(A, A^{-}\right)} e^{\lambda\left(\cos\left(A, A^{+}\right) - \cos\left(A, A^{-}\right)\right)}\right) \tag{7}$$

Among these, $sim(u,v)$ denotes the label of the sample pair $P_{uv} = \langle u, v \rangle$, $\cos(u,v)$ is the cosine similarity function between u and v, A^{+} is the positive sample, and A^{-} is the negative sample.

(3) Multi-task training method with collaborative PairWise strategy optimization

In this section, we propose a multi-task training framework by combining contrastive learning and PairWise strategy. This framework aims to simultaneously optimize the similarity scores and ranking losses of text pairs. Through this joint training method, the model not only learns to distinguish between similar and dissimilar text pairs but also more accurately reveals the relative relationships between text pairs at the semantic level, thereby significantly improving the accuracy and efficiency of text retrieval, as shown in Equation (8):

$$Loss_{Total} = \alpha Loss_{InfoNCE} + \left(1 - \alpha\right) Loss_{PairWise} \tag{8}$$

Among them, α is a weighting coefficient used to balance the contributions of the two loss functions.

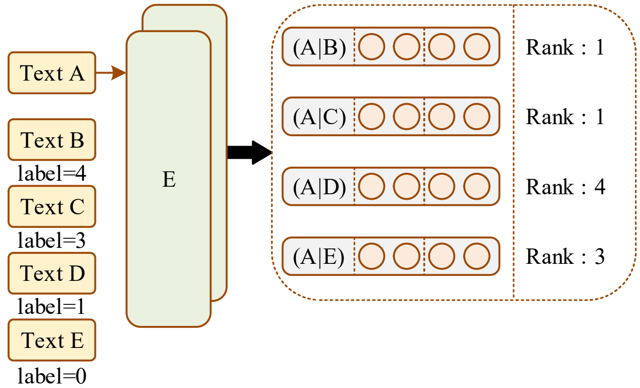


Figure 3: PairWise Task

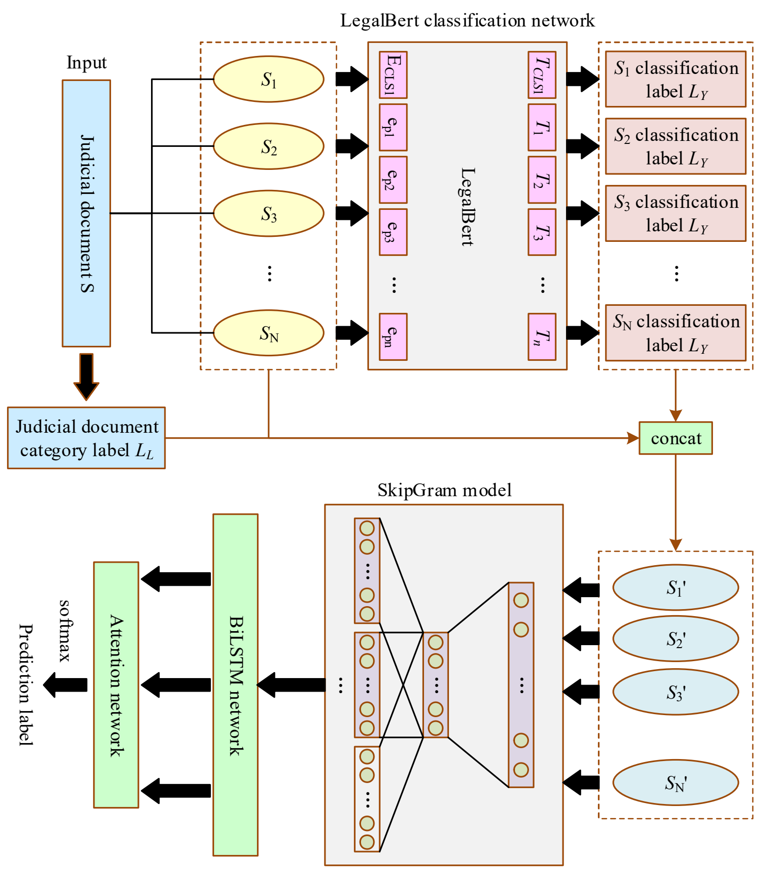


Figure 4: Dispute focus recognition model based on multi-feature fusion SKIPGRAM

III. Controversial Focus Identification Model Based on Multi-Feature Fusion

III. A. Definition of the task of identifying points of contention

Given a judicial document $S=\{S_1,S_2,S_3,\cdots,S_N\}$ composed of several sentences from both the plaintiff and the defendant. Here, N is the number of sentences from both sides. This paper uses a dispute focus identification model to identify the dispute focuses between the plaintiff and defendant in the judicial document, i.e., by constructing the mapping $\{S_1,S_2,S_3,\cdots,S_N\}\rightarrow\{S_{dispute1},S_{dispute2}\}$ mapping relationship, where $\{S_{dispute1},S_{dispute2}\}$ is the set of dispute points corresponding to the plaintiff and defendant, $dispute1\in\{1,2,\cdots,N\}$ and $dispute2\in\{1,2,\cdots,N\}$ with $dispute1\neq dispute2$.

III. B. Proposed method

This paper introduces the overall framework and detailed working process of the Multi-Feature Fusion SKIPGRAM-based Controversial Focus Identification Scheme (MFFS).

III. B. 1) Overall Architecture

The dispute focus identification model based on multi-feature fusion SKIPGRAM is shown in Figure 4. First, the judicial documents $S=\{S_1,S_2,S_3,\cdots,S_N\}$ into sentences based on punctuation marks, yielding the sentence list $[S_1,S_2,S_3,\cdots,S_N]$, where S_i represents the i th sentence in the judicial document. Next, the segmented sentences are input into the LegalBert classification network to obtain the sentence classification categories. The sentence classification categories and the document category are then fused back into the original sentences. Subsequently, the multi-feature fused plaintiff-defendant texts are input into the SkipGram model for training, followed by deep feature extraction using a bidirectional LSTM network and an Attention network. Finally, the softmax function is used to predict the final results, identifying the key points of contention between the plaintiff and defendant in the judicial document.

