Legal Judgment Prediction via Event Extraction with Constraints

Yi Feng[1], Chuanyi Li[1] and Vincent Ng[2]

1 State Key Laboratory for Novel Software Technology, Nanjing University, China

2 Human Language Technology Research Institute, University of Texas at Dallas, USA

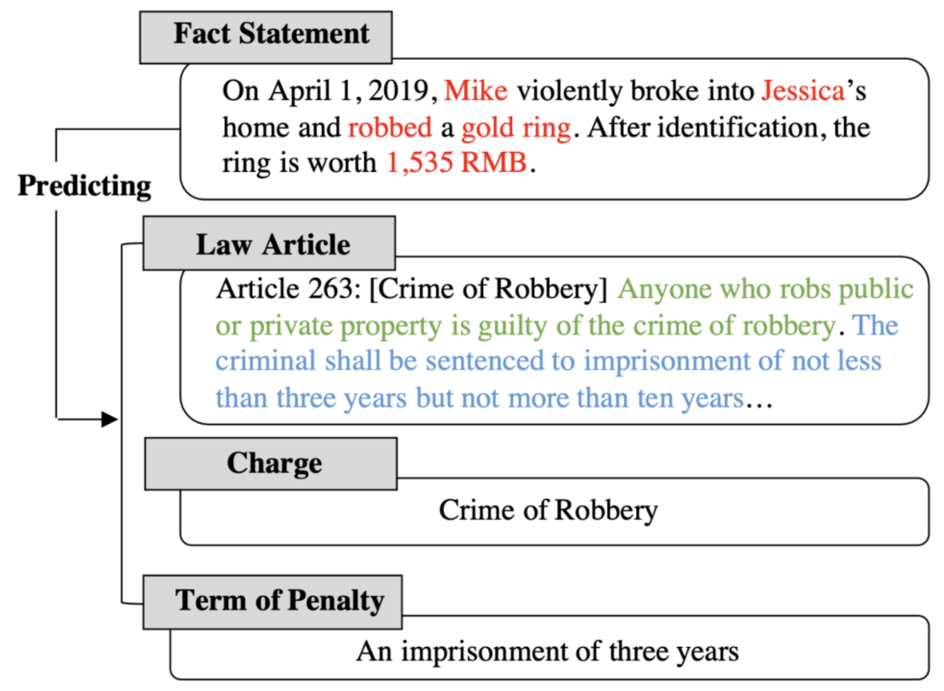
Legal Judgment Prediction (LJP)

- predicts the court's outcome given the facts of a legal case
- has been investigated in the context of different languages

Chinese Legal Judgment Prediction

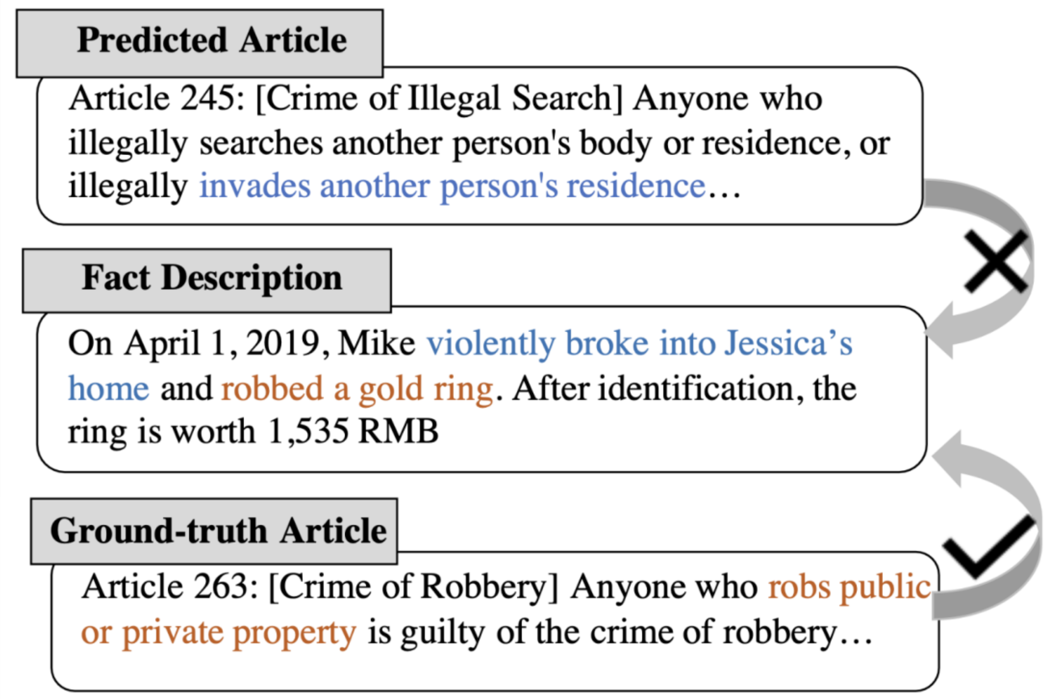
- Three subtasks: (1) **law article prediction**, (2) **charge prediction** and (3) **terms of penalty prediction**

- Input: a fact statement
- Outputs: law article -> charge -> term of penalty



Two Weaknesses of Existing LJP Models

- **Weakness 1:** failure to locate the key event information that determines the judgment results.
- **Example**: wrongly predicting the law article related to illegal search for a robbery case since many words describe the break-in process even though the main point is about robbery



Two Weaknesses of Existing LJP Models

- **Weakness 2**: inconsistent model outputs
- **Example**: wrongly predicting 5-7 years imprisonment, whereas the law article stipulates that the maximum prison term is 5 years.

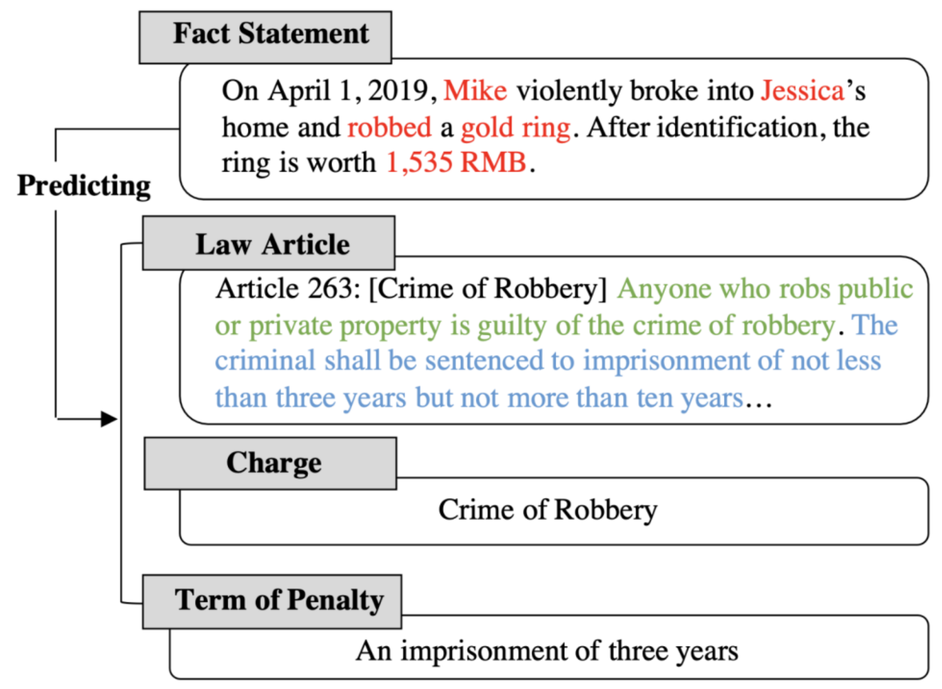
Fact Statement: The criminal Song gave birth to a baby boy in the bathroom of the Beijing-Shanghai Expressway Service Area at about 9:30 on March 29, 2016, and abandoned the baby boy in the bathroom.
Predicted: Article 261; Crime of abandoning babies; **5-7 years** imprisonment
Ground-truth: Article 261; Crime of abandoning babies; **9-12 months** imprisonment

Goal

- Improve Chinese legal judgment prediction by addressing the aforementioned weaknesses
  - failure to locate the key event information that determines the judgment results
  - inconsistent model outputs

Addressing Weakness 1 (Failure to locate key event info)

- **Observations**:
  - A law article consists of two parts: (1) the **event pattern**, which stipulates the behavior that violates the law and (2) the **judgment**, which describes the corresponding penalties
  - if we can detect the event pattern of a law article in the facts of a case, we can infer the judgment from the law article
- **Idea:** extract the fine-grained key event information and use it to match the event pattern.



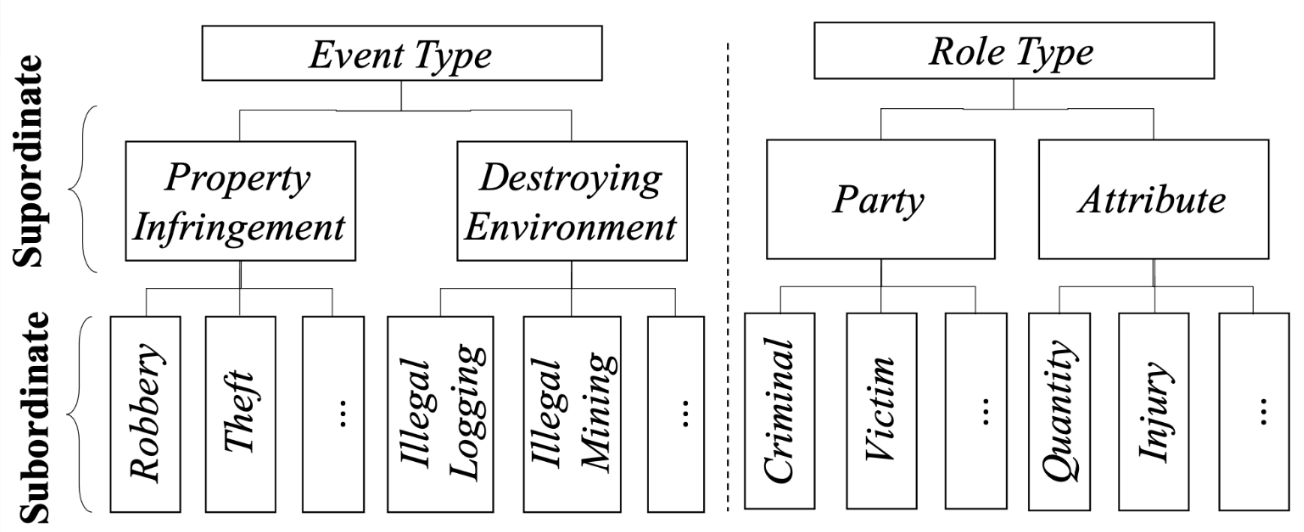
	Argument	Role
Who is the criminal?	Mike	Criminal
Who is the victim	Jessica	Victim
What happened?	robbed	Trigger-Rob
What were robbed?	gold ring	Property
What is the price of swag?	1,535 RMB	Quantity
Judgment Results: Article 263, Robbery, three-year imprisonment		

How to implement the idea?

- **Step 1**
  - Propose a hierarchical event definition referring to the hierarchy of law articles
- **Step 2**
  - Manually annotate a legal event dataset according to this definition
    - No existing datasets provide event annotations and judgments simultaneously

Defining the Event Hierarchy

- **Event definition**
  - Hierarchical events
  - Trigger and role types



Statistics: 4 superordinate and 16 subordinate roles; 6 superordinate and 15 subordinate trigger types

南京大學 NANJING UNIVERSITY

Collecting our Event-Annotated Dataset: LJP-E

- Step 1: judgment document collection
  - collect documents from the CAIL dataset
- Step 2: event trigger and argument role annotation
  - (1) highlight the salient words that correlate well with the event pattern of the law article
  - (2) select a trigger word and assign it a subordinate trigger type
  - (3) assign a subordinate role type to each of its arguments from a predefined role list

Statistics: 1367 cases in total

NANJING UNIVERSITY

Addressing Weakness 2 (Inconsistent outputs)

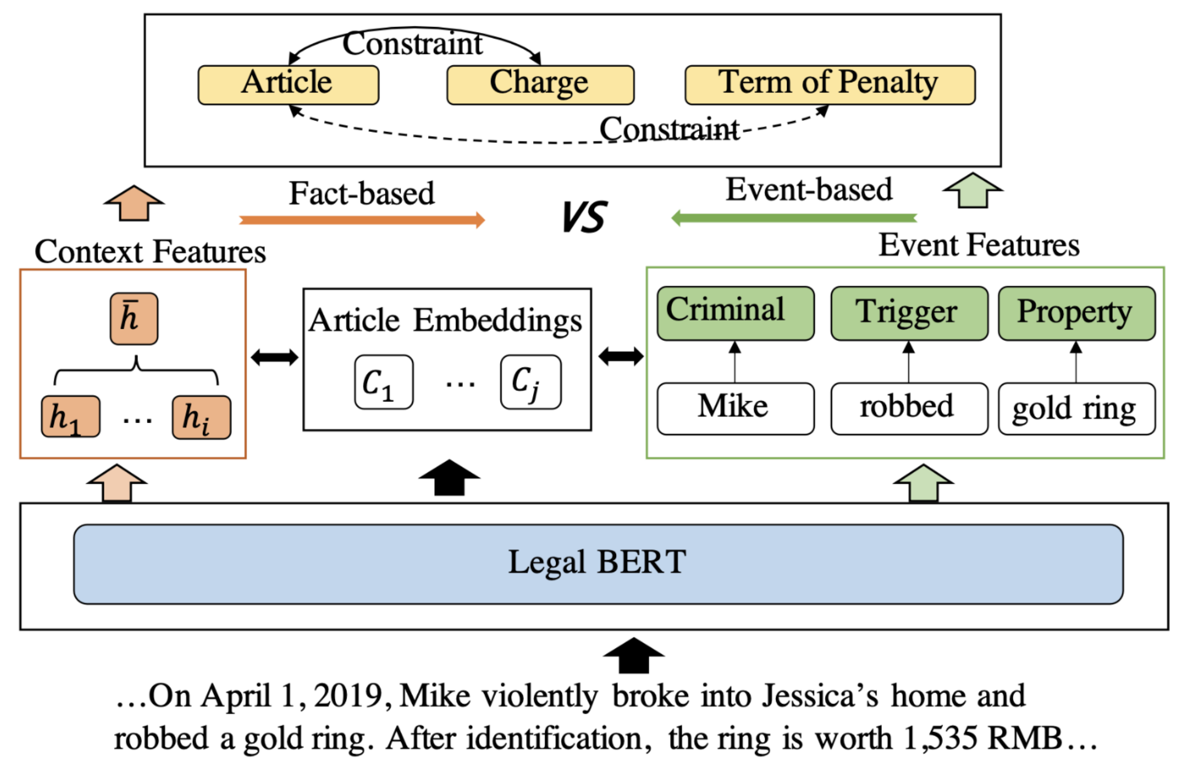
- Introduce cross-task consistency constraints
  - If a law article is detected, the allowable charges and range of term of penalty should be detected.

- Design output constraints on event extraction
  - Event-based constraints
    - