Towards Interactivity and Interpretability: A Rationale-based Legal Judgment Prediction Framework

Yiquan Wu[1], **Yifei Liu**[1], **Weiming Lu**[1*], **Yating Zhang**[2*]
Jun Feng[3], **Changlong Sun**[12], **Fei Wu**[1], **Kun Kuang**[1*]
[1]Zhejiang University, Hangzhou, China
[2]Alibaba Group, Hangzhou, China
[3]State Grid Zhejiang Electric Power Co., LTD, Hangzhou, China
{wuyiquan, liuyifei, luwm, kunkuang}@zju.edu.cn, yatingz89@gmail.com
changlong.scl@taobao.com, JuneFeng.81@gmail.com, wufei@cs.zju.edu.cn

Abstract

Legal judgment prediction (LJP) is a fundamental task in legal AI, which aims to assist the judge to hear the case and determine the judgment. The legal judgment usually consists of the law article, charge, and term of penalty. In the real trial scenario, the judge usually makes the decision step-by-step: first concludes the rationale according to the case's facts and then determines the judgment. Recently, many models have been proposed and made tremendous progress in LJP, but most of them adopt an end-to-end manner that cannot be manually intervened by the judge for practical use. Moreover, existing models lack interpretability due to the neglect of rationale in the prediction process. Following the judge's real trial logic, in this paper, we propose a novel Rationale-based Legal Judgment Prediction (RLJP) framework. In the RLJP framework, the LJP process is split into two steps. In the first phase, the model generates the rationales according to the fact description. Then it predicts the judgment based on the fact and the generated rationales. Extensive experiments on a real-world dataset show RLJP achieves the best results compared to the state-of-the-art models. Meanwhile, the proposed framework provides good interactivity and interpretability which enables practical use.

1 Introduction

Deep learning methods, especially the natural language process (NLP) techniques, have been employed to benefit the legal assistant systems in several aspects, such as controversy focus mining (Duan et al., 2019) and court hearing conversation generation (Ji et al., 2020). As one of the most important components of the judicial process, the task of legal judgment prediction (LJP) has been studied for decades (Kort, 1957; Keown, 1980; Lin et al., 2012; Chalkidis et al., 2019; Ma et al., 2021). LJP is defined to predict the judgment results (e.g., law article, charge, and term of penalty) of legal cases according to the fact descriptions. In the real trial scenario, as Fig. 1 shows, the judge makes the decision step-by-step: first concludes the rationale according to the case's facts and then determines the judgment. However, most existing methods focus on designing efficient features following an end-to-end framework (Luo et al., 2017; Zhong et al., 2018) while neglecting the necessity for a system to be interactive and interpretable in real trial scenarios.

To improve the practicality and effectiveness of LJP methods in real world applications, we face the following challenges: 1) **The prediction process cannot be manually intervened:** Legal tasks usually require high accuracy, but current LJP methods are commonly designed as an end-to-end manner in which the reasoning part is a "blackbox". In practical use, the judge is hard to intervene the trial logic during the prediction process, which may amplify the error and reduce the applicability of the method. 2) **The predicted judgment lacks interpretability:** The lack of interpretation is a common problem for deep learning models. However, it is especially important in the judicial field since every judgment should be given with a thorough explanation of reasoning to make the verdict more convincing.

Following the judge's trial logic, in this paper, we propose a novel Rationale-based Legal Judgment Prediction (RLJP) framework to incorporate the rationale into the LJP task. In the RLJP framework, the LJP process is divided into two steps. In the first step, it generates the charge rationale and penalty rationale respectively according to the fact description. Next, it predicts the judgment based on the fact and the corresponding rationales generated in the first step. Moreover, with the proposed TOPATT mechanism, our framework is able to better connect the two steps and also learn the topological dependencies among the three subtasks of judg-

*Corresponding Authors.

Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 4787–4799
December 7-11, 2022

Fact Description	After the hearing, the court held the facts as follows: on December 3, 2014, the defendant A went to victim B's house and said he wanted to borrow B's money. But B refused, so A pushed B to the ground, bound B with a red soft cloth belt, and robbed B's cash of 1100. After the case, the parents of the defendant A returned 1100 yuan to B. On December 7, 2014, the defendant A surrendered to the court.	
Rationale	**Charge Rationale**	For the purpose of illegal possession, the defendant A forcibly robbed B's property by means of violence.
	Penalty Rationale	After the case, the defendant A voluntarily surrendered and actively returned the stolen goods, and may be given a lighter punishment as appropriate.
Judgment	**Law Article**	Article 263
	Charge	Robbery
	Term of Penalty	42 months

Figure 1: An example of a robbery case. Given the fact description, the judge will elaborate the charge rationale to determine the charge and the penalty rationale to determine the term of the penalty. In addition, the reference law article will be indicated.

ment perdition. The advantages of this framework lie in its interactivity and interpretability. When the generated rationales are not completely correct, manual intervention can be taken to modify the generated rationale so as to change the predicted judgment. Meanwhile, the generated rationales can naturally serve as an explanation for the predicted judgment.

Since there is no public available legal document dataset that can directly support our experiments, we process the dataset released in the LAIC 2021 competition [1] through the guidance of legal experts. Extensive experiments demonstrate that our RLJP framework achieves sufficient applicability (e.g., interactivity and interpretability) with prediction accuracy improvement.

To sum up, our contributions are as follows:

- We investigate the problem of legal judgment prediction from the perspective of interactivity and interpretability.
- Simulating the judge's trial logic, we propose a novel rationale-based legal judgment prediction (RLJP) framework that splits the LJP process into two steps: rationale generation and judgment prediction.
- We demonstrate the effectiveness of RLJP in terms of both quantitative metrics and human evaluation. The experiments on a real-world legal document dataset show that our framework provides both interactivity and interpretability for practical use. Moreover, compared to SOTA models, it achieves best results on the task of legal judgment prediction, especially for the low frequency cases that RLJP outperforms the best baseline at a bigger margin.
- To help the reproducibility of proposed method, we make the code and dataset publicly available[2].

2 Related Work

2.1 Legal AI

Legal Artificial Intelligence (LegalAI) focuses on applying the technology of artificial intelligence, especially natural language processing, to benefit tasks in the legal domain (Zhong et al., 2020). In recent years, many researchers from both law and computer science fields have been exploring the potential and methods to perform judicial decisions and auxiliary tasks, aiming at helping lawyers and judges. Legal judgment prediction is the most common task in LegalAI (Chalkidis et al., 2019), where the model should predict the judgment results according to the case's fact description. Besides, there are also works on controversy focus mining (Duan et al., 2019), legal questions classification (Xiao et al., 2017), relevant case retrieval (Chen et al., 2013) and so on.

2.2 Legal Judgment Prediction

Legal judgment prediction aims to predict the judgment of legal cases according to the fact descriptions and has been studied for decades (Kort, 1957; Lin et al., 2012; Ma et al., 2021). The methods of earlier years are rule-based that require manually extracted features (Keown, 1980), which is simple and reliable, but the cost of extracting features is high. In recent years, deep learning has been proven to be effective in many domains (Shen et al., 2021; Liu et al., 2021, 2022; Li et al., 2022; Zhang et al., 2022). Thus, deep learning methods

[1] https://data.court.gov.cn/pages/laic2021.html, originally from https://wenshu.court.gov.cn/

[2] https://github.com/wuyiquan/RLJP

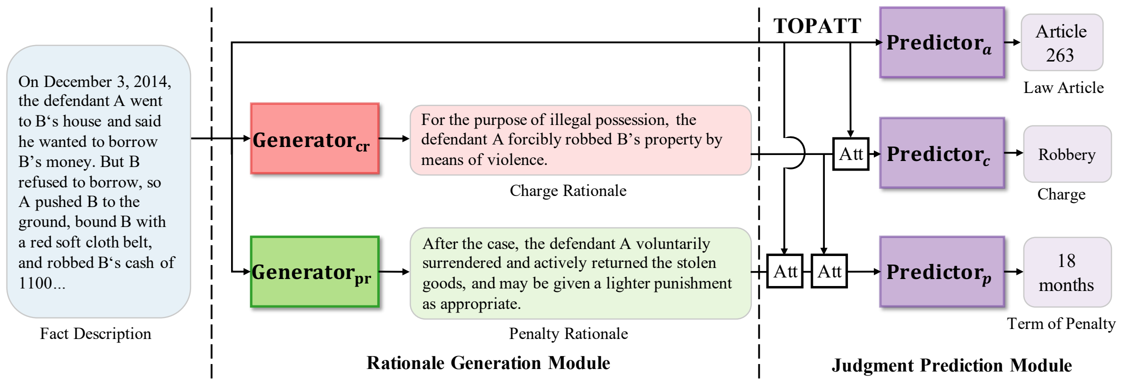

Figure 2: The overall framework of RLJP, which consists of a rationale generation module and a judgment prediction module. TOPATT mechanism (Sec.4.2.1) is proposed to better connect the two modules and learn the topological dependencies among the three subtasks of judgment prediction.

have also been applied to LJP and have achieved better performance (Zhong et al., 2020). In the meantime, these data-driven methods require far less labor. Luo et al. (2017) utilized an attention mechanism to integrate the prediction of law articles and charges. Zhong et al. (2018) considered the multiple predictions from the perspective of topology. Yue et al. (2021b) investigated the problem of rationale generation by fusing the generation and prediction. In such setting, however, if the generated rationales are inappropriate, they can't be manually intervened to revise the predicted judgment.

In this paper, following the judge's trial logic, we propose a RLJP framework to achieve both interactivity and interpretability with prediction accuracy improvement.

2.3 Rationale Models

With the increasing popularity of machine learning, understanding the reason for the prediction results becomes more and more important, especially when the model performs as a "blackbox" such as neural networks. Since there is no resolution in sight for explaining "blackbox" models, rationale models that self-explain while making decisions are often favored (Zheng et al., 2021). Lei et al. (2016) defines the rationale as some key sentences extracted from the input text and uses these extracted sentences to make predictions. Carton et al. (2018) introduces an adversarial method for producing high-recall explanations of neural text classifier decisions. Yu et al. (2019) introduces an introspective model which explicitly predicts and incorporates the outcome into the selection process. Chang et al. (2019) proposes a method to identify both factual and counterfactual rationales consistent with human rationalization. Yu et al. (2021) uses an additional attention-based predictor to overcome the concavity barrier of the extraction.

However, limited by the input text, these extraction-based rationale models are not practical and readable enough for users (Carton et al., 2020). In this paper, the rationales are generated from the input (fact description), and can be easily read and interpreted by the human judges.

3 Problem Formulation

In this work, we explore the problem of legal judgment prediction. We first clarify the definitions of the terms as follows.

Fact description consists of several descriptive sentences, which describes the identified facts (e.g., relevant events occurred in a case). Here, we denote the fact description as $f = \{w_t^f\}_{t=1}^{l_f}$, where l_f is the number of words in fact description.

Rationale is concluded from the fact description in order to determine and interpret the judgment. Notably, there exists corresponding rationales for both charge and term of penalty, named as **charge rationale** and **penalty rationale** respectively. Here, we denote the charge rationale as $cr = \{w_t^{cr}\}_{t=1}^{l_{cr}}$ and penalty rationale as $pr = \{w_t^{pr}\}_{t=1}^{l_{pr}}$, where l_{cr} is the number of words in charge rationale and l_{pr} is the number of words in penalty rationale.

Judgment includes the referred law articles, the charge and the term of penalty, which is determined by the judge according to the fact and rationales. Here, we denote the referred law article as a, the charge as c, and the term of penalty as p. The law article and the charge are in the form of labels, and the term of penalty is represented in months (numerical values).

In this work, we follow a "biomimic" design by simulating the thinking logic of judges when making decisions. We take the rationale into consideration, then the problem is defined as:

Problem 1 (Legal Judgment Prediction)**.** *Given the fact description f, our task is to generate the rationales (cr, pr) and predict the judgment (a, c, p).*

4 Rationale-based Legal Judgment Prediction (RLJP) Framework

In this section, we describe our rationale-based legal judgment prediction (RLJP) framework in detail. Fig. 2 shows the overall infrastructure of the proposed framework. It consists of a rationale generation module and a judgment prediction module. The generation module takes the fact description as input; the prediction module takes the fact description and the output of the generation module as input. Since the two modules work separately, when the generated rationales contain mistakes, manual intervention can be taken to make corrections before judgment prediction.

4.1 Rationale Generation Module

The rationale generation module aims to generate the charge rationale and penalty rationale according to the fact description.

4.1.1 Generation Model

The generation model takes the fact description as input and outputs rationales. We adopt two common generation models here:

• **BART** (Lewis et al., 2020) is a Transformer-based pretraining sequence-to-sequence model. Given the input f, the bidirectional encoder will encode it and pass it to an autoregressive generator.

• **PGN** (See et al., 2017) is a LSTM-based neural language generation model. The encoder will first encode the input, then an attention operation will be taken to integrate the input.

4.1.2 Generation Mode

The generation mode is independent of the generation model, so the generation models and the generation modes can be combined arbitrarily. We employ two generation modes here.

Separate generation mode (SGM). The input will be fed into two encoders: one for charge rationale generation, and the other for penalty rationale generation.

Joint generation mode (JGM). The two generators share one encoder.

4.2 Judgment Prediction Module

The judgment prediction module aims to predict the judgment based on the fact description and rationales. Firstly, we use three encoders that encode the fact description f, the charge rationale cr and the penalty rationale pr to get the corresponding sequences of hidden states $\mathbf{h^f}$, $\mathbf{h^{cr}}$ and $\mathbf{h^{pr}}$:

$$\begin{aligned} \mathbf{h^f} &= \mathrm{Encoder}(f), \\ \mathbf{h^{cr}} &= \mathrm{Encoder}(cr), \\ \mathbf{h^{pr}} &= \mathrm{Encoder}(pr), \end{aligned} \tag{1}$$

and we use a mean pooling to get the corresponding sentence-level representation h^f, h^{cr} and h^{pr}. Here, the encoder is replaceable.

4.2.1 TOPATT Mechanism

In practice, the judge will first decides the law articles, and then determines the charges according the law articles. Based on the referred law articles and the charges, the judge further concludes the term of penalty. Such topological dependencies among the three subtasks of legal judgment perdition have been have been applied in previous work Zhong et al. (2018). Different from the way that the dependencies are modeled on the decoder side, we propose a TOPATT mechanism during encoding process to integrate the inputs (e.g. fact and rationales) and learn the topological dependencies among the subtasks. Specifically, the charge rationale will conduct an attention operation on the fact description and the penalty rationale will perform an attention operation on the fact description and the charge rationale successively. The corresponding representation of charge rationale h^{cr}_{topatt} and penalty rationale h^{pr}_{topatt} is calculated as follow:

$$\begin{aligned} h^{cr}_{topatt} &= \mathrm{Att}(\mathbf{h^{cr}}, h^f), \\ h^{pr}_{topatt} &= \mathrm{Att}(\mathbf{h^{cr}}, \mathrm{Att}(\mathbf{h^{pr}}, h^f)). \end{aligned} \tag{2}$$

$\mathrm{Att}(\mathbf{x}, y)$ means do the attention operation in Bahdanau et al. (2015) on $\mathbf{x}$ and y, which is calculated as follow:

$$\begin{aligned} e_i &= v^T \tanh\left(W_x x_i + W_y y + b\right), \\ q &= \mathrm{softmax}\left(e\right), \end{aligned} \tag{3}$$

where v, W_x, W_y, b are learnable parameters. The output of $\mathrm{Att}(\mathbf{x}, y)$ is a weighted sum of $\mathbf{x}$:

$$x_{att} = \mathrm{Att}(\mathbf{x}, y) = \sum_i q_i x_i. \tag{4}$$

4.2.2 Judgment Prediction

Given the representation of the input, the three predictors predict the corresponding results. For the classification task (e.g., law articles and charge prediction), a fully connected layer and a softmax operation are employed in the predictor to get the probability distribution of the labels. As for the estimation of term of penalty, which is a regression task, it predicts the exact number through a fully

Consequently, the probability distribution of law articles P_a and charges P_c are calculated as:

$$\begin{aligned} P_a &= \mathrm{softmax}(\mathrm{FC}(h^f)), \\ P_c &= \mathrm{softmax}(\mathrm{FC}(h^{cr}_{topatt})), \end{aligned} \tag{5}$$

And the term of penalty p_{pred} is given by:

$$p_{pred} = \mathrm{round}(\mathrm{FC}(h^{pr}_{topatt})). \tag{6}$$

4.3 Training and Inference

The training processes of rationale generation and judgment prediction are independent. In the training of prediction module, the rationales are the ground truth of rationales.

We adopt the negative log-likelihood loss to optimize the generation module. For the charge rationale generator, the loss for time step t is the negative log-likelihood of its target word w^{cr}_t:

$$\mathcal{L}^{cr}_t = -logP(w^{cr}_t), \tag{7}$$

and the generation loss of the charge rationale generator is:

$$\mathcal{L}_{cr} = \frac{1}{T}\sum_{t=0}^{T}\mathcal{L}^{cr}_t, \tag{8}$$

where T is the length of the charge rationale. The loss of the penalty rationale generator $\mathcal{L}_{pr}$ is calculated in the same way. The overall generation loss is:

$$\mathcal{L}_{gen} = \mathcal{L}_{cr} + \mathcal{L}_{pr}. \tag{9}$$

For the judgment prediction module, cross-entropy is employed as the loss function for the classification task and log square error is utilized as the loss function for the regression task.

The loss of law article predictor $\mathcal{L}_a$ is given by:

$$\mathcal{L}_a = -\sum_{i=1}^{n_a} y_i \log\left(P_{a(i)}\right), \tag{10}$$

where $y_i = 1$ when $i = a$, otherwise, $y_i = 0$. a is the referred law article and n_a is the number of law articles. The loss of the charge predictor $\mathcal{L}_c$ is calculated in the same way as $\mathcal{L}_a$.

The loss of penalty predictor $\mathcal{L}_p$ is computed as:

$$\mathcal{L}_p = \mathrm{square}(log(p_{pred}+1) - log(p+1)), \tag{11}$$

Type	Result
# Sample	89768
# Law Article	95
# Charge	48
Avg. # tokens in fact description	402.3
Avg. # tokens in charge rationale	84.2
Avg. # tokens in penalty rationale	67.7

Table 1: Statistics of dataset.

where p is the real penalty. The overall prediction loss is:

$$\mathcal{L}_{pred} = \mathcal{L}_a + \mathcal{L}_c + \mathcal{L}_p. \tag{12}$$

At the inference stage, the generation module will generate the rationales first, then the prediction module will take the generated rationales as the input. To achieves the interactivity, after the generation, manual intervention can be taken to correct the rationales before the prediction.

5 Experiments

5.1 Dataset

Since there is no public available legal document dataset that can directly support our experiments, we process the dataset released in LAIC2021 competition[3] to get the rationales for training through the guidance of the legal experts as following steps: 1) Use keywords to extract the charge rationales and penalty rationales from the court's view automatically and objectively[4]. 2) Remove cases with too short rationales.

The dataset statistics are shown in Tab. 1. We randomly separate the dataset into a training set, a validation set, and a test set according to a ratio of 8 : 1 : 1.

5.2 Metric

5.2.1 Automatic Evaluation

Accuracy of rationale generation. To evaluate the performance of the generation, we adopt ROUGE and BLEU as the metrics. ROUGE[5] is a commonly used metric in the NLP task. We keep the results of ROUGE-1, ROUGE-2, and ROUGE-L. BLEU[6] (Papineni et al., 2002) is an automatic

[3] The origin dataset contains fact description, court's view and judgment results. Court's view contains rationales concluded from the fact. All the personal information is hidden.

[4] For example, the charge rationales are elaborated between two keywords: "The Court held that the defendant" and "The behavior has constituted", so the charge rationale can be extracted by matching the two keywords.

[5] https://pypi.org/project/rouge/

[6] http://www.nltk.org/api/nltk.test.unit.translate.html

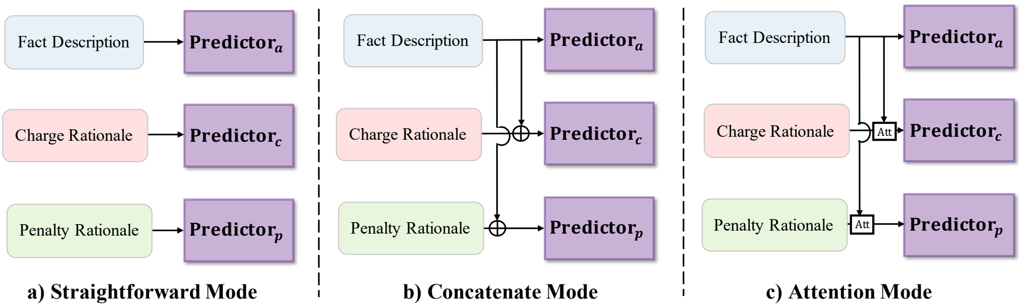

Figure 3: The other three integration modes of rationale.

Methods	Charge Rationale						Penalty Rationale					
	ROUGE			BLEU			ROUGE			BLEU		
	R-1	R-2	R-L	B-1	B-2	B-N	R-1	R-2	R-L	B-1	B-2	B-N
C3VG (Yue et al., 2021b)	67.5	48.8	65.8	67.8	61.8	57.4	50.3	31.5	45.6	47.0	40.2	34.2
BART-SGM	58.6	45.6	55.6	60.8	53.9	49.0	45.8	28.3	41.5	37.5	31.4	27.1
BART-JGM	61.1	47.9	58.1	63.4	56.5	51.6	42.2	23.6	36.2	40.7	32.9	27.2
PGN-SGM	75.5	66.4	74.0	76.0	71.6	68.8	58.3	42.2	53.3	54.7	48.0	42.7
PGN-JGM	**78.2**	**68.4**	**76.1**	**77.5**	**73.1**	**70.1**	**62.3**	**45.4**	**56.8**	**58.3**	**51.6**	**46.0**

Table 2: Results of rationale generation.

Methods	Charge Rationale		Penalty Rationale	
	Cons.	Flu.	Cons.	Flu.
C3VG	3.75	4.04	3.03	3.69
BART-SGM	3.80	3.86	3.12	3.56
BART-JGM	3.82	3.90	3.20	3.61
PGN-SGM	4.02	**4.60**	3.35	3.98
PGN-JGM	**4.05**	4.45	**3.41**	**4.03**

Table 3: Results of human evaluation.

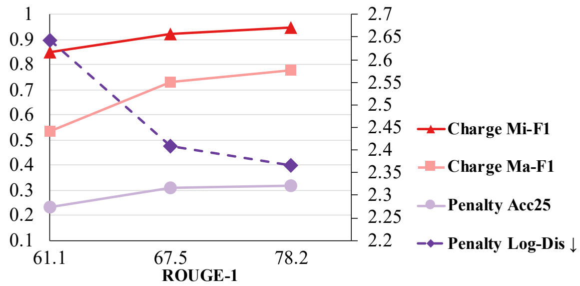

Figure 4: The relevance of the ROUGE-1 score and the performance of judgment prediction in RLJP(CNN).

evaluation for text generation tasks that highly correlates with human evaluation. We keep the result of BLEU-1 and BLEU-N(an average of BLEU-1, BLEU2, BLEU-3, and BLEU-4).

Accuracy of judgment prediction. To evaluate the performance of the prediction, we calculate the *Micro-F1* (Mi-F1) and *Macro-F1* (Ma-F1) for classification tasks (e.g., law article prediction and charge prediction). For the regression task (e.g., term of penalty prediction), we use log distance (Log-Dis) and Acc-25 to evaluate. The Log-Dis is calculated by $\log(|p_{true} - p_{predict}| + 1)$. Acc-25 is proposed in the LAIC 2021 competition where the predicted value will be considered as correct if it is within the upper and lower 25% range of the correct value.

5.2.2 Human Evaluation

We conduct a human evaluation to better analyze the quality of the generated rationales. First, we randomly sample 500 test cases, for each case, we present generated rationales from each method and the corresponding predicted judgment of RLJP(CNN)[7] with the ground truth to 5 human annotators with legal backgrounds. The evaluation is conducted following two perspectives: (1) **Consistency level**. Given the predicted results, annotators are asked to give a score on the consistency of the generated rationales (e.g., whether the penalty rationale is consistent with the predicted term of penalty). (2) **Fluency level**. Annotators are asked to give a score on the fluency of the generated rationales. For each perspective, 1 denotes the lowest score while 5 denotes the best score.

5.3 Baselines

We employ the following legal judgment prediction methods as baselines for comparison:

DPCNN (Johnson and Zhang, 2017) proposes a low-complexity deep CNN for text classification which can effectively model long-term dependencies in text. **BERT** (Devlin et al., 2019) is a masked

[7] We shuffle all the results to be fair for all the methods.

