

## *Retraction*

# **Retracted: Deep Learning for Decision Element Extraction in Fact Description of Legal Documents**

### **Advances in Multimedia**

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

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

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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

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

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

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

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

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

### **References**

- [1] X. Li, "Deep Learning for Decision Element Extraction in Fact Description of Legal Documents," *Advances in Multimedia*, vol. 2022, Article ID 6259071, 11 pages, 2022.

## Research Article

# Deep Learning for Decision Element Extraction in Fact Description of Legal Documents

Xia Li 

*Henan Institute of Economics and Trade, Zhengzhou 450018, China*

Correspondence should be addressed to Xia Li; [hnjmlxia@henetc.edu.cn](mailto:hnjmlxia@henetc.edu.cn)

Received 27 May 2022; Revised 9 July 2022; Accepted 15 July 2022; Published 9 September 2022

Academic Editor: Qiangyi Li

Copyright © 2022 Xia Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the work efficiency of judicial personnel and solve the problem of waste of judicial resources, this topic proposes a method of decision element extraction in the fact description of legal documents based on in-depth learning. Firstly, this paper briefly introduces the basic theory of deep learning, text mining technology, a neural network, and other theories and technologies, then expounds the decision element extraction model in the fact description of legal documents based on deep learning, such as the HMM model, the CRF model, and the Bert model, and finally expounds the establishment and implementation method of the decision element extraction model in the fact description of legal documents, so as to provide a guarantee for the work quality and efficiency of judicial personnel. The samples are according to the sample label frequency to obtain more balanced data, and we manually label keywords to obtain feature vectors to assist the model in improving the prediction results, but it also increases the statistical quantity mode of label co-occurrence. Although all modes can be included by using a larger matrix, the amount of calculation increases significantly. Therefore, the follow-up work mainly studies the important co-occurrence features that can be used and then adopts better dimensionality reduction methods to improve the final prediction results.

## 1. Introduction

Legal documents are mainly composed of case types, fact descriptions, and judgment results. A fact description is the core of a judicial case, which contains relatively key information such as the relationship between the plaintiff and the defendant, the cause and process of the matter, the degree of casualties involved in the case, and the number of losses. This information is an important basis for judgment prediction and is generally considered to be the judgment element of the case. The three element sentences in the fact description correspond to the label of having children after marriage (Dv1), the label of failing to fulfill family obligations (dv6), and the label of separation after marriage (dv13) in the judgment result, respectively. The three elements are the important bases of case judgment. Therefore, taking the fact description part of the case as the original data, it is of great significance to study the extraction of fact description elements. With the development of information technology

and the Internet in China, public and production departments have produced various forms of information documents, such as web pages and office documents. In particular, the information on web pages increases by tens of thousands every day [2]. Generally speaking, electronic information is provided to users in the form of text or natural language. How natural language information can be quickly adopted by information users after filtering, classification, and other processing, as shown in Figure 1. This problem has become increasingly urgent and must be solved. Research on information extraction for processing natural language is an important technology to meet the effective use of text information. With the construction of China's society ruled by the law and the enhancement of citizens' legal awareness, people resolve disputes through court proceedings to protect their legitimate rights and interests and safeguard social fairness and justice. Litigation cases have increasingly become the focus of social attention [3]. However, the amount of information of litigation cases is

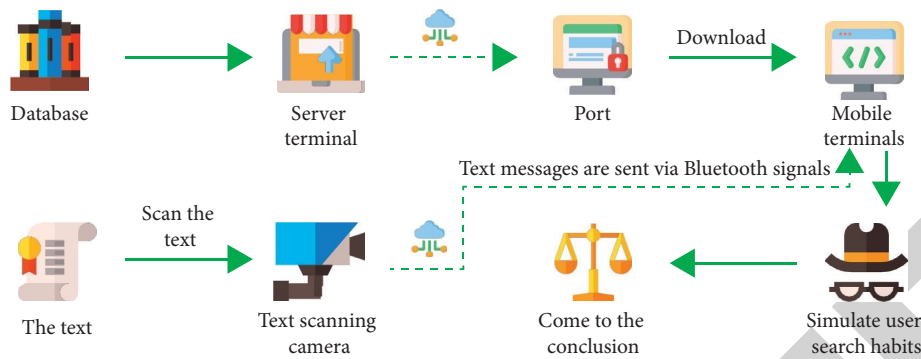


FIGURE 1: Legal document information extraction.

increasing exponentially. The relevant personnel in case judgment need to spend a lot of time reading case records and analyzing relevant historical cases, which makes them overburdened and inefficient. In addition, when facing different cases, judicial personnel rely on their accumulated experience and knowledge over the years to make a fair and reasonable judgment on the case, and this experience is inevitably subjective. In the face of a complex social environment, judicial personnel supplement the abstraction, lag, and uncertainty of legal provisions and also have a certain impact on the judgment results of cases [4]. These have further become obstacles to the construction and development of China's judicial field. Therefore, with the help of artificial intelligence technology, the development of an automatic auxiliary extraction system can greatly bring convenience to judicial personnel and provide solutions to the problems caused by the above objective or subjective reasons. In this paper, the decision element sentence extraction task is formalized into a multilabel classification model for fact description sentences. The main difficulties include the following: (1) The decision element classification task is a classic one-to-one or one-to-many problem. It is difficult to determine that each fact description belongs to one or more element labels by traditional methods. (2) The length of fact description sentences is uneven. The length of the shortest element sentence is 30–40 words, and the length of the longest element sentence will reach more than 300 words. The traditional model mostly uses fixed parameters as the vector dimension and fills the vector with 0 for the short sentence, which cannot effectively capture the characteristic representation of sentences with different lengths [5]. In order to weaken the negative impact of the length difference of different sentences on the effect of the model, the multihead self-attention mechanism (MAT) based on the mask method is further integrated into the BERT-CNN model.

On the basis of this research, this paper proposes a deep learning-based decision element extraction from the fact description of legal documents. First of all, a brief introduction to the basic theory of deep learning, text mining technology, the neural network, and other theories and technologies is introduced, and then the HMM model, the CRF model, the BERT model, and other deep learning-based judgment element extraction models in the description of legal documents are described. Finally, the facts of legal documents are

described. The establishment and implementation method of the judgment element extraction model in the description provides a guarantee of the quality and efficiency of judicial personnel. For different types of cases, the elements are different, and it is time-consuming and laborious to establish the element system for various cases in the judicial field. Therefore, it is our future research direction to study open element extraction in the judicial field based on unsupervised or semi-supervised data. In judicial field case judgment prediction, judicial case corpus often has a high imbalance problem and belongs to the task of multilabel classification.

## 2. Introduction to Relevant Theories and Technologies of In-Depth Learning

*2.1. Basic Theory of Deep Learning.* Using computers to process natural language is the process of converting textual information into machine language expressions. A complete natural language is composed of words in the case document. The words in the natural language are expressed as machine language through numerical representation and become word vectors [6]. The word vector was first expressed in a discrete way. With the progress of science and technology and the continuous research of scientific researchers and technicians, a distributed representation is proposed, which can better represent word vectors to solve semantic problems. According to the learning of word vectors, it can be known that the discrete representation of word vectors is the most basic representation method. The basic concept of discrete representation is to use high-dimensional sparse vectors to complete the representation of word vectors. Through the analysis of the dimension of the word vector, 1 or 0 is used to represent the position of any word. Generally, the current position is recorded as 1, otherwise, it is marked as 0 [7]. For example, there are three texts in a corpus as follows:

- (i) Text 1: I/like/blue sky
- (ii) Text 2: I/like/the sea
- (iii) Text 3: I/enjoy/delicious food

From the above three texts, we can see that this corpus is composed of six words. They are I, like, blue sky, sea, enjoy, and delicious food. Obviously, the dimension of the word

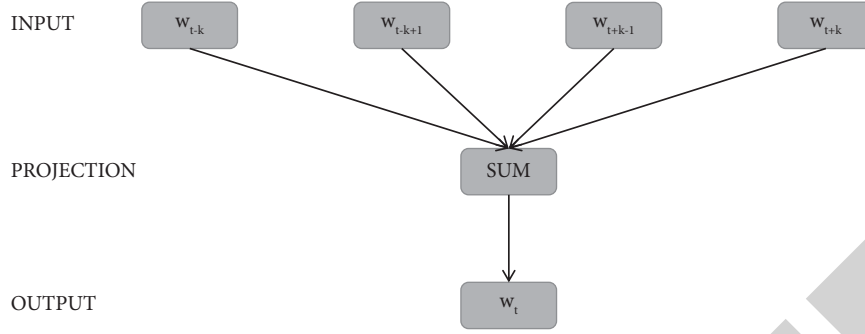


FIGURE 2: TheCBOW model structure.

vector is six. From the position of words in the corpus, discrete representation includes {"I" : [ 1 0 0 0 0 0 ]}, {"like" : [ 0 1 0 0 0 0 ]}, {"blue sky" : [ 0 0 1 0 0 0 ]}, {"sea" : [ 0 0 0 1 0 0 ]}, {"enjoy" : [ 0 0 0 0 1 0 ]}, and, {"delicious food" : [ 0 0 0 0 0 1 ]}. We can see that with the change in the corpus size, the vocabulary is also changing. The dimension of thesaurus will increase with the increase in the corpus, and the vector representation dimension of each word is also increasing. There is only one 1 in the discrete vector representation of each word, and the rest are 0. The number of 0 increases with the increase in the corpus [8].

With the development of deep learning, a more advanced representation method, distributed word vector representation has evolved. In the distributed word vector representation, the processing of a large number of text data is completed through the continuous training of the database. The trained model is easier to find the relevance of the context than the discrete representation [9]. The core idea of distributed representation is to grasp the correlation between vocabulary and context. The discrete vector in the word vector is processed by calculation, which is embedded in the lower dimensional space and expressed as a short vector with fixed value. The CBOW model predicts the possibility of the current word by using the words in the context, as shown in

$$P(w_t | \text{SUM}(w_{t-k}, w_{t-k+1}, \dots, w_{t+k-1}, w_{t+k})). \quad (1)$$

Under the condition that the current word  $w_t$  is known, the possibility of the current word  $w_t$  is predicted according to the context  $2k$  words  $w_{t-k}, w_{t-k+1}, w_{t+k-1}, w_{t+k}$  in the text. Calculate the probability that sentence  $s$  is a natural language, as shown in

$$P(S) = P(w_1, w_2, \dots, w_T) = \prod_{i=1}^T (P(w_i | \text{SUM}(w_{i-k}, w_{i-k+1}, \dots, w_{i+k-1}, w_{i+k}))). \quad (2)$$

$T$  represents the length of the text sequence, and  $P(s)$  represents the possibility of words occurring in this order ( $w_1, w_2, \dots, w_T$ ). Then, the possibility of joint occurrence of each word in the text sequence is  $P(w_1, w_2, \dots, w_T)$  [10]. The likelihood function of the case sequence is shown in

$$L(\theta) = \prod_{i=1}^s P(S_i, \theta). \quad (3)$$

It is further deduced as

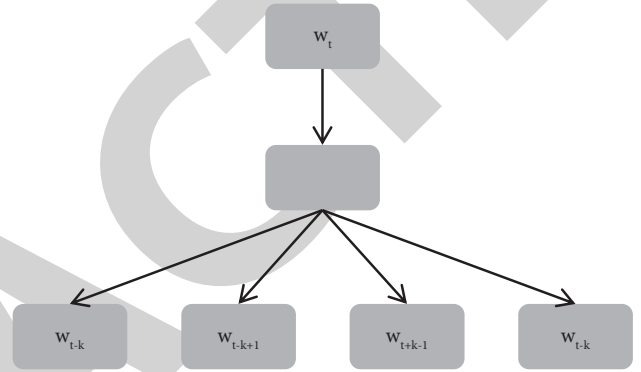


FIGURE 3: The skip Gram model structure.

$$L(\theta) = \frac{1}{T} \sum_{i=1}^S \sum_{t=1}^T \log P(w_t | \text{SUM}(w_{t-k}, w_{t-k+1}, \dots, w_{t+k-1}, w_{t+k}), \theta). \quad (4)$$

The structure of the CBOW model is shown in Figure 2.

The basic principle of the skip Gram model is just the opposite of the CBOW model. The model structure is shown in Figure 3.

In the text sequence, the possibility of the current word  $w_t$  is predicted by the context of  $2k$  words  $w_{t-k}, w_{t-k+1}, w_{t+k-1}, w_{t+k}$  [11].

**2.2. Text Mining Technology.** Text mining is also called text data mining. In recent years, with the rapid development of computer technology, text mining technology has made great development and has become a mainstream method in text processing. Text mining technology is essentially a text analysis technology, which produces high-quality text information in the process of text analysis and processing. High-quality text information is generated through text classification or text prediction, which takes the text as input data, identifies its features, generates structured data, and produces final evaluation and output [12]. According to the research, the generalized text mining operation process generally includes four aspects: text preprocessing task, core mining operation, presentation layer components, visualization function, and refinement operation processing. The text mining operation process is shown in Figure 4.

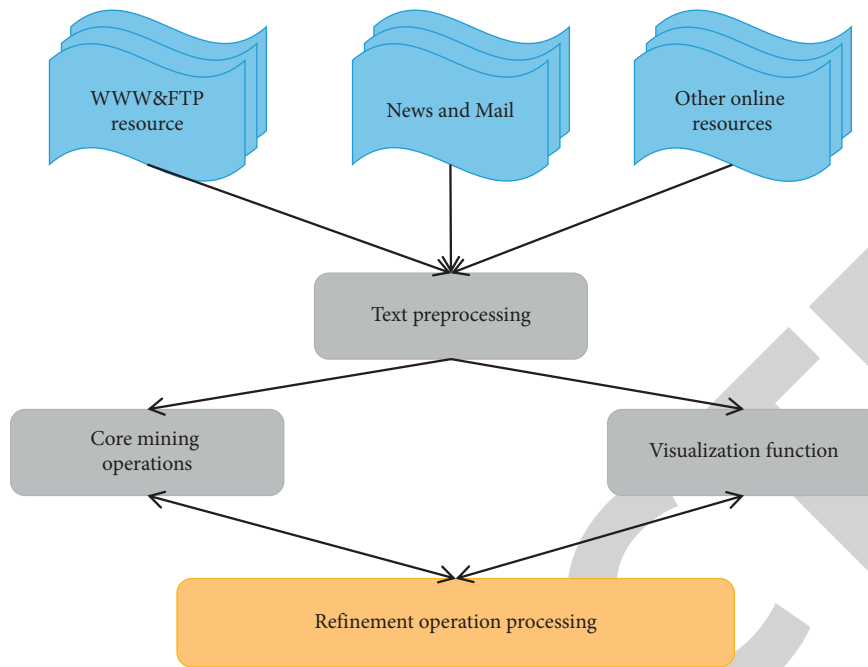


FIGURE 4: Operation flow of text mining.

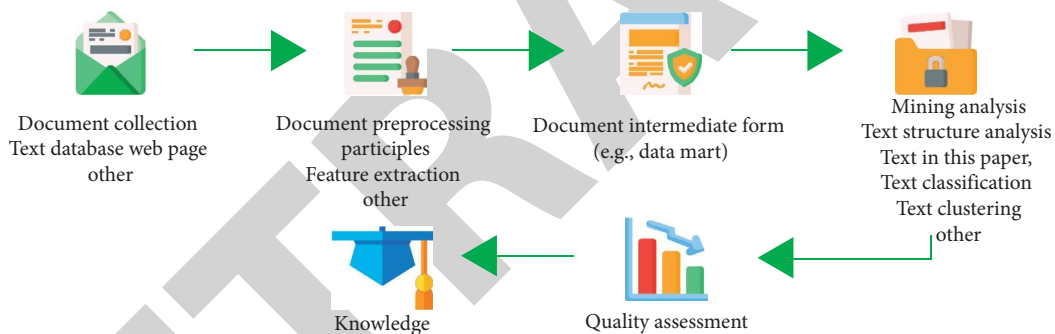


FIGURE 5: Typical text mining method flow.

According to the general text mining process, scholars summarize the typical text mining method process, which mainly includes three stages: preprocessing, text mining, and quality evaluation. Among them, Chinese text preprocessing often includes two stages: word segmentation and text feature representation. The technologies usually involved in text mining and analysis include text structure analysis, text classification, and text summarization. Quality assessment involves text association, distribution analysis, and trend prediction [13]. The typical text mining method is shown in Figure 5.

The types of civil case data elements of marriage and family, labor dispute, and loan contract are shown in Table 1.

Due to the large amount of case text data in the judicial field and most of them being presented in the form of standardized text, we need to pay attention to the relevant methods and technologies of text mining for the knowledge discovery of cases in the judicial field. In this paper, the tasks and applications of judicial text mining are mainly divided into three aspects: named entity recognition, element extraction, and

decision prediction in judicial text data [14]. For a named entity recognition task, it is a sequence annotation process. Feature extraction needs to be realized by text classification technology. The decision prediction task is oriented to extremely unbalanced data sets and belongs to a multilabel prediction task. Sampling technology is needed to alleviate the data imbalance. Sequence tagging is an important technology in natural language processing, which can solve the problem of classifying characters. However, different from classification, in addition to tagging characters, it also takes into account the relationship between output labels and labels. Through sequence tagging, tasks such as named entity recognition and relationship extraction can be realized. Its essence is to input a sequence and output a sequence. More commonly used models include the hidden Markov model (HMM) and conditional random field (CRF) [15].

**2.3. The Neural Network.** Neural networks are divided into feedforward neural networks, convolutional neural



TABLE 1: Types of data elements of civil cases of marriage and family, labor dispute, and loan contract.

Marriage and family	Labor dispute	Loan contract
Having children after marriage	Termination of labor relations	Assignment of creditor's rights
Child support with limited capacity	Pay wages	Loan amount $\times$ 10000 yuan
Joint property of husband and wife	Payment of economic compensation	Loan certificate
Payment of alimony	Failure to pay full labor remuneration	The lender is a financial institution
Real estate division	Existence of labor relations	Repayment of loan
Separation after marriage	No labor contract signed	Other organization funds of the company
Second divorce	Sign labor contract	Joint and several guarantee
Monthly alimony payment	Pay overtime pay	Demand repayment
Divorce granted	There is a mediation agreement	Interest payment
There is joint debt between husband and wife	Pay twice the labor contract not signed	Conclusion of guarantee contract
Personal property before marriage	Salary compensation	There is a written repayment commitment
Legal divorce	Pay industrial injury compensation	Invalidity of guarantee contract   cancellation
Failure to fulfil family obligations	Not brought at the labor arbitration stage	Refusal to perform repayment
There are illegitimate children	Illegal termination of labor relations without payment	Release the guarantor from warranty liability
Appropriate help	Damages	The guarantor is not liable for warranty
Nonperformance of divorce agreement	Economic layoffs	The pledgor is a company
Damages	No bonus payment	The lender fails to comply with the agreed date
Two years of estrangement	Illegal collection of property from workers	Loan amount
Children live with noncustodians	The employer shall notify the termination in advance	Assignment of debt by the debtor
Personal property after marriage	The legal personality has been lost	The agreed interest rate is unknown

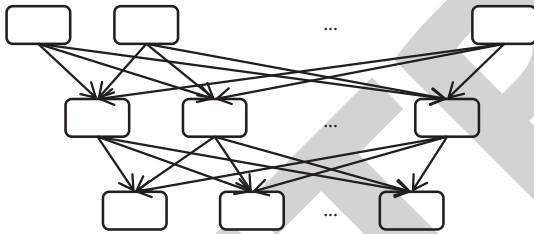


FIGURE 6: The feedforward neural network.

networks, and cyclic neural networks. The feedforward neural network is a network structure composed of multi-layer neurons. In practical applications, too many hidden layer nodes will consume a lot of computing power and time. Due to its relatively simple structure, it also has some shortcomings in feature extraction. The specific structure is shown in Figure 6.

The convolutional neural network is a deep learning model established with reference to human brain visual organization. The CNN generally has the following structures: input layer, convolution layer, excitation layer, pool layer, and full connection layer [16]. Firstly, the data passes through the input layer and then enters the convolution layer from the input layer. During convolution calculation, a matrix of fixed size, called the convolution kernel, is defined. The size of the matrix is the receptive field. Then, during convolution operation, the convolution kernel moves on the input matrix for point product operation to obtain the output result of the convolution layer. After the data passes through the convolution layer, it will enter the pooling layer. The function of the pooling layer is to reduce the dimension

of the vector and reduce the parameters and calculations in the neural network, which can prevent overfitting.

The cyclic neural network is mainly used to process text data of time series. In the research field of natural language processing, such as Chinese word segmentation, text classification, and reading comprehension, recurrent neural networks are often used [17]. Their characteristic is that the input text is closely related to the time series, which is also a symbolic difference from the feedforward neural network. The structure of the recurrent neural network is shown in Figure 7.

The model parameters to be updated in the network structure include  $U$ ,  $W$ , and  $V$ . From the expansion diagram, we can intuitively see that there are two kinds of input information for the text sequence at time  $t$ . Where  $tx$  is the input information of the text sequence at the current time,  $ST-1$  is the output information of the previous time, and the specific calculation of RNN is shown in

$$\begin{aligned} st &= f(Uxt + Wst - 1), \\ ot &= \text{softmax}(Vst). \end{aligned} \quad (5)$$

### 3. The Decision Element Extraction Model in Fact Description of Legal Documents Based on Deep Learning

3.1. *The HMM Model.* HMM is a probability generation model, which describes the process of generating state sequence from the hidden Markov chain and generating observation sequence from state sequence through a random process [18], as shown in Figure 8.

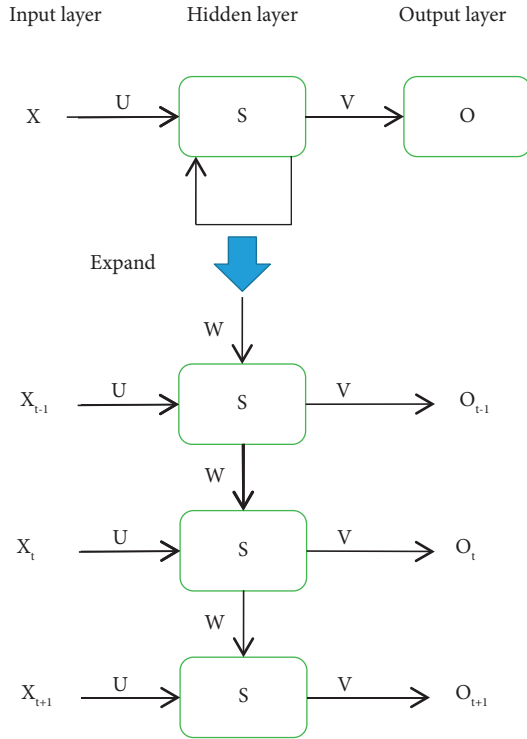


FIGURE 7: Structure of the cyclic neural network.

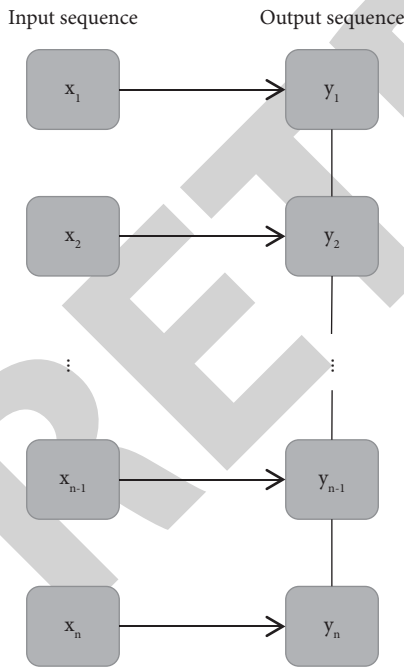


FIGURE 8: The HMM model.

The hidden Markov model (HMM) is a classical machine learning model, which can be used to solve the problem of sequence annotation. The input is the observation variable  $x = \{x_1, x_2, \dots, x_n\}$ , and the output is the corresponding state variable  $y = \{y_1, y_2, \dots, y_n\}$ . Suppose, at any time  $t$ , the observed value  $x_t$  is related to the current state value  $y_t$ , then at

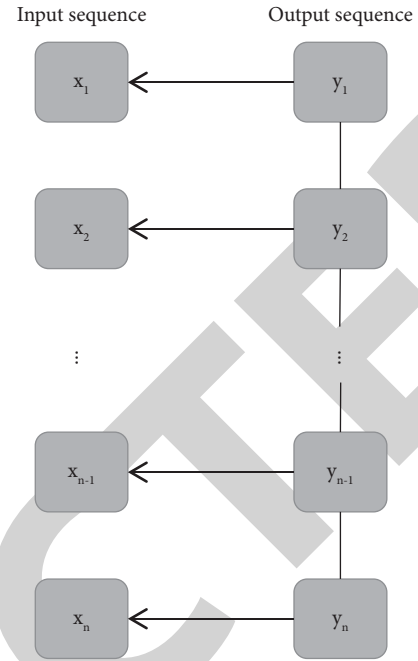


FIGURE 9: The CRF model.

any time  $t$ , the state  $y_t$  is only related to the previous state  $y_{t-1}$ , then the state transition matrix  $A = [a_{ij}]_{N \times N}$  represents the probability of state  $S_i$  transferring to state  $S_j$  at time  $t$ , the observation probability matrix  $B = [b_{ij}]_{N \times M}$  represents the large probability of generating observation  $O_j$  under state  $S_i$ , and the initial state probability distribution  $\Pi = (\Pi_1, \Pi_2, \dots, \Pi_n)$ , represents the probability of each state at the initial time, as shown in

$$\begin{aligned} a_{ij} &= P(y_{t+1} = s_j | y_t = s_i, 1 \leq i, j \leq N), \\ B_{ij} &= P(y_{t+1} = s_j | y_t = s_i, 1 \leq i \leq N, 1 \leq j \leq M), \\ \pi_i &= P(y_i = s_i), 1 \leq i \leq N. \end{aligned} \quad (6)$$

**3.2. The CRF Model.** CRF is a typical representative of statistical models. The linear chain conditional random field model is the most basic one of conditional random field topology, in which  $X$  and  $y$  have the same structure, which is often used in sequence annotation modeling. The linear chain conditional random field model is shown in Figure 9.

Let the input sequence be the case text  $x = \{x_1, x_2, \dots, x_n\}$ , and the output sequence  $y = \{y_1, y_2, \dots, y_n\}$ . The conditional probability calculation method of the output sequence  $y$  is as follows:

$$L_\lambda = \sum_{i=1}^M \log(P^\wedge(Y_i | X_i)) - \sum_k \frac{\lambda_k^2}{2\sigma_k^2}, \quad (7)$$

where  $\sum_k \lambda_k^2 / 2\sigma_k^2$  is the Gaussian a priori value and  $\sigma^2$  is the a priori variance.

The extraction process of key elements of the CRF model cases is shown in Figure 10.

If the unary feature combination is  $\%x [1, 1]$ , it is described as the part of speech feature of the next word. The

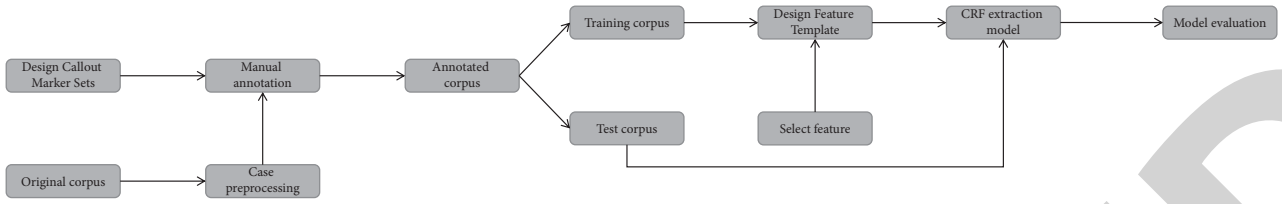


FIGURE 10: CRF model case key element extraction process.

```

#Unigram
U01:%x [-3, 0] U02:%x [-2, 0] U03:%x [-1, 0] U04:%x [-0, 0] U05:%x [1, 0] U06:%x [2, 0]
U07:%x [3, 0] U08:%x [-3, 0]/%x [-2, 0] U09:%x [-2, 0]/%x [-1, 0] U10:%x [-1, 0]/%x [0, 0]
U11:%x [0, 0]/%x [1, 0] U11:%x [0, 0]/%x [1, 0] U12:%x [1, 0]/%x [2, 0] U13:%x [2, 0]/%x [3, 0]
U14:%x [-3, 1] ..... U20:%x [3, 1] U21:%x [-3, 1]/%x [-2, 1] ..... U26:%x [2, 1]/%x [3, 1]

#Bigram
B
    
```

FIGURE 11: Feature template details.

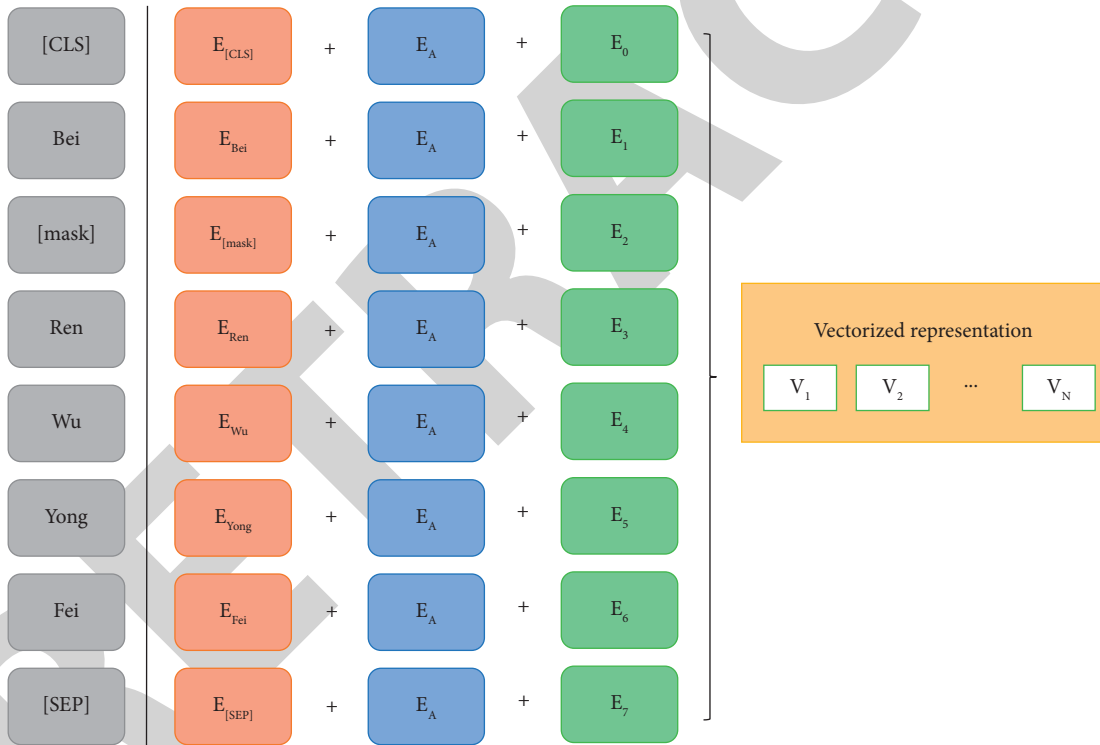


FIGURE 12: Input vectorization of the Bert model.

binary feature combination is  $\%x[0,0]/\%x[0,1]$ , which is described as the current word feature and other word part of speech features. The ternary feature combination is  $\%x[0,0]/\%x[0,1]/\%x[0,2]$ , which is described as the current word feature, other word part of speech features, and other word naming entity features. The details of the feature template are shown in Figure 11.

embeddings. The three embeddings corresponding to the word are superimposed to form the input of Bert. Location information obtained by model learning; [CLS] and [SEP] are the start flag and the end flag, respectively [19]. The input vectorization representation of the BERT model is shown in Figure 12.

The calculation formula of vectorization of text after superposition to express  $V_i$  is shown in

3.3. *The BERT Model.* Input the marked case text into the model. Each word in the text has three embeddings: token embeddings, segment embeddings, and positional

$$V_i = TiWt + SiWs + PiWp, \tag{8}$$



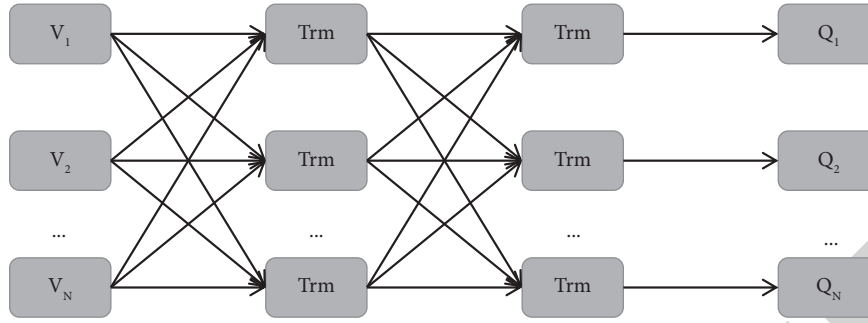


FIGURE 13: Principle of the Bert feature extraction.

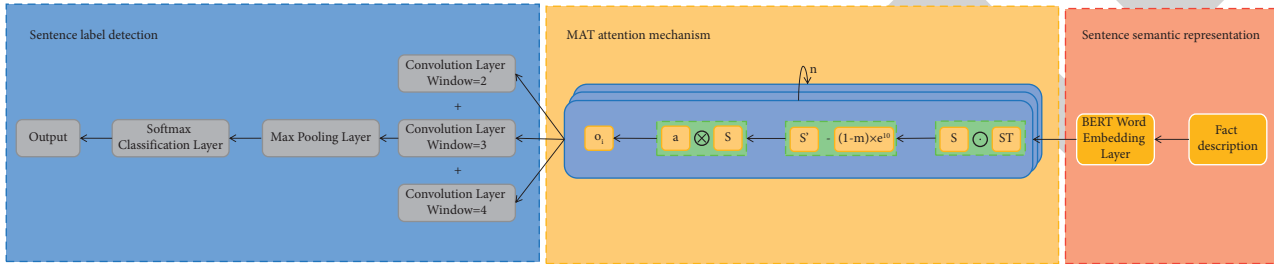


FIGURE 14: Structure of the sentence label prediction model.

where  $T_i$ ,  $S_i$ , and  $P_i$  are word vector, segmentation vector, and position vector, respectively, and  $W_t$ ,  $W_s$ , and  $W_p$  are adjustable parameters. The principle of the Bert feature extraction is shown in Figure 13.

#### 4. Establishment and Implementation of the Decision Element Extraction Model in Legal Document Fact Description Based on Deep Learning

**4.1. Problem Description.** The decision element extraction model mainly includes three parts: sentence semantic representation based on BERT, mat attention mechanism, and sentence label prediction based on the CNN. Sentence semantic representation uses the BERT model to generate sentence vector representation. The MAT attention mechanism weakens the filled part of the input vector and gives it the weight of the character level of the real vector. Sentence label prediction: after triple convolution and maximum pooling of input vectors, the Softmax classifier is used to predict the final label. The structure of the sentence label prediction model is shown in Figure 14.

**4.2. Mask-Based Multihead Self-Attention Mechanism.** The Bert model can fully express the relationship characteristics between words and sentences. Using the Bert model as a word embedding method can improve the effect of the model. The sentence length of legal judgment elements is mostly within 100 words, while the sentence length parameter with the best effect of the Bert model is usually 512, and the sentence length varies greatly [20]. Traditional models often use fixed parameters to specify all vector dimensions and use 0 filling to supplement the dimensions of short sentences. This

method does not accurately capture the effective features of sentences. In order to weaken the negative impact of this difference on the model effect, a multihead self-attention mechanism based on the mask method is proposed in this paper. Taking the sentence vector of “having a child after marriage” in the fact description as an example, the mask method weakens the filling vector that does not represent any meaning. During Softmax operation, the irrelevant 0 vector will hardly be assigned weight, effectively reducing the influence of irrelevant vectors on the real vector.

The specific implementation method of the multihead self-attention mechanism based on the mask method is as follows.

The fact matrix is transformed with the bes matrix to obtain the semantic description of the nonlinear model;  $S'$  is a combination matrix of fact description real vector and filled 0 vector. In order to minimize the influence of the filled 0 vector on the weight of the real sentence vector, this paper uses the mask method to weaken the filled vector to obtain matrix  $A$ . The weakening method is to quantitatively transform the fact description direction of the matrix by a high order of magnitude. The order of magnitude used in this paper is  $e10$ , so that the filling vector value of matrix  $S'$  is small enough, and the weight allocated to the filling vector during Softmax weight allocation is small enough to be ignored to achieve the effect of removal [21]. The obtained matrix  $a$  is normalized by Softmax to obtain the word vector correlation strength  $a'$  of the fact description matrix. The correlation strength  $a'$  is a one-to-one corresponding to the matrix  $S$ , and the weight of the vector is calculated to obtain the final fact description matrix  $o_i$ .  $N$  different  $o_i$  is obtained by using  $n$  different  $W_i$ . After splicing all the  $o_i$ , the fact description matrix  $Hatt$  with word meaning weight is obtained, and  $Hatt$  is used as the input part of convolution.

$$H^{\text{att}} = \text{concat}(O_1, O_2, \dots, O_n). \quad (9)$$

**4.3. Character Level Convolution.** In this paper, the convolution neural network with multiple convolution cores is used. The convolution neural network can accurately extract the local feature information in the convolution core by using the sliding convolution core, and the fusion of multiple convolution cores can expand the range of collected information. The Hatt matrix uses three filters, W1, W2, and W3, to generate local features for windows with sizes of h1, h2, and h3, respectively. If there are q, k, and v filters in W1, W2, and W3, respectively, the final eigenvector is

$$\begin{aligned} Z_1 &= [C1, 1, C2, 1, \dots, Cq, 1], \\ Z_2 &= [C1, 2, C2, 2, \dots, Cq, 2], \\ Z_3 &= [C1, 3, C2, 3, \dots, Cq, 3]. \end{aligned} \quad (10)$$

The whole feature extraction layer extracts the total features described by the facts. Zall and pooling methods select the local optimal features to obtain ZM, as shown in

$$\begin{aligned} Z_{\text{all}} &= \text{concat}(Z_1, Z_2, Z_3), \\ Z_M &= \max(Z_{\text{all}}). \end{aligned} \quad (11)$$

After passing through the full connection layer, ZM is used as the final classifier of the model to predict the types of decision elements.

**4.4. Prediction of Types of Judgment Elements.** This paper counts the label combinations involved in the case description in each legal document and converts all kinds of single-label combinations into an independent single label. Then the whole element category prediction problem is transformed into multiple single-label classification problems. In this paper, the cross-entropy loss function is used to calculate the loss value of label prediction probability, respectively. The loss value described by the overall fact is the sum of multiple label loss values [22]. The loss value formula of a single label is shown in

$$\begin{aligned} \text{loss}_i &= -p(xi)\log(q(xi)) + C\|xi\|_1, \\ q(xi) &= \text{softmax}(\log its i). \end{aligned} \quad (12)$$

After the Adam optimization function optimizes the lossall, the new lossall is inversely transferred to the model to make the model iterate and update the model parameters. The optimal parameters are used to predict the label to obtain the prediction probability q(xi). The label with the highest probability is the prediction label described by the fact.

## 5. Experiment and Analysis

**5.1. Data Set Construction.** In order to facilitate model processing, this paper transforms the fact description paragraphs in the data set into corresponding multiple fact description statements, combines multiple labels into a

TABLE 2: Number of data set samples.

Data	Training set	Validation set	Test set	Number of tags
Divorce	30523	3816	1500	36
Labour	26394	3310	1500	38
Loan	17824	2229	1500	39

single label, and adds “0” label corresponding nonelement statements. In order to improve the quality of training data and reduce the imbalance of the data set, this paper discards some labels and data that account for less than 0.1% of the time, such as “monthly maintenance payment” in divorce data and “labor arbitration stage is filed” in labor data. The number of data set samples is shown in Table 2.

**5.2. Evaluation Index.** The evaluation indexes used in the experiment include accuracy  $P$ , recall  $R$ , and  $F1$ . The evaluation indexes are as follows:

$$\begin{aligned} P &= \frac{TP + TN}{TP + FN + FP + TN}, \\ R &= \frac{TP}{TP + FP}, \\ F1 &= \frac{2 \cdot P \cdot R}{P + R}. \end{aligned} \quad (13)$$

**5.3. The Baseline Model.** Five baseline models were used in the experiment: the HMM model, the CRF model, the Bert model with the full connection layer model (BER-ALL), the Bert fused cyclic neural network model (BERT-LSTM), and the Bert fused convolutional neural network model (BERT-CNN) [23].

**5.4. Parameter Setting.** In terms of parameter setting, all models use the learning rate of 0.00005 on the training set to train five epochs. The learning attenuation rate is set to 0.9, dropout is set to 0.5, and the optimizer uses the Adam optimizer. The MAT method uses the multihead self-attention mechanism of 12 heads and 5 layers for weight division [24]. The window sizes of the triple convolution kernel in the convolution neural network are 2, 3, and 4 respectively.

**5.5. Experimental Results.** The decision element extraction experiments were carried out on five baseline models using three divided data sets [25]. Obviously, the results of the CNN model are better than the LSTM model on the three data sets. Similarly, in the experiment using Bert as the word embedding method, the results of the BERT-CNN model on the three data sets are significantly better than all baseline models. When using the Bert model as the word embedding method, the setting of the maximum sentence length parameter will directly affect the experimental results. In order to verify the impact of different sentence lengths on the results, this paper uses the BERT-CNN model to conduct

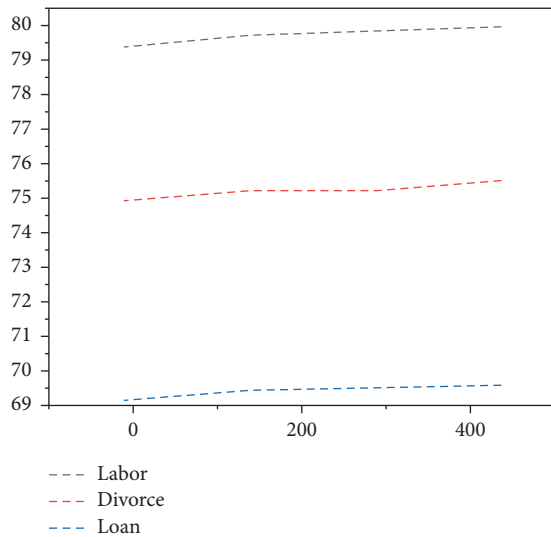


FIGURE 15: Effect of maximum sentence length on experimental results.

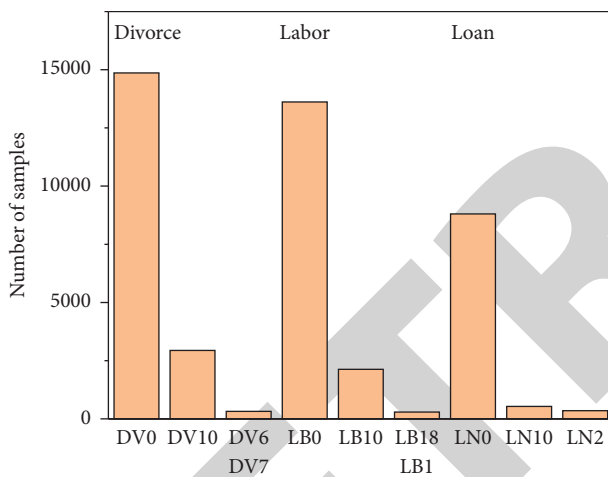


FIGURE 16: Error analysis.

comparative experiments of different sentence lengths [26, 27]. On the premise of the same other parameters, the maximum sentence length parameters are set to 64, 128, 256, and 512, respectively. The impact of the maximum sentence length on the experimental results is shown in Figure 15.

**5.6. Error Analysis.** In the three data sets, the model training bias caused by the uneven distribution of the number of samples corresponding to each label is an important reason affecting the final classification effect [28, 29]. The error analysis is shown in Figure 16.

## 6. Conclusions and Prospects

To sum up, this paper studies the element extraction task in the field of intelligent justice, formalizes the decision element sentence extraction task into a multilabel classification model of fact description sentences, and proposes a decision

element extraction method integrating Bert and the CNN (BERT-CNN). At the same time, in order to weaken the negative impact of the length difference of different sentences on the effect of the model, this paper further integrates the multihead self-attention mechanism (MAT) based on the mask method into the BERT-CNN model. Compared with the existing extraction methods, the model constructed in this paper can extract the characteristics of judicial data more accurately. However, the reference factors for judges to choose case judgment not only depend on the fact description of the case but also the relevant laws and regulations of the case. Therefore, exploring the factor extraction model integrated with relevant laws is also an important research point. In terms of named entity recognition in the judicial field, we will continue to expand the data corpus in the judicial field, construct the illegal element extraction corpus based on named entity recognition according to the legal ontologies NERCGP and NERFPPG, and complete the illegal element extraction on this basis. In the aspect of element extraction in the judicial field, the current element extraction method is mainly used to extract the defined elements of cases in the judicial field. First, classify different types of cases, then predefine the relevant elements according to the categories, and finally complete the extraction work. For different types of cases, the elements are different, and the establishment of an element system for various cases in the judicial field is time-consuming and laborious. Therefore, research on open factor extraction in the judicial field based on unsupervised or semi-supervised data is our future research direction. In the judicial field of case judgment prediction, there is often a high imbalance in the judicial field of case corpus, which belongs to the task of multilabel classification.

## Data Availability

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

## Conflicts of Interest

The author declares that there are no conflicts of interest.

## Acknowledgments

The research was supported by the Research on the Cultivation Path of Craftsman Spirit of Vocational College Students in Henan Province under the Background of Quality Improvement (No. 222400410180), supported by the Department of Science and Technology of Henan Province, and The Research on the Curriculum System of Labor Education for College Students in the New Era (No.2021YB0522), supported by the Education Department of Henan Province.

## References

- [1] X. Li, X. Kang, C. Wang, L. Dong, and S. Li, "A neural-network-based model of charge prediction via the judicial interpretation of crimes," *IEEE Access*, vol. 8, p. 1, 2020.

- [2] M. Shirinkina, "Information documents of executives: problems of identification and description," *Vestnik Volgogradskogo gosudarstvennogo universiteta. Serija 2. Jazykoznanije*, vol. 4, pp. 157–169, 2020.
- [3] I. Buansari, A. Supriatno, and M. Pangestu, "Translation procedure used in translating banking term in legal documents of bca," *Wanastra Jurnal Bahasa dan Sastra*, vol. 12, no. 2, pp. 247–258, 2020.
- [4] D. Ohreen, "The managerial use of empathy: missteps into the mind of others," *Philosophy of Management*, vol. 21, no. 2, pp. 135–161, 2021.
- [5] K. Ashihara, C. B. El Vaigh, C. Chu et al., "Improving topic modeling through homophily for legal documents," *Applied Network Science*, vol. 5, no. 1, pp. 77–20, 2020.
- [6] J. Lee, "Integration of digital twin and deep learning in cyber-physical systems: towards smart manufacturing," vol. 38, no. 8, pp. 901–910, 2020.
- [7] 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.
- [8] I. V. Tubalova and M. A. Nazemtseva, "Interpretation of the markers of freedom in slavic languages and in slavic legal document discourse," *Rusin*, vol. 62, pp. 159–175, 2020.
- [9] R. C. Kore, P. Ray, P. Lade, and A. Nerurkar, "Legal document summarization using nlp and ml techniques," *International Journal Of Engineering And Computer Science*, vol. 9, no. 5, pp. 25039–25046, 2020.
- [10] A. Erohina and V. Selifanov, "Analysis of legal documents in preparation for the development of programs and significant objects test procedures critical information infrastructure of the Rf (Pdis)," *Transaction of Scientific Papers of the Novosibirsk State Technical University*, vol. 2, pp. 57–65, 2020.
- [11] L. Shen, Z. Shao, Y. Yu, and X. M. Chen, "Hybrid approach combining modified gravity model and deep learning for short-term forecasting of metro transit passenger flows," *Transportation Research Record*, vol. 2675, no. 1, pp. 25–38, 2021.
- [12] S. Shriram, B. Nagaraj, J. Jaya, S. Shankar, and P. Ajay, "Deep learning-based real-time AI virtual mouse system using computer vision to avoid COVID-19 spread," *Journal of Healthcare Engineering*, vol. 2021, Article ID 8133076, 2028 pages, 2021.
- [13] A. R. Han, "A study on the regulations of the automated legal document drafting service," *Korea Association of the Law of Civil Procedure*, vol. 24, no. 3, pp. 391–449, 2020.
- [14] M. R. Masinter, "Doj regulation restricts reliance on guidance docs," *Campus Legal Advisor*, vol. 21, no. 5, p. 6, 2021.
- [15] O. O. Sydorenko, "The ratio between the concepts of a legal act and a legal document," *Actual Problems of Native Jurisprudence*, vol. 1, no. 1, pp. 18–22, 2021.
- [16] A. A. Legalov, "First description of the male of tsherepanovia globosa korotyae, 1992 (coleoptera, curculionidae) with new taxonomic data," *Ecologica Montenegrina*, vol. 41, no. 1, pp. 10–14, 2021.
- [17] F. Solihin, I. Budi, R. F. Aji, and E. Makarim, "Advancement of information extraction use in legal documents," *International Review of Law, Computers & Technology*, vol. 35, no. 3, pp. 322–351, 2021.
- [18] X. Zhao, X. Liu, J. Liu, J. Chen, S. Fu, and F. Zhong, "The effect of ionization energy and hydrogen weight fraction on the non-thermal plasma volatile organic compounds removal efficiency," *Journal of Physics D: Applied Physics*, vol. 52, no. 14, Article ID 145201, 2019.
- [19] Y. Hu, S. Yu, S. Qin, D. Chen, J. Chu, and Y. Yang, "How to extract traditional cultural design elements from a set of images of cultural relics based on f-ahp and entropy," *Multimedia Tools and Applications*, vol. 80, no. 4, pp. 5833–5856, 2021.
- [20] A. Mahjoub, F. Vardi, and R. Rad, "A multi-criteria decision based on adaptive routing algorithms with discrete operators for on-chip networks," *The Journal of Supercomputing*, vol. 78, no. 8, pp. 10906–10929, 2022.
- [21] D. Wang, K. Liu, and X. Zhang, "A generic indirect deep learning approach for multisensor degradation modeling," *IEEE Transactions on Automation Science and Engineering*, vol. 19, pp. 1924–1940, 2022.
- [22] R. Huang, "Framework for a smart adult education environment," vol. 13, no. 4, pp. 637–641, 2015.
- [23] J. Scott and J. O'Shea, "How legal documents translated outside institutions affect lives, businesses and the economy," *International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique*, vol. 34, no. 5, pp. 1331–1373, 2021.
- [24] T. G. Kopaneva, "Features of legal documents translation," *Humanities and Social Sciences*, vol. 83, no. 6, pp. 180–186, 2020.
- [25] H. Xie, Y. Wang, Z. Gao, B. P. Ganthia, and C. V. Truong, "Research on frequency parameter detection of frequency shifted track circuit based on nonlinear algorithm," *Nonlinear Engineering*, vol. 10, no. 1, pp. 592–599, 2021.
- [26] J. Wang and P. Srikantha, "Stealthy black-box attacks on deep learning non-intrusive load monitoring models," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3479–3492, 2021.
- [27] B. Ma, J. Wu, S. Song, and W. Liu, "Assuring privacy-preservation in mining medical text materials for COVID-19 cases - a natural language processing perspective," *Open Journal of Internet of Things (OJIOT)*, vol. 6, no. 1, pp. 6–13, 2020.
- [28] J. Wu, S. Guo, H. Huang, W. Liu, and Y. Xiang, "Information and communications technologies for sustainable development goals: state-of-the-art, needs and perspectives," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2389–2406, March 2018.
- [29] J. Wu, S. Guo, J. Li, and D. Zeng, "Big data meet green challenges: big data toward green applications," *IEEE Systems Journal*, vol. 10, no. 3, pp. 888–900, Sept. 2016.