III. B. 2) Model Details Description

(1) Based on the LegalBert classification network module

Due to the limited features of the plaintiff's text and the defendant's text in judicial documents, there is little correlation between the features of different texts. Based on this, this paper constructs a label classification network based on LegalBert and defines five categories: action, compensation, penalty, appraisal, and other explanations. The plaintiff's text and the defendant's text are then fed into the LegalBert classification network for category prediction.

The five categories—action, compensation, penalty, appraisal, and other explanations—are used to input the plaintiff's text and the defendant's text into the LegalBert classification network for category prediction. The plaintiff and defendant text sequences $\{S_1,S_2,S_3,\cdots,S_N\}$ are sequentially input into the LegalBert classification network. Taking S_i as an example, $S_i=\{L_1,L_2,L_3,\cdots,L_n\}$ After LegalBert encoding, $E_p=\left(e_{p1},e_{p2},e_{p3},\cdots,e_{pn}\right)$ is obtained, with the calculation formula shown in Equation (9), where n is the number of words in the sentence. Next, the softmax classifier is used to predict the label probabilities of E_p, yielding the sentence classification label L_y, with the calculation formula shown in Equation (10).

$$E_p = LegalBert(S_i) \tag{9}$$

$$L_y = soft\max(E_p) \tag{10}$$

When judicial documents are published, they are categorized into corresponding case types, such as administrative cases, assault cases, civil cases, inheritance cases, etc. The case category is used as a feature L_L and concatenated with the sentence classification label L_y at the end of the pleading text to obtain a new judicial document text sequence $S'=\{S_1',S_2',S_3',\cdots,S_T'\}$ is obtained, where $S_i'=\{S_i,L_Y,L_L\}$

Next, we will provide a detailed introduction to the new judicial document $S'=\{S_1',S_2',S_3',\cdots,S_T'\}$ after being represented by the Skip-Gram encoding module, bidirectional LSTM, and Attention network, the final output results are obtained.

(2) Based on the Skip-Gram model encoding module

The Skip-Gram model can vectorize text. After merging the features, the plaintiff's text and the defendant's text are $S'=\{S_1',S_2',S_3',\cdots,S_T'\}$ into the Skip-Gram model, and then extract each sentence $S_{i'}$ from the plaintiff and defendant texts in sequence. Next, use the word vector encoder of the Skip-Gram model layer to convert $S_{i'}=\{L_1,L_2,L_3,\cdots,L_n,L_y,L_L\}$ into encoding layer vectors $E=\{v_{c_1}',v_{c_2}',v_{c_3}',\cdots,v_{c_i}',\cdots,v_{c_T}',v_{L_Y}',v_{L_L}'\}$. Here, v_{c_1}' is the vectorized result of the i th word in S_i'. where $v_{L_y'}$ and $v_{L_L'}$ are the vectorized results of the sentence classification category and the judicial document category, respectively.

In order to better integrate the features of the sentence classification category and the judicial document category, the word vector representation is optimized by recalculating the probabilities of the central word and the context words. In this paper, each word is treated as a central word l. The set of surrounding words c that appear in the context of the central word is defined as D_p, and the set of surrounding words that do not appear in the context of the central word is defined as D_n. Based on the Skip-Gram model's syntax for predicting context words from central words, the objective function L is redefined as shown in Equation (11).

$$L = \sum_{(l,c)\in D_p} \log \frac{1}{1+e^{-v_c' v_l}} + \sum_{(l,c)\in D_n} \log \frac{1}{1+e^{-v_c' v_l}} + \sum \frac{1}{1+e^{-v_{L_Y}' v_l}} + \sum \frac{1}{1+e^{-v_{L_Y}' v_l}} \tag{11}$$

Among them, v_l is the vectorized representation corresponding to the central word, v_c' is the vectorized representation corresponding to the surrounding word c, and v_{L_Y}' is the vectorized representation of the sentence classification category L_y, and v_{L_L}' is the vectorized representation of the case category label L_L.

Traversing each word in sentence S_i' as the central word, obtaining the objective function L corresponding to each word, calculating the gradient of objective function L corresponding to each word, when the gradient of objective function L is largest, the vectorized representation of all surrounding words c in objective function L is represented as the optimal vectorized representation of the i th word L_i in sentence S_i', and the optimal vectorized representation (e_y, e_L) of the vectorized representation v_{L_Y}' and v_{L_L}' in sentence S_i' is used as the sentence classification category L_y and case category label L_L to obtain the sentence S_i' Optimal vector representation $E_i = \left[e_1, e_2, \cdots, e_i, \cdots, e_T, e_y, e_L\right]$.

(3) Based on the BiLSTM-Attention representation module

The optimal vector representations e_i of each word in the vector representation E of each sentence are successively input into the bidirectional LSTM network for training, yielding the forward and backward output vector representations $\vec{h}_i$ and $\overleftarrow{h}_i$, calculated as shown in Equations (12)-(13):

$$\vec{h}_i = \overrightarrow{LSTM}\left(\vec{h}_{i-1}, e_i\right) \tag{12}$$

$$\overleftarrow{h}_i = \overleftarrow{LSTM}\left(\overleftarrow{h}_{i+1}, e_i\right) \tag{13}$$

In this context, $\overrightarrow{LSTM}$ and $\overleftarrow{LSTM}$ denote the forward and backward LSTM neural networks, respectively, and denote the corresponding forward and backward LSTM network word vectorization outputs. $\vec{h}_{i-1}$ is the vectorized representation of the word input to the vectorized representation output $\vec{h}_i$ of the previous time step in the forward network, $\overleftarrow{h}_{i+1}$ is the vectorized representation output of the word at the next time step in the backward $\overleftarrow{LSTM}$ network, which is the vectorized representation input $\overleftarrow{h}_i$ of the word.

Through the forward network, $E_i = \left[e_1, e_2, \cdots, e_i, \cdots, e_T, e_y, e_L\right]$ into $\left[\vec{h}_1, \vec{h}_2, ..., \vec{h}_i, ..., \vec{h}_T, \vec{h}_y, \vec{h}_L\right]$ by the forward network. Convert to $\left[\overleftarrow{h}_1, \overleftarrow{h}_2, \cdots, \overleftarrow{h}_i, \cdots, \overleftarrow{h}_T, \overleftarrow{h}_y, \overleftarrow{h}_L\right]$.

Next, concatenate the vectorized representation of the forward output, the vectorized representation of the reverse output $\overleftarrow{h}_i$, and the optimal vectorized representation e_i as the hidden state representation h_i of the i th word in the sentence. The concatenation formula is shown in Equation (14):

$$h_i = concat\left(\vec{h}_i, \overleftarrow{h}_i, e_i\right) \tag{14}$$

The hidden state representation H of the sentence output by the BiLSTM model layer is obtained by concatenating the h_i sequences, as shown in Equation (15):

$$H = \left[h_1; h_2; \cdots; h_n\right] \tag{15}$$

The hidden state representation H of the sentence is input into the Attention layer, which outputs the attention weight matrix A. The calculation formula is shown in Equation (16):

$$A = soft\max\left(W_{s2} \tanh\left(W_{s1} H^T\right)\right) \tag{16}$$

Multiply the hidden state representation of the sentence by the attention weight matrix to obtain the matrix Z of the sentence, as shown in equation (17):

$$Z = AH \tag{17}$$

(4) Prediction output layer

Each sentence matrix is fed into the fully connected layer as input, and the probability y of each sentence is output. The two sentences with the highest probabilities are identified as the focal points of the dispute between the plaintiff and the defendant. The fully connected layer uses the softmax activation function to calculate the probabilities corresponding to each sentence, with the probability calculation formula shown in Equation (18):

$$y = soft\max\left(WZ + b\right) \tag{18}$$

In particular, W and b are the weight matrix and bias of the fully connected layer, respectively.

III. C. Deliberation and ruling on points of contention

In cases heard by a panel of arbitrators, arbitrators may submit their opinions through the system. During the decision-making process, the system automatically provides case analysis, legal provisions, and similar precedents for arbitrators to reference when rendering their decisions. Additionally, based on the case and its prior documents, the system automatically generates a draft of the ruling document, which includes the document's header, party information, case background, arbitration claims and their facts and reasons, findings of fact, the Commission's opinion, and the closing statement. The system-generated draft is then input into a dispute focus identification model based on multi-feature fusion to identify and output potential dispute focuses in the ruling document. The “The Commission's Findings” section comprises three parts: “Reasoning for the Ruling,” “Legal Basis for the Ruling,” and “Main Text of the Ruling.” This section is the core of the ruling document and the most time-consuming and labor-intensive part for arbitrators. Therefore, the system intelligently generates the “reasoning section” of the ruling document based on the applicant's claims and the handling of the case. It also intelligently generates the “legal basis” and “main text of the ruling” based on the arbitrator's handling of the arbitration claims, combined with the labor and personnel dispute legal knowledge graph. For the “Judgment Disposition” section, arbitrators may make modifications. If a judgment is rendered without precedents, i.e., outside the deviation range, the arbitrator must provide an explanation.

IV. Model Performance Verification and Evaluation

IV. A. Similar Case Matching Model

This section sets up rule matching experiments, model comparison experiments, and model module ablation tests to evaluate the performance of case matching models for collaborative sentence pairing tasks from multiple dimensions.

IV. A. 1) Rule matching experiment

The experiments were conducted using the dataset provided by the 2022 National Legal Knowledge Competition Information Extraction Track, which includes 50 categories of legal elements such as courts, judgments, case numbers, defendant names, plaintiff's statements of facts, and charges, totaling 6,200 data entries. The performance metrics of accuracy, recall, and F1 were tested for the rule matching method proposed in this paper. The experimental results of the rule matching experiment for the model in this paper are shown in Table 1, with the following features: (L1) court, (L2) judgment document, (L3) case number, (L4) client, (L5) procuratorate, (L6) legal provisions, (L7) presiding judge, (L8) judge, (L9) judicial assistant, (L10) trial date, and (L11) court clerk.

Table 1: Rule matching experiment results

Legal elements	Accuracy rate (%)	Racall rate (%)	F1 value (%)
L1	98.29	99.44	98.01
L2	100.00	100.00	100.00
L3	100.00	100.00	100.00
L4	98.11	98.78	99.88
L5	98.47	99.00	99.26
L6	99.81	99.52	98.53
L7	98.99	98.81	98.17
L8	99.93	99.44	98.34
L9	100.00	100.00	100.00
L10	100.00	100.00	100.00
L11	99.43	98.84	99.25
Average values of each item	99.37	99.44	99.22

Table 1 shows that the average accuracy rate for extracting prominent legal elements such as (L1) court, (L2) judgment document, and (L3) case number was 99.37%, the average recall rate was 99.44%, and the average F1 score was 99.22%. Additionally, in (L1) courts, (L2) judgments, (L3) case numbers, (L4) clients, (L5) procuratorates, (L6) legal provisions, (L7) presiding judges, (L8) judges, (L9) judicial assistants, (L10) trial dates, and (L11) court clerks. In the case of the client and legal provisions, the extraction was not fully successful. This is because the rule matching was exact matching, and in a small number of documents, there were instances where these two types of elements were written with either more or fewer words. Such situations affected the accuracy of the legal element rule extraction.

IV. A. 2) Comparison with similar models

Three metrics—P@5, P@10, and MAP—were selected as accuracy evaluation criteria, while NDCG@10, NDCG@20, and NDCG@30 were used as ranking metrics. The model for matching similar cases in this paper was compared with the baseline model on the LcCaRD test set. (BM25, LMIR, BERT-wwmext, OpcnCLaP, RoBERTa, Lcgal-RoBERTa, BM25+BERT-wwmext, BM25+OpcnCLaP, BM25+RoBERTa, BM25+Lcgal-RoBERTa) on the LcCaRD test set are shown in Table 2. As can be seen, compared to other similar baseline models, the proposed model achieved the most superior evaluation performance across six evaluation metrics, with scores of 61.79, 57.29, 67.02, 91.14, 92.95, and 96.62, respectively.

Table 2: The performance of multiple models on the test set

Model	P@5	P@10	MAP	NDCG @10	NDCG @20	NDCG @30
BM25	46.00	48.50	55.61	77.01	80.85	90.21
LMIR	54.50	50.50	57.46	79.38	83.02	91.98
BERT-wWm$_{ext}$	47.87	48.87	61.90	83.51	84.08	87.80
OpcnCLaP	41.96	41.87	54.31	80.27	82.89	85.52
RoBERTa	43.76	41.26	55.37	74.11	77.55	81.03
Lcgal-RoBERTa	45.93	45.76	55.89	75.87	77.75	82.16
BM25+BERT-wWm$_{ext}$	48.12	49.02	61.24	85.62	88.50	94.63
BM25+OpcnCLaP	43.01	42.12	54.17	83.13	88.23	94.18
BM25+RoBERTa	44.45	41.12	56.76	81.59	87.26	93.54
BM25+Lcgal-RoBERTa	45.86	45.88	56.90	83.60	87.36	93.30
Textual	61.79	57.29	67.02	91.14	92.95	96.62

IV. A. 3) Ablation testing

This section explores the impact of different components of the proposed model on similar case retrieval through a series of ablation experiments. In all ablation experiments, the same case dataset, parameter settings, and experimental environment were used. The ablation combinations were set as follows: (G1) complete model, (G2) pooling+attention, (G3) pooling+embedding, and (G4) attention+embedding. The ablation test results on the test set are shown in Figure 5.

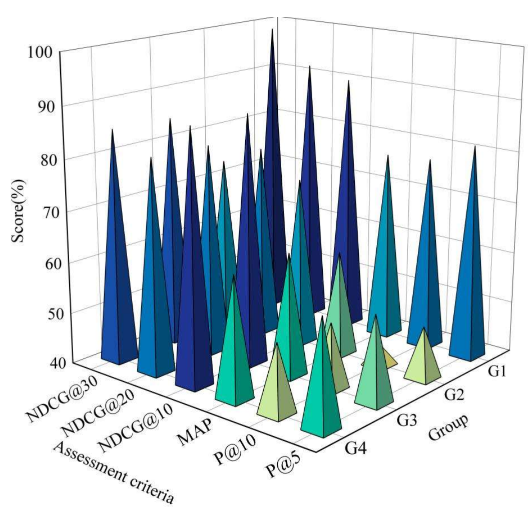


Figure 5: Ablation tests on the test set

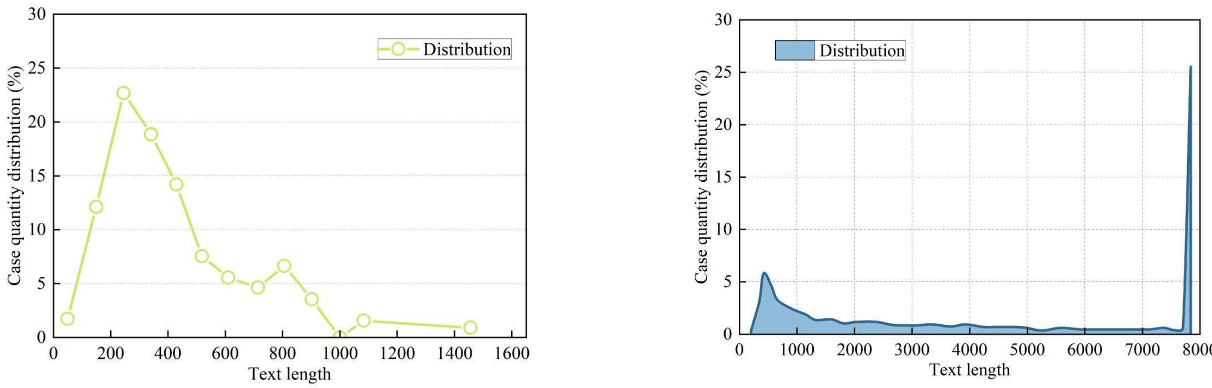
It can be seen that on the whole, only the (G1) complete model has a score of 70.00% or above in each evaluation index. The overall score of the model without the attention interaction layer (G3) is lower than 80.00%, and the indicator P@5 is only 44.087%. This results show that the attention mechanism is very effective in the case matching model, which promotes the information interaction between the two text representations output by the twin network, and enhances the text representation ability and retrieval effect.

IV. B. Dispute Focus Identification Model

In testing the performance of the dispute focus identification model, the selected dataset and evaluation metrics are consistent with those of the case matching model. The selected baseline models for comparison are: TF-IDF, BM25, SBERTq512, OpcnCLaP, Lawformcr, BERT-PLI, and BERT-LBIAq100.

IV. B. 1) Data Set Analysis

Based on the dispute focus identification model proposed in this paper, Figure 6(a) shows the distribution of text lengths in the query cases, and Figure 6(b) shows the distribution of text lengths in the candidate cases. In the query cases, the text lengths are mainly concentrated in the range of 0-600, accounting for about 60% of the total. In the candidate cases, the text lengths are mainly concentrated around 8000.



(a) Query the percentage distribution of the text length of the case

(b) The percentage distribution of text length of the candidate cases

Figure 6: The percentage distribution of text length in the LcCaRD test set

IV. B. 2) Comparison of recognition performance

Table 3 shows the comparison of the performance evaluation of the controversy focus identification model proposed in this paper with 7 similar baseline models, and the overall evaluation scores of the 8 models in index NDCG@10 and NDCG@20 are 75.00 or above, and the score of the model in this paper is as high as 88.62 in the NDCG@20 indicators. In terms of the performance of other indicators, although the overall concentration is between 40.00~70.00, the evaluation score of the model in this paper is still better than that of similar models, and the scores on P@5, P@10, MAP and NDCG@30 indicators are 63.26, 52.74, 67.65 and 63.26 respectively.

Table 3: Performance evaluation and comparison of the model on the LcCaRD dataset

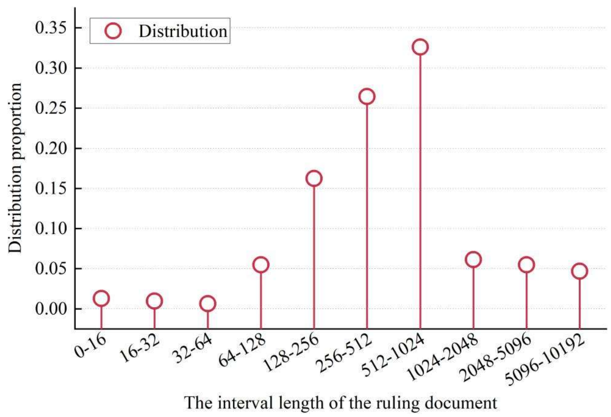
Model	P@5	P@10	MAP	NDCG @10	NDCG @20	NDCG @30
TF-IDF	40.23	40.23	48.76	75.68	80.43	40.23
BM25	43.29	48.75	52.39	76.78	81.72	43.29
$SBERT_{q512}$	53.12	47.04	56.64	79.66	80.13	53.12
OpcnCLaP	48.03	43.67	54.33	81.38	83.97	48.03
Lawformcr	53.27	49.61	62.02	83.95	85.23	53.27
BERT-PLI	45.63	43.02	49.21	75.07	80.07	45.63
$BERT\text{-}LBIA_{q100}$	52.41	46.15	62.84	79.82	83.03	52.41
Textual	63.26	52.74	67.65	86.45	88.62	63.26

V. Establishment and testing of a labor dispute resolution system and mechanism

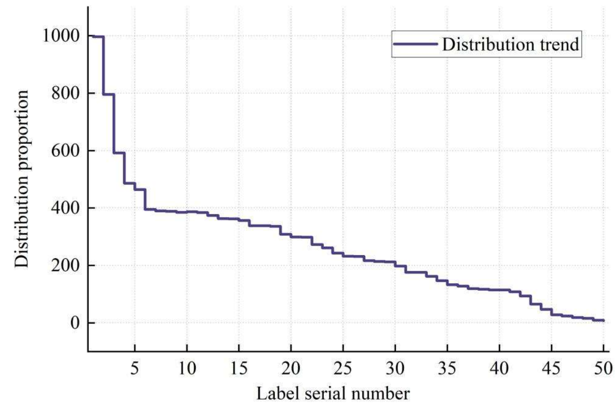
To obtain legal basis for labor disputes, a legal data statistical analysis module was designed in this chapter. Based on the case matching model for the collaborative sentence ranking task, a case retrieval module was designed. Based on the dispute focus identification model with multi-feature fusion, a dispute identification module was established. The three modules were integrated to form a labor dispute resolution system.

V. A. Legal Data Statistical Analysis Module

Using the legal data statistical analysis module proposed in this paper, we analyzed the characteristics of the dataset provided by the 2022 National Legal Knowledge Competition Information Extraction Track. Among these, the dispute focus labels were selected from the 50 most frequently occurring keywords in current labor dispute practices. The distribution of the length of judicial documents is shown in Figure 7(a), and the trend in the distribution of dispute focus labels within the documents is shown in Figure 7(b). The length of judicial documents primarily ranges between 256 and 1024, accounting for over 50.00% of the total. The top five keywords in dispute focus labels—labor contracts, labor remuneration, working hours, social insurance, and non-compete clauses—are most frequently distributed in judicial documents, with each exceeding 400 occurrences.



(a) The distribution of interval lengths in judicial documents

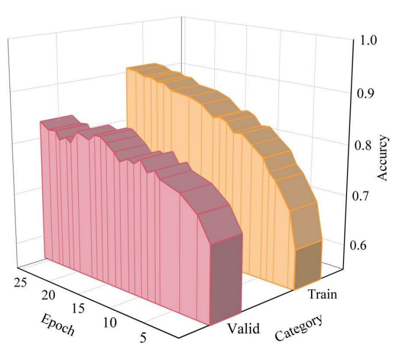


(b) Distribution trend of controversial labels

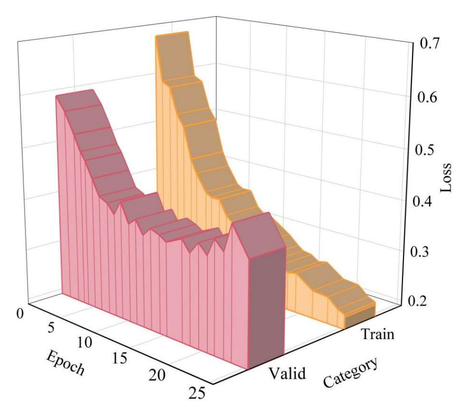
Figure 7: The application performance of the statistical analysis module

V. B. System Model Training and Performance Analysis

The model for the labor dispute resolution system described in this paper was trained using the training set and validation set. The accuracy and loss rates of the model during the training process are shown in Figures 8(a) and 8(b), respectively. Epoch represents the number of training rounds, Accuracy represents the model accuracy, and Loss represents the model loss rate. As can be seen, on the validation set, as the number of training rounds gradually increases, the model's accuracy ultimately reaches 83.01%, with a loss rate as low as 38.70%. On the training set, after 25 training rounds, the model's accuracy ultimately reaches 90.69%, with a loss rate of only 21.40%.



(a) Accuracy performance



(b) Loss performance

Figure 8: Accuracy and loss in the model training process

Six similar models—CNN, BiLSTM, BiGRU, BLA, BGA, and CBLA—were selected as comparison models. Table 4 shows a comparison of the final accuracy and loss rates of these models with those of the labor dispute resolution system model described in this paper. Among the seven models, the model described in this paper achieved the highest accuracy of 0.8579 and the lowest loss rate of 0.3964, verifying the reliability of the proposed model in assisting with labor dispute resolution.

Table 4: The comparison results with similar models

Model	Accuracy rate	Loss rate
CNN	0.8640	0.4526
BiLSTM	0.8288	0.4594
BiGRU	0.8345	0.4537
BLA	0.8517	0.4304
BGA	0.8625	0.4519
CBLA	0.8517	0.4552
Textual	0.8579	0.3964

VI. Conclusion

In terms of the reference of similar cases in labor dispute resolution, this paper adds the attention mechanism module to the expression matching model, and synthesizes contrastive learning and training ideas based on PairWise to form a collaborative PairWise multi-task training method, and builds a similar case matching model for collaborative sentence-to-order sorting tasks. The average accuracy of the model is as high as 99.37% for the obvious legal elements in terms of rule matching. In the six indicators of P@5 (61.79), P@10 (57.29), MAP (67.02), NDCG@10 (91.14), NDCG@20 (92.95) and NDCG@30 (96.62), the performance of similar models far exceeded that of similar models.

For the identification and adjudication of labor disputes, this paper proposes a multi-feature fusion dispute focus recognition model based on LegalBert classification network module, Skip-Gram model, BiLSTM-Attition representation module and prediction output layer module. The model achieves the best score of 88.62 in the NDCG@20 index, and has superior recognition performance.

Based on the proposed model and the statistical analysis module of legal data, a labor dispute resolution system mechanism is established. After 25 rounds of training, the accuracy of the system model reached 90.69% while the loss was only 21.40%, which has high feasibility and reliability, and is an effective attempt to apply artificial intelligence to labor dispute resolution mechanism.

References

[1] Zhou, G., Chu, G., Li, L., & Meng, L. (2020). The effect of artificial intelligence on China’s labor market. China Economic Journal, 13(1), 24-41.

[2] Mfanafuthi, M., Nyawo, J., & Mashau, P. (2019). Analysis of the impact of artificial intelligence and robotics on human labour. Gender and Behaviour, 17(3), 13877-13891.

[3] Webb, M. (2019). The impact of artificial intelligence on the labor market. Available at SSRN 3482150.

[4] Mukherjee, A. N. (2022). Application of artificial intelligence: benefits and limitations for human potential and labor-intensive economy–an empirical investigation into pandemic ridden Indian industry. Management Matters, 19(2), 149-166.

[5] Huo, Q., Ruan, J., & Cui, Y. (2024). “Machine replacement” or “job creation”: How does artificial intelligence impact employment patterns in China's manufacturing industry?. Frontiers in Artificial Intelligence, 7, 1337264.

[6] Wang, X., Chen, M., & Chen, N. (2024). How artificial intelligence affects the labour force employment structure from the perspective of industrial structure optimisation. Heliyon, 10(5).

[7] Su, Z., Togay, G., & Côté, A. M. (2021). Artificial intelligence: a destructive and yet creative force in the skilled labour market. Human Resource Development International, 24(3), 341-352.

[8] Carneiro, D., Novais, P., Andrade, F., Zeleznikow, J., & Neves, J. (2014). Online dispute resolution: an artificial intelligence perspective. Artificial Intelligence Review, 41, 211-240.

[9] Hampole, M., Papanikolaou, D., Schmidt, L. D., & Seegmiller, B. (2025). Artificial intelligence and the labor market (No. w33509). National Bureau of Economic Research.

[10] Zarifhonarvar, A. (2024). Economics of chatgpt: A labor market view on the occupational impact of artificial intelligence. Journal of Electronic Business & Digital Economics, 3(2), 100-116.

[11] Damioli, G., Van Roy, V., & Vertesy, D. (2021). The impact of artificial intelligence on labor productivity. Eurasian Business Review, 11, 1-25.

[12] Zhou, H., Wang, L., Cao, Y., & Li, J. (2025). The impact of artificial intelligence on labor market: A study based on bibliometric analysis. Journal of Asian Economics, 101926.

[13] Carbonero, F., Davies, J., Ernst, E., Fossen, F. M., Samaan, D., & Sorgner, A. (2023). The impact of artificial intelligence on labor markets in developing countries: a new method with an illustration for Lao PDR and urban Viet Nam. Journal of evolutionary economics, 33(3), 707-736.

[14] Ermakova, E. P., & Frolova, E. E. (2021). Using artificial intelligence in dispute resolution. In Smart technologies for the digitisation of industry: Entrepreneurial environment (pp. 131-142). Singapore: Springer Singapore.

[15] Miyagusuku, J. T. (2024). La inteligencia artificial y las relaciones laborales. IUS ET VERITAS, (69), 55-69.

[16] Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. Journal of Labor Economics, 40(S1), S293-S340.

[17] Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., ... & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. Proceedings of the National Academy of Sciences, 116(14), 6531-6539.

[18] Guan, J., Yu, Z., Liao, Y., Tang, R., Duan, M., & Han, G. (2024). Predicting Critical Path of Labor Dispute Resolution in Legal Domain by Machine Learning Models Based on SHapley Additive exPlanations and Soft Voting Strategy. Mathematics, 12(2), 272.

[19] Zeleznikow, J. (2021). Using artificial intelligence to provide intelligent dispute resolution support. Group Decision and Negotiation, 30(4), 789-812.