Modes	Charge		Term of Penalty		Law Article	
	Mi-F1	Ma-F1	Log-Dis↓	Acc25	Mi-F1	Ma-F1
Fact-only	93.01	70.74	2.391	29.05	91.57	45.12
Straightforward	93.96	75.32	2.766	18.44	91.57	45.12
Concatenate	93.81	73.56	2.390	27.92	92.41	47.68
Attention	94.43	75.76	2.466	27.39	91.83	46.99
TOPATT	**94.72**	**77.89**	**2.365**	**31.69**	**93.62**	**51.76**

Table 4: Judgment prediction results of RLJP(CNN) based on different rationale integration modes.

Methods	Charge		Term of Penalty		Law Article	
	Mi-F1	Ma-F1	Log-Dis↓	Acc25	Mi-F1	Ma-F1
CNN (LeCun et al., 1989)	93.01	70.74	2.391	29.05	91.57	45.12
LSTM (Sutskever et al., 2014)	93.87	70.12	2.328	31.19	92.49	45.07
Transformer (Vaswani et al., 2017)	94.39	76.66	2.448	26.35	93.40	48.84
DPCNN (Johnson and Zhang, 2017)	94.83	73.77	2.367	28.16	92.61	46.79
TopJudge (Zhong et al., 2018)	94.69	76.22	2.281	31.81	93.17	48.79
BERT (Devlin et al., 2019)	94.89	76.99	2.319	31.49	93.43	48.41
RoBERTa (Liu et al., 2019)	94.71	78.22	2.293	32.10	94.01	48.87
LADAN (Xu et al., 2020)	94.73	78.24	**2.262**	32.60	94.52	53.46
NeurJudge (Yue et al., 2021a)	94.39	77.47	2.288	31.91	92.35	52.94
C3VG (Yue et al., 2021b)	90.41	77.09	-	-	-	-
RLJP(CNN)	94.72	77.89	2.365	31.69	93.62	51.76
RLJP(LSTM)	94.48	76.96	2.405	29.67	94.50	52.69
RLJP(Transformer)	**95.21**	**79.94**	2.265	**33.15**	**95.10**	**55.81**
RLJP(Intervention)	99.18	93.14	1.962	48.95	95.10	55.81

Table 5: Results of judgment prediction. Note that PGN-JGM is adopted for rationale generation in RLJP.

language modeling (MLM) pretraining method that is pre-trained on a large corpus and fine-tuned on the downstream task. **RoBERTa** (Liu et al., 2019) is a robustly optimized BERT pretraining approach that carefully measures the impact of many key hyperparameters and training data size. **TopJudge** (Zhong et al., 2018) formalizes the dependencies among the subtasks of LJP and makes judgment predictions through topological learning. **NeurJudge** (Yue et al., 2021a) splits the fact description into two parts and encodes them separately. **LADAN** (Xu et al., 2020) uses a graph distillation operator to extract discriminative features for distinguishing confusing law articles. **C3VG** (Yue et al., 2021b) fuses the rationale generation and charge prediction, where the generated rationales are not interactive.

We use CNN (LeCun et al., 1989), LSTM (Sutskever et al., 2014) and Transformer (Vaswani et al., 2017) to implement the RLJP framework respectively. To simulate the human intervention, we use real rationales to replace the generated rationales in RLJP (Intervention).

Moreover, to evaluate the effectiveness of the proposed TOPATT mechanism, we compare it with the other three integration modes for rationale (see Fig. 3), which are described in Appendix A.1 in detail.

5.4 Experiment Results

We analyze the experimental results in this section, and the parameter settings are shown in Appendix.

Results of rationale generation: From Tab. 2 and Tab. 3, we have the following observations: 1) Joint generation mode (JGM) achieves better performance than the separate generation mode (SGM), which indicates the shared encoder has a promoting effect on rationale generation. 2) The performance of charge rationale generation is better than that of penalty rationale generation, which may be owing to the complexity of the discourse of penalty rationale. 3) The performance of PGN-JGM is better than C3VG, which proves the advantage of splitting the rationale generation and result prediction. 4) The Kappa coefficient between any two human annotators is over 0.78 (substantial agreement), which indicates the quality of human evaluation.

Fig. 4 proves that the better the generated rationales are (see the horizontal axis of ROUGE-1 score), the greater it benefits for the judgment prediction (see the increased performance of the corresponding subtasks in judgment prediction). We thus choose the PGN-JGM in the follow-up experiments for judgment prediction.

Results of judgment prediction: According to the results shown in Tab. 4, we can conclude

Methods	Highest-10 (80.50%)		Lowest-10 (0.88%)		Lowest-15 (1.86%)		Lowest-20 (3.25%)	
	Mi-F1	Ma-F1	Mi-F1	Ma-F1	Mi-F1	Ma-F1	Mi-F1	Ma-F1
LADAN	96.57	95.37	51.16	45.30	59.28	53.45	69.1	62.04
RLJP(Transformer)	**97.05**	**95.54**	**59.49**	**51.49**	**62.87**	**57.64**	**72.94**	**65.97**

Table 6: Comparison of the performance between high-frequency charges and low-frequency charges.

<table>
<tr><td>Fact Description</td><td colspan="6">After hearing, it was found that from August 2012 to October 2013, the defendant A cheated the victim B for a total of 166000 yuan on the grounds that he could help the victim B handle the case of lighter punishment for the production and sale of fake and shoddy products by his relatives. After the incident, the defendant A voluntarily surrendered and returned 100000 yuan to the victim B.</td></tr>
<tr><td rowspan="4">Rationale</td><td rowspan="2">Charge Rationale</td><td>GT</td><td colspan="4">The court held that the defendant A, for the purpose of illegal possession, used the means of fabricating facts and concealing the truth to defraud others' property for many times, with a huge amount, and his behavior constituted the crime of fraud.</td></tr>
<tr><td>GEN</td><td colspan="4">The court held that the defendant A made up facts, concealed the truth and defrauded others' property for the purpose of illegal possession. The amount was large, and his behavior constituted the crime of fraud.</td></tr>
<tr><td rowspan="2">Penalty Rationale</td><td>GT</td><td colspan="4">The defendant A voluntarily surrendered after the case and truthfully confessed the facts of the crime, may be given a lighter punishment.</td></tr>
<tr><td>GEN</td><td colspan="4">The defendant A confessed the facts of his crime.</td></tr>
<tr><td colspan="2"></td><td>GT</td><td>CNN</td><td>LADAN</td><td>RLJP(CNN)</td><td>RLJP(Intervention)</td></tr>
<tr><td rowspan="3">Judgment</td><td>Law Article</td><td>Article 262</td><td>Article 262 ✅</td><td>Article 262 ✅</td><td>Article 262 ✅</td><td>Article 262 ✅</td></tr>
<tr><td>Charge</td><td>Fraud</td><td>Robbery ❌</td><td>Robbery ❌</td><td>Fraud ✅</td><td>Fraud ✅</td></tr>
<tr><td>Term of Penalty</td><td>36 months</td><td>32 months ❌</td><td>30 months ❌</td><td>42 months ❌</td><td>36 months ✅</td></tr>
</table>

Figure 5: Case study. GT refers to the ground truth and GEN refers to the generated results.

that: 1) Compared to the Fact-only mode that uses the fact as the only input, the utilization of rationale can benefit all the subtasks of the judgment prediction. 2) TOPATT achieves the best performance compared to other three integration modes, which indicates that constructing the topological dependencies among the corresponding subtasks is effective to achieve better judgment prediction.

Tab. 5 depicts the main results of the comparison among our proposed framework and the baseline methods. We have the following observations: 1) By applying our RLJP framework, simple models (e.g., CNN, LSTM, Transformer) gain significant improvement and achieve comparable performance to the state-of-the-art methods (e.g., TopJudge, NeurJudge, LADAN). Moreover, RLJP(Transformer) achieves the best performance in almost all the evaluation metrics. 2) The performance of the baseline C3VG is unsatisfactory, which proves the fusion of the two stages of rationale generation and judgment prediction does not benefit the final result prediction. 3) RLJP framework brings more improvements to the subtasks of law article prediction (2.35% increase in Ma-F1 compared to the best baseline LADAN).

As Tab. 6 shows, the improvement of performance is more significant in low-frequency charges (e.g., Ma-F1 of Lowest-10 boosting from 45.30% to 51.49%), which proves the RLJP can also mitigate the impact of unbalanced data distribution and make the model more robust.

Evaluation of interactivity and interpretability: Tab. 5 proves the interactivity of our two-step framework that by intervening in the generated rationales, RLJP(Intervention) achieves much better performance in the final judgment prediction. The remarkable improvement (e.g., Charge Mi-F1 boosting from 95.21% to 99.18%) shows the importance of interactivity in practical use. Tab. 3 shows the qualitative metrics by human evaluations. The rationales generated by RLJP (e.g., the methods PGN-SGM and PGN-JGM) achieve better consistency performance than the ones generated by the baseline method C3VG, which proves the feasibility of rationale generation in the proposed two-step framework.

Overall, based on the rationales, our method provides interactivity and interpretability for prediction and achieves the best performance on all the subtasks of legal judgment prediction.

5.5 Case Study

Fig. 5 shows an intuitive comparison among the selected methods. In RLJP, given the fact description, the rationale generation module firstly generates the charge rationale and penalty rationale respectively. Then the prediction module predicts the judgment results based on the given rationales and facts. As the case shows, there exists some mistakes in the generated penalty rationale since it ignores the surrender of the defendant. Consequently, the prediction of the term of penalty is incorrect. After manual intervention (e.g., correcting the penalty rationale), the term of penalty prediction becomes correct, which proves the interactivity of RLJP. However, as for the other two end-to-end models (CNN and LADAN), their predicted charges and terms of penalty are both incorrect and there is no way to intervene them due to the unexplainability of their models.

6 Ethical Discussion

While AI is gaining adoption in legal justice, ethical issues have also gained increasing attention since any subtle miscalculation may trigger serious consequences (Wu et al., 2020). In such circumstances, our framework is more appropriate to offer suggestions to the judges rather than making final judgment without humans intervention. Indeed, our purpose is to provide assistance to the judges and improve their work efficiency. The system allows the judges to adjust the generated rationales and make modifications if needed.

7 Conclusion and Future Work

In this paper, we propose a novel rationale-based legal judgment prediction (RLJP) framework to solve the task of legal judgment prediction in criminal cases by thoroughly taking interactivity and interpretability into consideration. Referring to the judge's practical trial logic, the RLJP splits the classical end-to-end prediction into two phases: In the first stage, it generates the rationale according to the fact description. Next, it predicts the judgment based on the fact and the generated rationale. Experiments on a real-world dataset clearly show that the proposed framework provides good interactivity and interpretability which enables practical use, and also achieves better performance on the task of legal judgment prediction compared to the state-of-the-art models.

Based on the RLJP framework, in the future, we can explore the following directions: (1) Using the content of the law articles to improve the prediction of the term of penalty. (2) Providing more references (e.g., similar cases) for the task of rationale generation.

8 Limitations

In this section, we discuss the limitations of our work as follows:

- We limit the proposed method in judicial domain where there exists high quality of text data (e.g., verdict) and the "rationale" can be extracted directly from the legal document. As for the application to other field, it may require manual annotations.
- Since the proposed model RLJP is a two phase learning process where the first stage of rationale generation may suffer the disturbance in quality due to the noise in the input fact, which may thus influence the final prediction results. The possible solution is to decompose the input in combination with the event extraction task.

Acknowledgments

This work was supported in part by National Key Research and Development Program of China (2022YFC3340900, 2020YFC0832505), the Key Research and Development Program of Zhejiang Province, China (No. 2021C01013), National Natural Science Foundation of China (No. 62006207, No. 62037001), the Starry Night Science Fund of Zhejiang University Shanghai Institute for Advanced Study (SN-ZJU-SIAS-0010), Project by Shanghai AI Laboratory (P22KS00111), Program of Zhejiang Province Science and Technology (2022C01044) and the Fundamental Research Funds for the Central Universities (226-2022-00142, 226-2022-00051).

Finally, we would like to thank the anonymous reviewers for their helpful feedback and suggestions.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Samuel Carton, Qiaozhu Mei, and Paul Resnick. 2018. Extractive adversarial networks: High-recall explanations for identifying personal attacks in social media posts. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 3497–3507. Association for Computational Linguistics.

Samuel Carton, Anirudh Rathore, and Chenhao Tan. 2020. Evaluating and characterizing human rationales. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 9294–9307. Association for Computational Linguistics.

Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. 2019. Neural legal judgment prediction in english. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 4317–4323. Association for Computational Linguistics.

Shiyu Chang, Yang Zhang, Mo Yu, and Tommi S. Jaakkola. 2019. A game theoretic approach to classwise selective rationalization. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 10055–10065.

Yen-Liang Chen, Yi-Hung Liu, and Wu-Liang Ho. 2013. A text mining approach to assist the general public in the retrieval of legal documents. *J. Assoc. Inf. Sci. Technol.*, 64(2):280–290.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Xinyu Duan, Yating Zhang, Lin Yuan, Xin Zhou, Xiaozhong Liu, Tianyi Wang, Ruocheng Wang, Qiong Zhang, Changlong Sun, and Fei Wu. 2019. Legal summarization for multi-role debate dialogue via controversy focus mining and multi-task learning. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019*, pages 1361–1370. ACM.

Changzhen Ji, Xin Zhou, Yating Zhang, Xiaozhong Liu, Changlong Sun, Conghui Zhu, and Tiejun Zhao. 2020. Cross copy network for dialogue generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 1900–1910. Association for Computational Linguistics.

Rie Johnson and Tong Zhang. 2017. Deep pyramid convolutional neural networks for text categorization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 562–570. Association for Computational Linguistics.

R Keown. 1980. Mathematical models for legal prediction. *Computer/lj*, 2:829.

Fred Kort. 1957. Predicting supreme court decisions mathematically: A quantitative analysis of the “right to counsel” cases. *American Political Science Review*, 51(1):1–12.

Yann LeCun, Bernhard E. Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne E. Hubbard, and Lawrence D. Jackel. 1989. Backpropagation applied to handwritten zip code recognition. *Neural Comput.*, 1(4):541–551.

Tao Lei, Regina Barzilay, and Tommi S. Jaakkola. 2016. Rationalizing neural predictions. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 107–117. The Association for Computational Linguistics.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 7871–7880. Association for Computational Linguistics.

Mengze Li, Tianbao Wang, Haoyu Zhang, Shengyu Zhang, Zhou Zhao, Jiaxu Miao, Wenqiao Zhang, Wenming Tan, Jin Wang, Peng Wang, Shiliang Pu, and Fei Wu. 2022. End-to-end modeling via information tree for one-shot natural language spatial video grounding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 8707–8717. Association for Computational Linguistics.

Wan-Chen Lin, Tsung-Ting Kuo, Tung-Jia Chang, Chueh-An Yen, Chao-Ju Chen, and Shou-de Lin. 2012. Exploiting machine learning models for chinese legal documents labeling, case classification, and sentencing prediction. *Int. J. Comput. Linguistics Chin. Lang. Process.*, 17(4).

Nayu Liu, Xian Sun, Hongfeng Yu, Fanglong Yao, Guangluan Xu, and Kun Fu. 2022. Abstractive summarization for video: A revisit in multistage fusion network with forget gate. *IEEE Transactions on Multimedia*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Zhenguang Liu, Peng Qian, Xiang Wang, Lei Zhu, Qinming He, and Shouling Ji. 2021. Smart contract vulnerability detection: From pure neural network to interpretable graph feature and expert pattern fusion.

In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 2751–2759. ijcai.org.

Bingfeng Luo, Yansong Feng, Jianbo Xu, Xiang Zhang, and Dongyan Zhao. 2017. Learning to predict charges for criminal cases with legal basis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 2727–2736. Association for Computational Linguistics.

Luyao Ma, Yating Zhang, Tianyi Wang, Xiaozhong Liu, Wei Ye, Changlong Sun, and Shikun Zhang. 2021. Legal judgment prediction with multi-stage case representation learning in the real court setting. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 993–1002. ACM.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, July 6-12, 2002, Philadelphia, PA, USA*, pages 311–318. ACL.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 1073–1083. Association for Computational Linguistics.

Kai Shen, Lingfei Wu, Siliang Tang, Yueting Zhuang, Zhen He, Zhuoye Ding, Yun Xiao, and Bo Long. 2021. Learning to generate visual questions with noisy supervision. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 11604–11617.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 3104–3112.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008.

Yiquan Wu, Kun Kuang, Yating Zhang, Xiaozhong Liu, Changlong Sun, Jun Xiao, Yueting Zhuang, Luo Si, and Fei Wu. 2020. De-biased court's view generation with causality. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 763–780. Association for Computational Linguistics.

Guangyi Xiao, Even Chow, Hao Chen, Jiqian Mo, Jingzhi Guo, and Zhiguo Gong. 2017. Chinese questions classification in the law domain. In *14th IEEE International Conference on e-Business Engineering, ICEBE 2017, Shanghai, China, November 4-6, 2017*, pages 214–219. IEEE Computer Society.

Nuo Xu, Pinghui Wang, Long Chen, Li Pan, Xiaoyan Wang, and Junzhou Zhao. 2020. Distinguish confusing law articles for legal judgment prediction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 3086–3095. Association for Computational Linguistics.

Mo Yu, Shiyu Chang, Yang Zhang, and Tommi S. Jaakkola. 2019. Rethinking cooperative rationalization: Introspective extraction and complement control. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 4092–4101. Association for Computational Linguistics.

Mo Yu, Yang Zhang, Shiyu Chang, and Tommi S. Jaakkola. 2021. Understanding interlocking dynamics of cooperative rationalization. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 12822–12835.

Linan Yue, Qi Liu, Binbin Jin, Han Wu, Kai Zhang, Yanqing An, Mingyue Cheng, Biao Yin, and Dayong Wu. 2021a. Neurjudge: A circumstance-aware neural framework for legal judgment prediction. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 973–982. ACM.

Linan Yue, Qi Liu, Han Wu, Yanqing An, Li Wang, Senchao Yuan, and Dayong Wu. 2021b. Circumstances enhanced criminal court view generation. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 1855–1859. ACM.

Rongzhi Zhang, Yue Yu, Pranav Shetty, Le Song, and Chao Zhang. 2022. Prompt-based rule discovery and boosting for interactive weakly-supervised learning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 745–758, Dublin, Ireland. Association for Computational Linguistics.

Yiming Zheng, Serena Booth, Julie Shah, and Yilun Zhou. 2021. The irrationality of neural rationale models. *CoRR*, abs/2110.07550.

Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. 2018. Legal judgment prediction via topological learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 3540–3549. Association for Computational Linguistics.

Haoxi Zhong, Chaojun Xiao, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. 2020. How does NLP benefit legal system: A summary of legal artificial intelligence. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 5218–5230. Association for Computational Linguistics.

A Appendices

A.1 Rationale Integration Modes

Here, we show the detailed description of the rationale integration modes in the Fig. 3.

• **Straightforward mode**. The rationales are directly fed to the predictor to get the corresponding results.

• **Concatenate mode**. The representation of rationales will be concatenated with the representation of fact description.

$$\begin{aligned} h_{con}^{cr} &= \mathrm{Concat}(h^{cr}, h^{f}), \\ h_{con}^{pr} &= \mathrm{Concat}(h^{pr}, h^{f}), \end{aligned} \tag{13}$$

where Concat means concatenate operation.

• **Attention mode**. The rationales do an attention operation(Vaswani et al., 2017) based on the representation of fact description to get the final representation.

$$\begin{aligned} h_{att}^{cr} &= \mathrm{Att}(\mathbf{h^{cr}}, h^{f}), \\ h_{att}^{pr} &= \mathrm{Att}(\mathbf{h^{pr}}, h^{f}), \end{aligned} \tag{14}$$

where Att means attention operation.

A.2 Experiment Parameters

Our experiment is carried out on two V100 GPU, and all the baseline models adopt the parameters in their original papers. Baselines including NeurJudge, LADAN and C3VG runs 16 epochs in training. BERT and RoBERTa runs 10 epochs in training. The other models runs 100 epochs in training. All the models evaluate once on test set. Tab. 7 shows the number of parameters of all methods.

Model	# of Param
CNN	3.5M
LSTM	7.6M
Transformer	7.9M
DPCNN	6.8M
BERT	110M
RoBERTa	110M
TopJudge	9.5M
NeurJudge	8.3M
LADAN	10.7M
C3VG	54.6M
RLJP(CNN)	19.5M
RLJP(LSTM)	18.7M
RLJP(Transformer)	21.9M

Table 7: The number of parameters of all methods.

A.3 Case Demonstration

Here, we demonstrate another two cases in Fig. 6 and Fig. 7.

<table>
<tr><td>Fact Description</td><td colspan="6">After hearing, it was found that at about 22:00 on January 26, 2015, the defendants A rushed to the east lane of xxx Road in xxx County, beat the victim B who passed by, and robbed him of a mobile phone, a MP3, a pack of cigarettes and a small amount of cash. According to the forensic identification of xxx County Public Security Bureau, the crown of the victim B upper left tooth was broken, and the injury was slight. Identified by xxx County Price Certification Center: the robbed mobile phone is worth 400 yuan. After the incident, the mobile phone has been chased back and the close relatives of the defendant A compensated the victim for the loss of 1500 yuan, and the victim B expressed understanding to the defendant A.</td></tr>
<tr><td rowspan="4">Rationale</td><td rowspan="2">Charge Rationale</td><td>GT</td><td colspan="4">The court held that the defendant A gang up to forcibly rob others' property by violence for the purpose of illegal possession.</td></tr>
<tr><td>GEN</td><td colspan="4">The court held that the defendant A robbed others' property by violent means for the purpose of illegal possession.</td></tr>
<tr><td rowspan="2">Penalty Rationale</td><td>GT</td><td colspan="4">The defendant A pleaded guilty with a good attitude. According to the facts of the crime, the nature and circumstances of the crime and the degree of harm to society, a lighter punishment and probation can be applied.</td></tr>
<tr><td>GEN</td><td colspan="4">The defendant A confessed his crime truthfully, which is a confession and can be given a lighter punishment as appropriate. According to the criminal circumstances and repentance of the defendant A, probation can be applied to him.</td></tr>
<tr><td colspan="2"></td><td>GT</td><td>CNN</td><td>LADAN</td><td>RLJP(CNN)</td><td>RLJP (Intervention)</td></tr>
<tr><td rowspan="3">Judgment</td><td>Law Article</td><td>Article 263</td><td>Article 263 ✅</td><td>Article 263 ✅</td><td>Article 263 ✅</td><td>Article 263 ✅</td></tr>
<tr><td>Charge</td><td>Robbery</td><td>Robbery ✅</td><td>Robbery ✅</td><td>Robbery ✅</td><td>Robbery ✅</td></tr>
<tr><td>Term of Penalty</td><td>36 months</td><td>56 months ❌</td><td>42 months ❌</td><td>33 months ❌</td><td>36 months ✅</td></tr>
</table>

Figure 6: Case demonstration 1.

<table>
<tr><td>Fact Description</td><td colspan="6">After hearing, it was found that at about 16:00 on January 29, 2016, the defendant A drove a Mazda car No. xxx in xxx Township, xxx County. When he wanted to stop at his door, he mistakenly took the accelerator as the brake and hit his neighbor the victim B sitting at his door, resulting in the invalid rescue and death of the victim B. The defendant A voluntarily surrendered to the traffic police brigade of xxx Public Security Bureau on January 30, 2016. After the incident, the defendant A and the close relatives of the victim reached a civil compensation agreement, which has been partially fulfilled and obtained the understanding of the close relatives of the victim.</td></tr>
<tr><td rowspan="4">Rationale</td><td rowspan="2">Charge Rationale</td><td>GT</td><td colspan="4">The court held that the defendant A's negligence caused the death of the victim B.</td></tr>
<tr><td>GEN</td><td colspan="4">The court held that the defendant A's negligence in driving a motor vehicle outside the scope of public transport management caused one death.</td></tr>
<tr><td rowspan="2">Penalty Rationale</td><td>GT</td><td colspan="4">The circumstances of the crime are relatively minor. Since the defendant A voluntarily surrendered after committing a crime and truthfully confessed his crime. A lighter punishment may be imposed according to law.</td></tr>
<tr><td>GEN</td><td colspan="4">The defendant A voluntarily surrendered after committing a crime and truthfully confessed his crime. He surrendered and should be given a lighter punishment.</td></tr>
<tr><td colspan="2"></td><td>GT</td><td>CNN</td><td>LADAN</td><td>RLJP(CNN)</td><td>RLJP (Intervention)</td></tr>
<tr><td rowspan="3">Judgment</td><td>Law Article</td><td>Article 233</td><td>Article 233 ✅</td><td>Article 233 ✅</td><td>Article 233 ✅</td><td>Article 233 ✅</td></tr>
<tr><td>Charge</td><td>Involuntary Manslaughter</td><td>Involuntary Manslaughter ✅</td><td>Involuntary Manslaughter ✅</td><td>Involuntary Manslaughter ✅</td><td>Involuntary Manslaughter ✅</td></tr>
<tr><td>Term of Penalty</td><td>12 months</td><td>15 months ❌</td><td>13 months ❌</td><td>12 months ✅</td><td>12 months ✅</td></tr>
</table>

Figure 7: Case demonstration 2.