World Wide Web (2022) 25:381–400
https://doi.org/10.1007/s11280-021-00873-8

# Charge prediction modeling with interpretation enhancement driven by double-layer criminal system

Lin Li[1] · Lingyun Zhao[1] · Peiran Nai[1] · Xiaohui Tao[2]

Received: 5 December 2020 / Revised: 13 February 2021 / Accepted: 25 February 2021 /
Published online: 11 May 2021

**Abstract**

With the rapid development of artificial intelligence and the increasing demand for legal intelligence, using AI methods to predict legal judgments has become a hot spot in recent years. Charge prediction is one of the core tasks of Legal Judgment Prediction (LJP). It aims to predict charge from complicated legal facts, so as to help the court make judgments or provide legal professional guidance to non-professionals. In the field of legalAI, interpretability is crucial compared to others. Reasonable interpretability can eliminate hidden dangers such as gender discrimination and provide support for judges' decisions. However, how to add the legal theory framework to the modeling to improve the interpretability is a challenge, which has few researches at present. To address this problem, we use Double-layer Criminal System as a guide to build Charge Prediction modeling called DCSCP which aims to predict charges in the criminal law of China. In general, our characteristic is to achieve multi-granularity inference of legal charges by obtaining the subjective and objective elements from the fact descriptions of legal cases. Specifically, our approach is performed in two steps: (1) extract the objective elements from the fact description and use them to generate candidate charges to achieve coarse-grained prediction; (2) extract the subjective elements from the fact description, and design the first-order predicate logic inference to realize the fine-grained charge inference in combination with the candidate charges. Experimental results show that our DCSCP can provide interpretable predictions, and it can maintain performance compared to other state-of-the-art charge prediction models.

**Keywords** Charge prediction · Double-layer criminal system · Inference · Interpretability

This article belongs to the Topical Collection: *Special Issue on Web Information Systems Engineering 2020*
Guest Editors: Hua Wang, Zhisheng Huang, and Wouter Beek

✉ Lin Li
cathylilin@whut.edu.cn

Extended author information available on the last page of the article.

## 1 Introduction

In recent years, with the rapid development of artificial intelligence, the combination of artificial intelligence and other domains has become a mainstream trend [8, 12, 18, 25, 36]. In the legal domain, using AI methods to predict legal judgments has become a hot topic [29, 39, 43, 48]. Among them, charge prediction belongs to the Legal Judgment Prediction (LJP), which focuses on predicting the charge based on the fact description of a specific case and is active in the criminal law. Charge prediction plays a very important role for legal professionals and non-professionals. For legal professionals such as judges, it will not replace the judge's decision but will help the judge make a judgment and give recommendations. For non-professionals, charge prediction will give them professional legal guidance.

Charge prediction has been studied for decades, and its methods are also emerging endlessly. From the earliest methods based on mathematical statistics [22, 33] to the present method based on deep learning [30, 44], we can see the continuous progress in the intellectualization of charge prediction. Although deep learning technology has made a great contribution to the performance, it has the drawback that it is not fully interpretable. Interpretability which means the ability of AI systems to explain their predictions has made considerable progress in many domains [2, 7, 9].

The legal domain pays more attention to interpretability than other domains. In the field of justice, every decision is crucial, and wrong decisions will produce unjust, false and wrong cases. In the judicial system, one of the most important litigation is fairness. However, in historical cases, there are still unfair cases due to racial or gender discrimination [49]. At the same time, most of the existing models are based on learning, which means that those may be biased in the training process [10]. Therefore, if the model is not interpretable, it is difficult for us to be convinced. At the same time, the model may be biased due to learning discrimination cases, and even lead to misjudgments. There have also been some interpretability related studies in recent years. Ye et al. used the generation of court opinions as the basis for interpretability [45]. Jiang et al. used deep reinforce learning to extract rationales to improve interpretability [19]. Li et al. used the interpretable Markov Logic Networks to predict judgments of divorce cases [26].

Although there have been some interpretability studies of charge prediction, to the best of our knowledge, there are few studies that explicitly use a professional legal framework to improve interpretability. However, there are some well-established and widely accepted theories in the legal domain. Crime Constitution Theory is one of them for criminal law convictions. Judges will strictly follow Crime Constitution Theory when making judgments on charges. Therefore, in this paper, we use Crime Constitution Theory as a professional legal framework to predict charges in criminal law. Among them, Double-layer Criminal System [46] is a mainstream Crime Constitution Theory [28]. The most notable feature of Double-layer Criminal System is that the criminal theory system is composed of two layers: *objective illegality* and *subjective responsibility* [46].

In this paper, we use the charge judgment process in Double-layer Criminal System to guide the Charge Prediction modeling called DCSCP which aims to apply in the process of criminal judgment to guide the modeling, thereby enhancing the interpretability of the model. Our work is inspired by the judge's decision process: When the judge makes a decision, the judge will use Crime Constitution Theory as the guide, such as Double-layer Criminal System. In this system, the judge will analyze objective illegality and subjective responsibility based on the fact descriptions of a legal case. After that, the judge will extract the objective elements and subjective elements to convict. At the same time, the judge will inevitably take law articles into consideration when making decisions. In the Double-layer Criminal System, it is relatively easy to obtain subjective elements, while the more difficult part is the automatic extraction of objective elements and the inference of charges. In response to the above difficulties, we propose an approach as shown in Figure 1.

Our approach is performed in two steps: (1) Extract the objective elements in the fact description and achieve coarse-grained prediction as shown in the upper part of Figure 1. This step corresponds to the objective illegality in the Double-layer Criminal System. We use a pre-training model and fine-tune it to train $N$ ($N$ is the total number of charges in the dataset) classifiers. For each model, we extract the objective elements of the fact description and judge whether the fact description is guilty. Therefore, candidate charges are generated to achieve coarse-grained charges prediction. Besides, distance supervision is used in this step to solve the problem of lack of annotation data. (2) Extract the subjective elements in the fact description, and realize the fine-grained charge inference as shown in the lower part of Figure 1. This step corresponds to the subjective responsibility in the Double-layer Criminal System. For each charge, the subjective element is $negligence$ or $intent$. Therefore, we regard subjective element extraction as a text classification task. The charges are mapped by subjective elements to label the subjective element for each charge. Then, a pre-training model is used to implement text classification. Finally, first-order predicate logic inference [41] is used to get the final charge. It has good interpretability and is easy to be understood and accepted by people.

Our approach is different from traditional charge prediction methods in two aspects. One is that compared with the existing interpretable methods, our approach is driven by the current most mainstream Crime Constitution System. The other is that we design two-step modeling to automatically extract key elements and implement multi-granularity inference by following this theory.

As indicated above, the advantage of our approach is multi-fold:

- Simulated conviction process with interpretation enhancement. This paper proposes a charge prediction approach that uses Double-layer Criminal System as the theoretical guidance. This approach simulates the entire process of judge conviction. As judges have done, our approach refers to legal articles and consider objective and subjective elements respectively, and finally infer the charge.

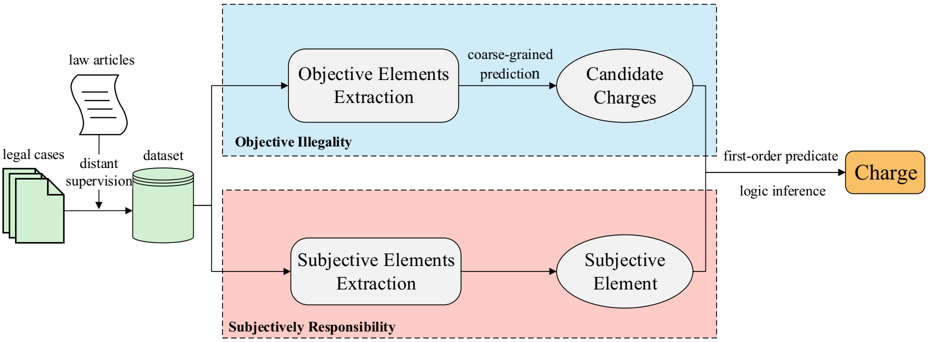


**Figure 1** Architecture of interpretable charge prediction modeling driven by double-layer criminal system

- Visualization of the charge prediction process. Compared with the popular end-to-end model modelings [43], our approach can obtain objective and subjective elements and output them separately as auxiliary information for the judge's conviction. This will be more applicable to legal professionals. For non-legal professionals, it is helpful for them to get to know the legal basis of their interesting cases.
- Comparable performance. Experiments are conducted on the one of largest public dataset called CAIL2018 [42]. The results show that our model has comparable performance with other state-of-the-art charge prediction models. At the same time, ours has better performance than an interpretable model.

The rest of this paper is organized as follows. Firstly, related work is discussed in Section 2. Secondly, in Section 3 the enlightenment of our approach is introduced and two of its modules are proposed: objective illegality driven coarse-grained charge prediction module and subjectively responsibility driven fine-grained inference module. Then, our experimental results are in Section 2. Finally, we conclude this paper in Section 5.

## 2 Related work

### 2.1 Charge prediction

The charge prediction is a subtask of LJP and has a long history. Early work usually uses methods based on mathematical statistics. Nagel used statistical mathematics to realize the scientific prediction of litigation results [33]. Segal used probit to estimate the parameters and proposed a prediction model [37]. Lauderdale et al. used information about substantive similarity among cases, to estimate judicial preferences [24]. The combination of mathematical methods and legal rules makes the predicted results interpretable. With the rapid development of deep learning, many researchers began to combine the legal field with neural networks. In recent years, the work can be roughly divided into two aspects. One is to use new models and technologies to improve results. Xu et al. proposed an end-to-end model, $LADAN$, to solve the charge prediction [43]. Hu et al. propose an attention-based neural network for charge prediction [17]. The other is to use legal knowledge to enhance model performance. Duan et al. proposed an external knowledge enhanced multi-label charge prediction approach [39]. Luo et al. proposed a neural network method based on attention, combined with relevant legal articles [30]. Although the use of deep learning technology improves the performance of charge prediction, there is certainly room for improvement in the interpretability of the above researches.

### 2.2 Interpretability in charge prediction

The charge prediction based on mathematical statistics is good at interpretable, but the accuracy is relatively poor. Compared with them, machine learning show higher performance and weaker interpretability [11]. For example, the larger the coefficient corresponding to the SVM [35] feature, the greater the importance of this feature. After deep learning technology improved the performance of charge prediction, researchers began to turn their attention to the interpretability or explainability of deep learning. Interpretability means the ability of AI systems to explain their predictions, has attracted more and more attention [19]. Hendricks et al. differentiate the interpretability between *introspection explanation* which explains how a model determines its final output and $justification\ explanation$ which produces sentences detailing how the evidence is compatible with a system output [13]. Ye et al. used generating court opinions to improve interpretability [45]. Jiang et al. used a deep reinforcement learning method to extract rationales as the basis for interpretability [19]. Zhong et al. present a model based on reinforcement learning to visualize the prediction process and give interpretable judgments [49]. Liu proposed a framework, JUMPER, inspired by the cognitive process of text reading, that models text classification as a sequential decision process [27]. All of the above work has made considerable contributions to the interpretability of charge prediction. However, to the best of our knowledge, the existing work is not guided by Crime Constitution Theory, which leads to its lack of legal professionalism. This is the reason why we propose an approach driven by Double-layer Criminal System.

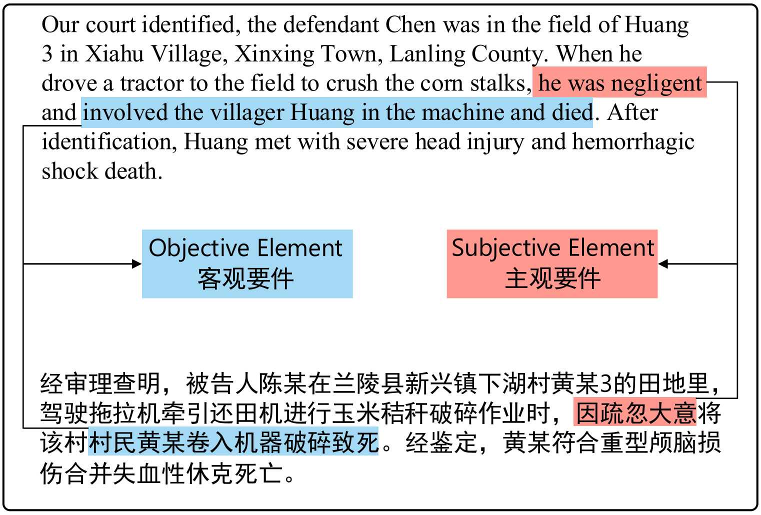
## 3 Our modeling

### 3.1 Enlightenment of our modeling

Double-layer criminal system is currently one of the most mainstream Crime Constitution Theory [28]. Double-layer Criminal System began with the discovery of negative constituent elements, which was proposed and initiated by Baumgarten in 1913 [15]. Double-layer Criminal System in this paper is proposed by Zhang [46]. Its most notable feature is that the criminal theory system is composed of two layers: *objective illegality* and *subjective responsibility*: objective illegality includes objective elements and objective obstacles; subjective responsibility includes subjective elements and subjective obstacles. Among them, objective elements and subjective elements play a decisive role in predicting charge. whereas obstacles affect sentencing and whether it constitutes a crime [40]. Given the above, we use Double-layer Criminal System as a guide and simplify Double-layer Criminal System: subjective and objective elements are selected as the basis for charge prediction.

With this enlightenment, we observed that using Double-layer Criminal System for charge prediction is mapped to two systems in the process of human cognition described by the “dual processes theory” [1]. It is believed that human cognition is divided into two systems. System 1 is an intuitive and unconscious thinking system whose operation depends on experience and association. System 2 is a unique human logical reasoning ability. It is a manifestation of human advanced intelligence. Among them, the extraction of objective and subjective elements is intuitive and based on experience, so it is mapped to the “dual processes theory” System1. The inference after the acquisition of the elements is mapped to System2.

Figure 2 is an example of fact description in the legal case. In the actual conviction process, the judge considers based on the facts provided by the prosecution and defense and adopt the Double-layer Criminal System. The judge extracts the corresponding objective element basis from the fact description and generates several candidate charges in his mind based on the objective elements. In the blue part of Figure 2, the judge extracts “Huang met with a severe head injury and hemorrhagic shock death.” as an objective basis. At this time, the judge may consider the charge of negligence causing death, intentional homicide, intentional injury, etc. as candidate charges. This process is based on experience and correlation, and can be regarded as a System 1 in the “two-process theory”. After that, the judge analyzed the subjective elements of the offender. As shown in the red part of Figure 2, the judge learns that the subjective element of the offender is negligence based on the sentence “he was negligent and involved the villager Huang in the machine and died.” This process



Our court identified, the defendant Chen was in the field of Huang 3 in Xiahu Village, Xinxing Town, Lanling County. When he drove a tractor to the field to crush the corn stalks, he was negligent and involved the villager Huang in the machine and died. After identification, Huang met with severe head injury and hemorrhagic shock death.

经审理查明，被告人陈某在兰陵县新兴镇下湖村黄某3的田地里，驾驶拖拉机牵引还田机进行玉米秸秆破碎作业时，因疏忽大意将该村村民黄某卷入机器破碎致死。经鉴定，黄某符合重型颅脑损伤合并失血性休克死亡。

**Figure 2** An example of fact description in the legal case

is based on logical reasoning and can be regarded as a System 2 in the "two-process theory". Finally, the judge considers the candidate charges and subjective elements and infers the final charges. In this case, the charge is negligence causing death.

## 3.2 Overview of our approach

Considering the importance of Crime Constitution Theory to charge prediction, we propose an approach, called DCSCP (Double-layer Criminal System-driven Charge Prediction) as shown in Figure 3. Our approach is driven by the Double-layer Criminal System and is

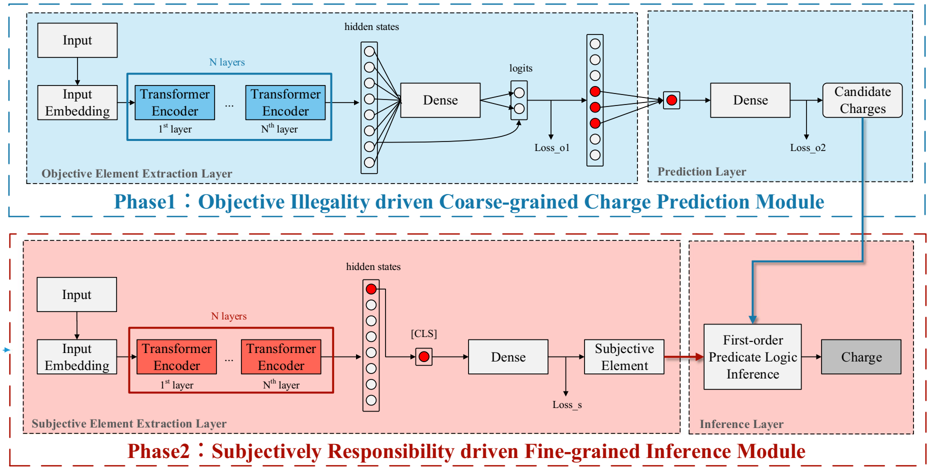


**Figure 3** The overview of DCSCP approach

committed to interpretability. Therefore, we implement multi-granularity charge prediction from both subjective and objective aspects. At the same time, our previous research proved that DCSCP is a general model so that it can meet the needs of non-professionals and professionals. In terms of subjective element extraction, this article regards it as a sentiment analysis problem. We have used similar techniques to complete the extraction of key entities in the financial domain [47]. In terms of extracting objective elements, this article uses QA technology. Before that, we had proposed a model called DeCES to solve the Legal MRC problem [34].

Our approach consists of two parts. (1) The first part at the top of Figure 3 is the objective illegality driven coarse-grained charge prediction module, which uses a pre-training model to extract the objective elements of the legal cases, and then uses One-vs-Rest classification to realize the coarse-grained charge prediction. In this module, distant supervision is used to solve the problem of the lack of labeled data. Law articles and their corresponding objective elements are used to label the objective elements of legal cases. (2) The second part at the bottom of Figure 3 is a subjectively responsibility driven fine-grained inference module, which uses text classification to extract the subjective elements of the legal cases, and then uses first-order predicate logic inference, combined with subjective elements to complete fine-grained charge prediction. Among them, the first part and the extraction of subjective elements in the second part map to System1 in Section 3.1. The fine-grained inference in the second part maps System2.

### 3.3 Objective illegality driven coarse-grained charge prediction module

This module imitates the preliminary judgments made by judges using Double-layer Criminal System. When convicting by using Double-layer Criminal System, the judge will analyze the objective elements of the case by describing the facts of the case, and make coarse-grained predictions based on the objective elements analyzed to obtain the candidate set of charges. Inspired by this, as shown in the upper part of Figure 3, this module is composed of two layers, called objective element extraction layer and prediction layer.

#### 3.3.1 Objective element extraction layer

In this layer, the fact description of the legal case is input to the multi-layer encoders, and then the hidden states of the last layer of the encoder are taken out and input to the fully connected layer to obtain objective elements. Among them, multilayer encoders can use pre-trained models such as ELECTRA.

We treat the above tasks as text classification tasks. Since there are hundreds of criminal charges and uneven data distribution, the direct use of multi-classification models may not achieve good results. Therefore, we use one-vs-rest classification to achieve multiple classifications. Specifically, this module contains $N$ models ($N$ is the number of criminal charges). Each model contains an objective element extraction layer and a prediction layer, and corresponds to one charge to determine whether the input data is this charge.

Take a model as an example. In the objective element extraction layer, we use the fact description $f$ obtained from the previous module as the model input. Input the hidden states $\mathbf{h} \in \mathbb{R}^{m \times s}$ of the twelfth layer to the dense layer, where $m$ is the maximum input length of the model and $s$ is the hidden size. Then input $\mathbf{h}$ into the dense layer to get $\mathbf{logits} \in \mathbb{R}^{2 \times s}$ as

$$\vec{logits} = \left[\vec{l_{start}}\vec{l_{end}}\right] \tag{1}$$

where $\mathbf{l_{start}}$ represents the value of each word as the start of objective elements. Then, we select the largest values as model predictions. At this time, we get the $Loss_{o1}$ of this module for this layer as shown in (2).

$$Loss_{o1} = -\sum_{i=1}^{m} (start_i \log P(\mathbf{l_{start}}) + end_i \log P(\mathbf{l_{end}})) \tag{2}$$

According to $\mathbf{l_{start}}$ and $\mathbf{l_{end}}$, we can get the predicted start position $pos_s$ and end position $pos_e$, then we get the fact description sentences corresponding to $pos_s$ and $pos_e$, that is, objective elements.

### 3.3.2 Prediction layer

In this layer, we average the vectors corresponding to the positions of the objective elements generated in the previous layer to obtain a sentence-level vector. This vector is then used to input to the dense layer, using GELUs as an activation function to determine the charge corresponding to the objective element.

However, during training, there will be unexpected situations where $pos_s \le pos_e$. This situation is mainly due to two reasons: The first is that we limit the maximum input length, which may result in the objective elements not being input. The second reason is that the previous step predicted the wrong $pos_s$ and $pos_e$.

$$\mathbf{I} = \begin{cases} mean(\sum_{i=pos_s}^{pos_e} \mathbf{h_i}),\ pos_s > pos_e \\ \\ \mathbf{h_0},\ pos_s \le pos_e \end{cases} \tag{3}$$

Therefore, in order to ensure that the prediction goes smoothly, the input of the prediction layer $\mathbf{I}$ is shown in (3). In the prediction layer, if $pos_s \le pos_e$, we will use $\mathbf{h_0}$ as input $\mathbf{I}$, which is the [CLS] token. On the contrary, if $pos_s > pos_e$, we will take out the hidden states corresponding to the objective elements $\mathbf{h_{se}} \in \mathbb{R}^{(pos_e - pos_s)\times s}$, and get its mean vector $\mathbf{I} \in \mathbb{R}^{1\times s}$ as the sentence-level vector. After that, $\mathbf{I}$ is input into the dense layer for classification to determine whether the objective requirements constitute the charge corresponding to this model. Its activation function is GELUs. After the above operations, the loss function $Loss2$ corresponding to each piece of data is shown in (4).

$$Loss_{o2} = -c \log P(\hat{c}) - (1-c) \cdot \log(1 - P(\hat{c})) \tag{4}$$

When $c = 1$, it means that this piece of data corresponds to the charge of the model, and when $c = 0$, it means that this piece of data does not correspond to the charge of this model. $P(\hat{c})$ represents the probability that this piece of data corresponds to this model.

In summary, the reward function of this model $f(\theta)$ is shown in (5). And we use Adam [21] as the optimizer of this module.

$$f(\theta) = minimize\left(\alpha Loss_{o1} + \beta Loss_{o2} + \lambda \left\|\theta\right\|_2^2\right) \tag{5}$$

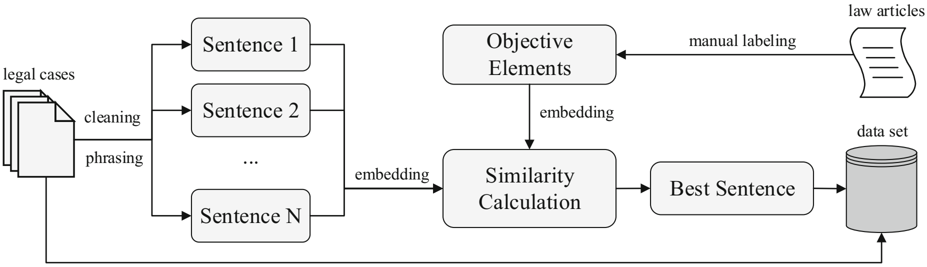


**Figure 4** The architecture of distant supervision

where $\alpha, \beta, \lambda \in \mathbb{R}$ are three hyperparameters of reward function, and $\theta$ are model parameters. After each piece of data undergoes $N$ ($N$ is the total number of all charges) times of the above operations, a candidate set $C = \{c_1, c_2, ..., c_k\}$ of possible charges of the data will be generated.

## 3.4 Training processes of coarse-grained charge prediction module

### 3.4.1 Generating label data

Due to the existing dataset, such as the data of the competition CAIL2018 [42], the label data usually only contains: fact, accusation, relevant articles, the term of imprisonment, punish of money, criminals, and other information. If we want to train an objective illegality-driven coarse-grained charge prediction module, it is impossible to directly use the current dataset. We lack the label data of objective elements. Therefore, it is inspired by the application of distant supervision in relation extraction [32]. We also use distant supervision to construct a training dataset. The specific architecture of this module is shown in Figure 4.

We use distant supervision to label the legal cases with the objective elements. Therefore, we manually labeled the objective elements $o_e$ for each charge according to the latest revision of the PRC criminal law from website.[1] As shown in Table 1, the legal experts marked the objective elements of each charge based on the law article and their own legal expertise.

For each legal case, we used the clause tool to divide the fact descriptions $F$ into multiple fact sentences $\{f_1, f_2, \cdots, f_n\}$. For each fact description $f_i$, it is embedded into a 300-dimensional vector using Skip-Gram with Negative Sampling (SGNS) [31]. In the legal cases we extracted, each $F$ is corresponding to a charge. Therefore, for the objective elements $o_e$ of each charge, we also used SGNS to embed it into a 300-dimensional vector. After that, we calculated the text similarity between $f_i$ and $o_e$, and use $f^*$ with the largest similarity as the objective element of the legal case, where $f^* \in F$. In this paper, we use Word Mover's Distance (WMD) [23] as a model for text similarity calculation. Since WMD is based on the linear programming algorithm, it is interpretable. Therefore, compared to calculation methods such as cosine distance, WMD enhances the interpretability of our approach.

[1] https://www.zuiming.net/

**Table 1** An example of legal experts labeling data. All shown examples are translated from Chinese for illustration

| Charge | Deforestation |
|---|---|
| Law Article | If you violate the provisions of the Forestry Law and deforest forests or other trees in large quantities, you will be sentenced to not more than three years of fixed-term imprisonment, criminal detention or control, and a fine or a fine alone; if you violate the provisions of the Forestry Law in large quantities, you will be sentenced to not less than three years and not more than seven years of fixed-term imprisonment and a fine. |
| Objective Elements | This charge manifests itself objectively in the act of violating national forest protection laws and regulations by arbitrarily harvesting forests or other trees owned or managed by the unit, as well as those on my own mountain, without the approval and issuance of a logging permit by the forestry administrative department and other competent departments as stipulated by law, or by holding a logging permit, but inviolation of the location, number, species and manner stipulated in the logging permit. |

### 3.4.2 Training algorithm

The training algorithm of Objective Illegality driven Coarse-grained Charge Prediction Module is shown in Algorithm 1.

**Algorithm 1** Training the coarse-grained charge prediction module.

```
Require: Fact description of legal case $F = \{f_1, f_2, \cdots, f_n\}$.
while Training do
    Run ELECTRA [4] with input $f$ while the output is $\mathbf{h}$.
    Run dense layer with input $\mathbf{h}$ while the output is logits in (2).
    for $i \leftarrow 0$ to max length do
        $pos_s \leftarrow Max(\mathbf{l_i^{start}})$
        $pos_e \leftarrow Max(\mathbf{l_i^{end}})$
    end for
    if $pos_s > pos_e$ then
        $\mathbf{I} \leftarrow mean(\sum_{i=pos_s}^{pos_e} \mathbf{h_i})$
    else
        $\mathbf{I} \leftarrow \mathbf{h_0}$
    end if
    Run dense layer with input $\mathbf{I}$ while the output is the prediction of charge.
    The final total loss function is in (5).
end while
```

Due to the use of One-vs-Rest classification, we train models with the same number of charges. For each model, the training data is divided into multiple batches. For each minibatch, we input its fact description, the start and end positions of the objective elements, and whether it is the current charge in the model. And in each step, the objective element extraction and the current charge judgment are trained at the same time, and the sum of the loss functions of the two is backpropagated in (5). Since the proportions of the two loss functions are different, we added two weight coefficients $\alpha$ and $\beta$ to control the proportion. Finally, we use Adam as the optimizer.

### 3.5 Subjectively responsibility driven fine-grained inference module

This module is composed of two layers, called subjective element extraction layer and inference layer. As shown in the lower part of Figure 3.

#### 3.5.1 Subjective element extraction layer

If a fact constitutes a crime, then its subjective element should be $intent$ or $negligence$. Every charge has subjective elements, and similarly, every legal case also has subjective elements. Given the above, we consider the extraction of subjective elements as a text classification task, or more precisely, as a two-class classification task. In this module, we use ELECTRA [4] as a pre-training model to implement text classification.

In this layer, like ELECTRA, we take the fact description $f$ in the legal case as input, and then embed it to the 12-layer transformer encoder [38]. In the last layer, we take the first tensor of hidden states $\mathbf{h_0}$ out to the downstream classification task, which is the [CLS] token mentioned in BERT [6]. Then we input $\mathbf{h_0}$ to a dense layer, and use GELUs [14] as the activation function (6).

$$\mathbf{logits} = GELUs(W_o\mathbf{h_0} + b_o) \tag{6}$$

Finally, we input **logits** to a linear layer to get the subjective elements, where $\hat{y}$ is the predictions that the sample is $intentional$. We define $y$ is the sample label as (7).

$$y = \begin{cases} 1, \text{subjective elements} = intent \\ 0, \text{subjective elements} = negligence \end{cases} \tag{7}$$

The loss function of this module. is the cross-entropy loss function as shown in (8).

$$Loss_s = -\sum_{i=1}^{N} y_i \log P(\hat{y}_i) + (1 - y_i) \cdot \log(1 - P(\hat{y}_i)) \tag{8}$$

In this layer, our train dataset comes from manual annotations by legal experts. Because the subjective element of each charge is either $intent$ or $negligence$. In the existing datasets, each legal case has its own charge. Therefore, we can label each legal case with the subjective elements according to the mapping relationship between the crime and the subjective elements.

In the training process, we reward $loss_s$ with L2 regularity and use Adam [21] to optimize this layer.

#### 3.5.2 Inference layer

In this module, we use first-order predicate logic inference to complete the final fine-grained charge inference. We combine the subjective elements output in Section 3.5.1 and the candidate set $C = \{c_1, c_2, ..., c_k\}$ of possible charges in Section 3.3.1. After that, we use the (9) inference to get the final charge of one legal case.

- $C(x)$: $x$ is the candidate charge.
- $Eq(x)$: The predicted value of the subjective element of $x$ is consistent with the subjective element corresponding to its charge.

**Table 2** The details of CAIL2018 dataset

| Dataset | Training | Valid | Test | Label |
|---|---|---|---|---|
| 1,936,794 | 1,897,957 | 17,131 | 217,016 | 202 |

- $Max(x, y)$: The probability that $x$ is the charge is greater than or equal to $y$.
- $Charge(x)$: $x$ is the final charge.

$$(\exists x)(C(x) \wedge Eq(x) \wedge (\forall y)(C(y) \wedge Eq(y) \wedge Max(x, y))) \rightarrow Charge(x) \tag{9}$$

Taking Figure 2 as an example, we can obtain the candidate charges $C = \{$ Negligence causing Death, Intentional Homicide, Intentional Injury $\}$ through Section 3.3.2 and its corresponding probability $P = \{ 0.95, 0.96, 0.87 \}$, and we can also obtain the subjective elements $negligence$ through Section 3.5.1. Thus, we can know $(\exists x)(C(x) \wedge Eq(x)) =$ $\{$ Negligence causing Death $\}$ through the mapping of the charge and the subjective elements. Similarly, it's not difficult for us to know $(\exists x)(\forall y)(C(y) \wedge Eq(y) \wedge Max(x, y)) =$ $\{$ Negligence causing Death, Intentional Injury $\}$. Finally, after the above inference, we can get $Charge(x) = \{$ Negligence causing Death $\} \wedge \{$ Negligence causing Death, Intentional Homicide $\} = \{$ Negligence causing Death $\}$.

# 4 Experiments

## 4.1 Dataset and evaluation measure

### 4.1.1 Dataset

In this paper, CAIL2018 dataset [42] are used for experiments. It is collected from China Judgement Online [2] and contains more than 2.6 million criminal law cases. It is currently a public and widely recognized dataset, and has been used as an experimental dataset in many studies to evaluate model performance [39, 43, 48, 49]. The first stage and exercise contest data are used for our experiments. The distribution of the dataset is shown in Table 2. However, in a real conviction process, judges take into account the opinions of both the prosecution and the defense. In this dataset, many legal cases only have the prosecution's opinion. Therefore, we only selected the samples that begin with "Our court identified" to compose our dataset. This kind of samples account for about 17% of the total samples. Table 3 is an example of a legal case.

In addition, CAIL2018 contains 202 different charges, and the distribution is quite uneven. According to statistics, the amount of data contained in the top50 charges accounts for about 92.7% of the total data. Therefore, we only selected the samples in top50 charges to compose our dataset. Figure 5 shows the distribution of top50 charge in training data. The five charges with the largest proportion are Theft, Reckless Driving, Intentional Injury, Traffic Accident and Smuggling, Selling, Transporting, Manufacturing Drugs.

In summary, we selected 455,445 samples from CAIL2018 top50 charges. For each sample, the fact description starts with "Our court identified" and corresponds to one charge. Its

[2] https://wenshu.court.gov.cn

**Table 3** An example of a legal case

| | |
|---|---|
| Fact | Our court identified, The defendant Zhen Moujia to Industrial and Commercial Bank of China (ICBC) Zhangye branch for card number x x x x credit card. Since the defendant Zhen Moujia held the card many times large malicious overdraft, by the Industrial and Commercial Bank of China Zhangye branch many times, the defendant Zhen Moujia more than the bank still does not return the prescribed period. |
| Charge | Credit Card Fraud |

composition is shown in Table 3. Among them, 414,780 samples are used for training, 3,980 samples for adjusting hyperparameters, and 36,685 samples for testing the final score.

### 4.1.2 Evaluation measures

In this paper, we use Macro-Precision, Macro-Recall and Macro-F1 as evaluation measure as

$$Macro - F1 = \sum\nolimits_{i=1}^{N} F1_i \tag{10}$$

$$F1_i = \frac{2(Precision_i * Recall_i)}{Precision_i + Recall_i} \tag{11}$$

where Macro-F1 is the average value of the sum of $F1_i$ for each charge. $Precision_i$ refers to the number of predicted correct charges divided by the total number of positive samples of the predicted data. $Recall_i$ refers to the number of predicted correct charges divided by the total number of positive samples in the label data.

## 4.2 Experimental settings

In this paper, our experimental environment and device information are as follows. First, our system is CentOS Linux release 7.2.1511 (core), and we have 4 TITAN Xp GPUs, each of which has 12G of memory. Secondly, our experimental environment is python3.6,

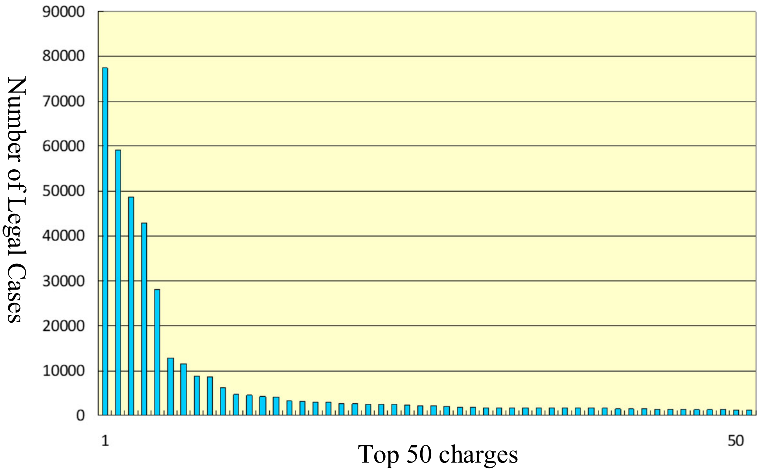


**Figure 5** The charge distribution of dataset

CUDA version is 10.0, and cuDNN version is 7.6.5. Finally, our deep learning framework is pytorch1.3.1.

In the distant supervision module in Section 3.3, we use the trained SGNS[3] as a text embedding tool. Among them, SGNS uses the data of the People's Daily News for training and generates a 300-dimensional word vector. The text similarity calculation uses the WMD algorithm in gensim[4].

In the modules in Sections 3.4 and 3.5, we use ELECTRA-180g-base[5] as the pre-training model, which was released on October 22, 2020. ELECTRA-180g-base uses 12-layer transformer encoding, its hidden size is 768, and it has 102M parameters. To prevent overfitting, we set the probability of hidden layer dropout to 0.3 in both modules. At the same time, we also added the L2 regularity as shown in (9), where $\lambda$ is set to 1e-2. Considering the limited GPU memory, we set the batch size of both modules to 16. In the subjective element extraction layer, the maximum input length is set to 256, and in the coarse-grained charge prediction module, the maximum length is set to 300.

## 4.3 Experimental results and analysis

### 4.3.1 Baselines

To evaluate the performance and interpretability of our DCSCP, we implemented several baselines to compare these two aspects.

**SVM** [35]. SVM (Support Vector Machines) is one of the commonly used classification models in machine learning. Compared with the neural network model, SVM has certain interpretability, that is, the larger the coefficient corresponding to the feature, the greater the importance of this feature.

**DPCNN** [20]. DPCNN (Deep Pyramid Convolutional Neural Networks) is a wide and effective Convolutional Neural Network for deep text classification at the word-level. It is mainly composed of the region embedding layers and convolution blocks. Compared with previous models, it can effectively extract remote relationship features in the text, and the complexity is not high.

**GRU** [3]. GRU (Gate Recurrent Unit) is a type of RNN (Recurrent Neural Network). Like LSTM (Long-Short Term Memory) [16], it is also proposed to solve problems such as long-term memory and gradients in backpropagation. Compared with LSTM, there is one less "gating" inside the GRU, which has fewer parameters than LSTM, but can also achieve functions equivalent to LSTM.

**ELECTRA** [4]. The model structure of ELECTRA is the same as BERT [6], consisting of 12 layers of Transformer Encoders[38]. Compared with BERT, ELECTRA proposes a new pre-training framework, using a combination of generator and discriminator. Experiments prove that it performs well in downstream tasks such as text classification.

**QAjudge** [49]. QAjudge is proposed to be used to implement Legal Judgment Prediction. It is based on reinforcement learning to visualize the prediction process and give interpretable judgments. Compared with the state-of-the-art model, it is interpretable, but its prediction performance is relatively inferior.

[3] https://github.com/Embedding/Chinese-Word-Vectors

[4] https://radimrehurek.com/gensim

**Table 4** Experimental results of DCSCP with others

| | Measures | Macro-Precision | Macro-Recall | Macro-F1 |
|---|---|---|---|---|
| w/ Interpretability | SVM | 73.9 | 56.2 | 63.8 |
| | QAjudge(K=3) [49] | 89.2 | 88.9 | 88.5 |
| | QAjudge(K=9) [49] | 92.6 | 92.5 | 92.3 |
| | Ours | **92.8** | **96.8** | **94.5** |
| w/o Interpretability | DPCNN | 95.3 | 95.3 | 95.2 |
| | GRU | 95.4 | 95.7 | 95.5 |
| | ELECTRA | **96.1** | **96.7** | **96.3** |

The entries in bold represent the highest values

### 4.3.2 Experiment results

The results of the comparative experiment between our DCSCP and the existing model are shown in Table 4. To be fair, we use the same data pre-processing and training methods, and the optimizer uses Adam. From Table 4 we observe:

(1) When considering interpretability, our modeling has a higher score than SVM and QAjudge. Our model is 2.2% higher in Macro-F1 than QAjudge. At the same time, our modeling has high Macro-Recall and relatively low Macro-Precision. This is because our modeling is more inclined to obtain more candidate charges when training.

(2) Compared with other models without interpretability, the Macro-F1 of our model is slightly lower than other deep learning models. Among them, our performance is only 2.2% lower than the best model. This is because we have sacrificed part of the performance of the model to a certain extent in order to extract the objective elements of the fact description and improve the interpretability. As shown in (5), this module has two loss functions. In order to extract objective elements, this makes our model more difficult to train.

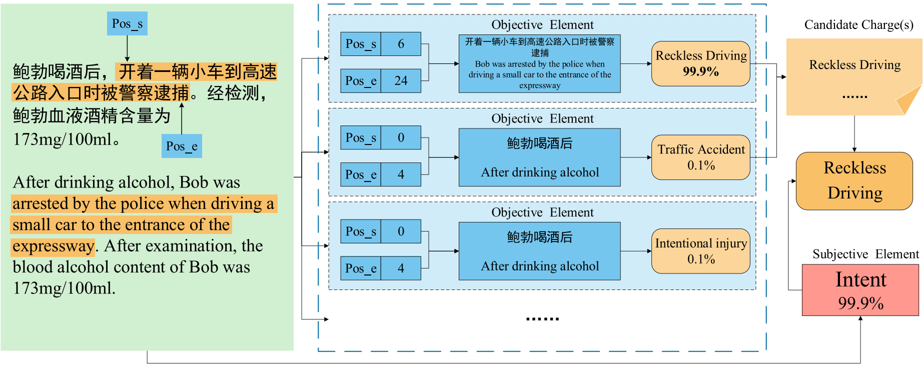
### 4.3.3 Interpretable comparative analysis

In this section, we compare and analyze the interpretability of DCSCP and baselines. It is not difficult to know that in baselines, DPCNN, GRU, and ELECTRA are not interpretable. The interpretability of SVM is also unsatisfactory. Therefore, we compare the interpretability of DCSCP and QAjudge. Since interpretability cannot be quantitatively evaluated like performance, we use a specific example for comparative analysis.

As shown in Table 5, this is an example from the QAjudge paper [49]. Through this example, it can be seen that QAjudge uses the form of Question-Answer to obtain information continuously, so as to achieve charge prediction and is interpretable. However, QAjudge did not use the professional legal framework as a guide. From a legal point of

**Table 5** An example of judgment process from the QAjudge

| Fact Description: After drinking alcohol, Bob was arrested by the police when driving a small car to the entrance of the expressway. After examination, the blood alcohol content of Bob was 173mg/100ml | |
|---|---|
| 1. Is the case related to traffic? | Yes |
| 2. Did an accident occur? | No |
| 3. Did the party drink alcohol? | Yes |
| Judgment Results: Reckless Driving | |



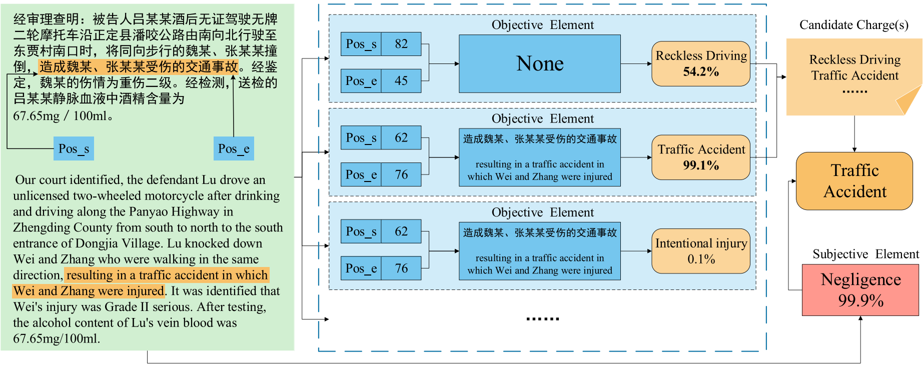
鲍勃喝酒后，开着一辆小车到高速公路入口时被警察逮捕。经检测，鲍勃血液酒精含量为173mg/100ml。

After drinking alcohol, Bob was arrested by the police when driving a small car to the entrance of the expressway. After examination, the blood alcohol content of Bob was 173mg/100ml.

**Figure 6** The same example of judgment process as QAjudge from ours

view, its judgment process is not professional enough to provide judges with corresponding assistance.

In Figure 6, we translated the example of QAjudge into Chinese and use DCSCP for charge prediction. We can find that our model can accurately judge the simpler example like Figure 6. At the same time, we show a more complicated example of the charge prediction process of our model as shown in Figure 7. DCSCP divides the sample into objective illegality and subjective responsibility driven by Double-layer Criminal System. In the first phase, corresponding to Section 3.3, DCSCP extracts objective elements and obtains charges with a probability greater than 50% to generate candidate charges. In the second phase, corresponding to Section 3.5, DCSCP analyzes subjective elements of the sample to be negligence, with a probability of 99.9%. According to the mapping relationship between charges and subjective element we know the subjective element of Reckless Driving is intent, the subjective element of Traffic Accident is negligence. Therefore, it is concluded that the charge is Traffic Accident. Compared with the judgment process of QAjudge, DCSCP also has visual interpretability. The difference is that DCSCP is driven by a professional legal framework and can give reasonable explanations for more complicated cases.



经审理查明：被告人吕某某酒后无证驾驶无牌二轮摩托车沿正定县潘咬公路由南向北行驶至东贾村南口时，将同向步行的魏某、张某某撞倒，造成魏某、张某某受伤的交通事故。经鉴定，魏某的伤情为重伤二级。经检测，送检的吕某某静脉血液中酒精含量为67.65mg / 100ml。

Our court identified, the defendant Lu drove an unlicensed two-wheeled motorcycle after drinking and driving along the Panyao Highway in Zhengding County from south to north to the south entrance of Dongjia Village. Lu knocked down Wei and Zhang who were walking in the same direction, resulting in a traffic accident in which Wei and Zhang were injured. It was identified that Wei's injury was Grade II serious. After testing, the alcohol content of Lu's vein blood was 67.65mg/100ml.

**Figure 7** A complicated example of judgment process from ours

**Table 6** Experimental results of subjective elements extraction

| Measure | Accuracy |
|---|---|
| SVM | 91.6 |
| DPCNN | 98.2 |
| GRU | 98.3 |
| Ours | **99.6** |

The entries in bold represent the highest values

### 4.3.4 Error analysis

Through random sampling to check the results, as well as the analysis of the above experiments, we classify the causes of errors into two categories:

(1) The objective elements are not accurate. There are two reasons why the objective elements are not correct. The first reason is that the input length of the model we used is limited, and some of the legal cases are too long. Through statistics of the training data, we found that the data whose objective element position exceeds the maximum input length accounted for 20.54% of the total data. If the location of the objective element is greater than the input length, it must not be extracted. The second reason is that the labels of objective elements have a few errors. Since distance supervision is used to generate labeled data, the labeled data is not strictly correct, and there may be multiple objective elements, which leads to certain errors.

(2) The subjective elements are not accurate. The correctness of the subjective elements directly determines the final result of the charge prediction. To evaluate the extraction performance of subjective elements, we use Accuracy as the metric. The reason for using Accuracy instead of $F1$ as the measure is that we want to be as correct as possible whether the subjective element is $intent$ or $negligence$. Although the extraction of subjective elements has reached 99.6% Accuracy as shown in Table 6. It still limits the upper bound of our approach performance to a certain extent.

## 5 Conclusion and future work

In this paper, we target the challenge of adding an interpretable legal theory framework to the modeling for charge prediction. We propose DCSCP which is driven by Double-layer Criminal System and implements multi-granularity inference based on objective illegality and subjective responsibility. DCSCP is driven by a mainstream Crime Constitution Theory, therefore it has legal professionalism and interpretability. Experimental results prove that DCSCP has comparable performance compared with state-of-the-art charge prediction models. At the same time, compared with the existing interpretable charge prediction models, DCSCP has better interpretability. As a result, DCSCP can provide judges with an interpretable basis for judgment or provide legal guidance to non-professionals.

In the future, our work still has room for improvement. First, we will further refine the guidance of Double-layer Criminal System on modeling. More factors such as obstacles will be added to further improve interpretability and performance. Secondly, for the acquisition of labeled data, we will seek more methods to replace distance supervision. The distance supervision limits the performance upper bound of the objective element extraction module. Finally, we will explore more suitable classification models and modify the loss function to improve performance.

## References

1. Baddeley, A.: Working memory. Science **255**(5044), 556–559 (1992)
2. Bau, D., Zhou, B., Khosla, A., Oliva, A., Torralba, A.: Network dissection: Quantifying interpretability of deep visual representations. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6541–6549 (2017)
3. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv:1406.1078 (2014)
4. Clark, K., Luong, M.T., Le, Q.V., Manning, C.D.: Electra: Pre-training text encoders as discriminators rather than generators. arXiv:2003.10555 (2020)
5. Cui, Y., Che, W., Liu, T., Qin, B., Wang, S., Hu, G.: Revisiting pre-trained models for chinese natural language processing. arXiv:2004.13922 (2020)
6. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805 (2018)
7. Dong, Y., Su, H., Zhu, J., Zhang, B.: Improving interpretability of deep neural networks with semantic information. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4306–4314 (2017)
8. Du, J., Michalska, S., Subramani, S., Wang, H., Zhang, Y.: Neural attention with character embeddings for hay fever detection from twitter. Health Information Science and Systems **7**(1), 1–7 (2019)
9. Ghoshal, B., Tucker, A.: Estimating uncertainty and interpretability in deep learning for coronavirus (covid-19) detection. arXiv:2003.10769 (2020)
10. Grgic-Hlaca, N., Redmiles, E.M., Gummadi, K.P., Weller, A.: Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. In: Proceedings of the 2018 World Wide Web Conference, pp. 903–912 (2018)
11. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine Learning **46**(1-3), 389–422 (2002)
12. He, J., Rong, J., Sun, L., Wang, H., Zhang, Y., Ma, J.: A framework for cardiac arrhythmia detection from iot-based ecgs. World Wide Web **23**(5), 2835–2850 (2020)
13. Hendricks, L.A., Akata, Z., Rohrbach, M., Donahue, J., Schiele, B., Darrell, T.: Generating visual explanations. In: European Conference on Computer Vision, pp. 3–19. Springer (2016)
14. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv:1606.08415 (2016)
15. Hirsch, H.J.: Die Lehre von den negativen Tatbestandsmerkmalen. Röhrscheid (1960)
16. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997)
17. Hu, Z., Li, X., Tu, C., Liu, Z., Sun, M.: Few-shot charge prediction with discriminative legal attributes. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 487–498 (2018)
18. Jiang, H., Zhou, R., Zhang, L., Wang, H., Zhang, Y.: Sentence level topic models for associated topics extraction. World Wide Web **22**(6), 2545–2560 (2019)
19. Jiang, X., Ye, H., Luo, Z., Chao, W., Ma, W.: Interpretable rationale augmented charge prediction system. In: Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations, pp. 146–151 (2018)
20. Johnson, R., Zhang, T.: Deep pyramid convolutional neural networks for text categorization. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 562–570 (2017)
21. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv:1412.6980 (2014)
22. Kort, F.: Predicting supreme court decisions mathematically: A quantitative analysis of the “right to counsel” cases. Am. Polit. Sci. Rev. **51**(1), 1–12 (1957)
23. Kusner, M., Sun, Y., Kolkin, N., Weinberger, K.: From word embeddings to document distances. In: International Conference on Machine Learning, pp. 957–966 (2015)
24. Lauderdale, B.E., Clark, T.S.: The supreme court’s many median justices. Am. Political Sci. Rev. **106**(4), 847–866 (2012)
25. Li, H., Wang, Y., Wang, H., Zhou, B.: Multi-window based ensemble learning for classification of imbalanced streaming data. World Wide Web **20**(6), 1507–1525 (2017)
26. Li, J., Zhang, G., Yan, H., Yu, L., Meng, T.: A markov logic networks based method to predict judicial decisions of divorce cases. In: 2018 IEEE International Conference on Smart Cloud (SmartCloud), pp. 129–132. IEEE (2018)
27. Liu, X., Mou, L., Cui, H., Lu, Z., Song, S.: Jumper: Learning when to make classification decision in reading. In: IJCAI (2018)
28. Liu, Y.: Criticism on the flatness of criminal constitution system. Legal Research, 5 (2011)

29. Long, S., Tu, C., Liu, Z., Sun, M.: Automatic judgment prediction via legal reading comprehension. In: China National Conference on Chinese Computational Linguistics, pp. 558–572. Springer (2019)
30. Luo, B., Feng, Y., Xu, J., Zhang, X., Zhao, D.: Learning to predict charges for criminal cases with legal basis. arXiv:1707.09168 (2017)
31. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
32. Mintz, M., Bills, S., Snow, R., Jurafsky, D.: Distant supervision for relation extraction without labeled data. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, pp. 1003–1011 (2009)
33. Nagel, S.S.: Applying correlation analysis to case prediction. Tex. L. Rev., 1006 (1963)
34. Nai, P., Li, L., Tao, X.: A densely connected encoder stack approach for multi-type legal machine reading comprehension. In: International Conference on Web Information Systems Engineering, pp. 167–181. Springer (2020)
35. Platt, J.: Sequential minimal optimization: A fast algorithm for training support vector machines (1998)
36. Sarki, R., Ahmed, K., Wang, H., Zhang, Y.: Automated detection of mild and multi-class diabetic eye diseases using deep learning. Health Information Science and Systems **8**(1), 1–9 (2020)
37. Segal, J.A.: Predicting supreme court cases probabilistically: The search and seizure cases, 1962-1981. Am. Political Sci. Rev. **78**(4), 891–900 (1984)
38. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)
39. Wei, D., Lin, L.: An external knowledge enhanced multi-label charge prediction approach with label number learning. arXiv:1907.02205 (2019)
40. Wen-hao, C.: Promote of essential criminal law and value in double-layer criminal system. Journal of Henan Judicial Police Vocational College, 9 (2018)
41. Whitehead, A.N., Russell, B.: 1913. principia mathematica. Cambridge **2**, 1925–1927 (1910)
42. Xiao, C., Zhong, H., Guo, Z., Tu, C., Liu, Z., Sun, M., Feng, Y., Han, X., Hu, Z., Wang, H., et al.: Cail2018: A large-scale legal dataset for judgment prediction. arXiv:1807.02478 (2018)
43. Xu, N., Wang, P., Chen, L., Pan, L., Wang, X., Zhao, J.: Distinguish confusing law articles for legal judgment prediction. arXiv:2004.02557 (2020)
44. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E.: Hierarchical attention networks for document classification. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1480–1489 (2016)
45. Ye, H., Jiang, X., Luo, Z., Chao, W.: Interpretable charge predictions for criminal cases: Learning to generate court views from fact descriptions. arXiv:1802.08504 (2018)
46. Zhang, M.: Principles of Criminal Law. The Commercial Press (2017)
47. Zhao, L., Li, L., Zheng, X.: A bert based sentiment analysis and key entity detection approach for online financial texts. arXiv:2001.05326 (2020)
48. Zhong, H., Guo, Z., Tu, C., Xiao, C., Liu, Z., Sun, M.: Legal judgment prediction via topological learning. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3540–3549 (2018)
49. Zhong, H., Wang, Y., Tu, C., Zhang, T., Liu, Z., Sun, M.: Iteratively questioning and answering for interpretable legal judgment prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 1250–1257 (2020)

## Affiliations

**Lin Li[1] · Lingyun Zhao[1] · Peiran Nai[1] · Xiaohui Tao[2]**

Lingyun Zhao
124665532@whut.edu.cn

Peiran Nai
neng245547874@whut.edu.cn

Xiaohui Tao
Xiaohui.Tao@usq.edu.au

[1] School of Computer Science and Technology, Wuhan University of Technology, Wuhan, China

[2] School of Sciences, University of Southern Queensland, Toowoomba, Australia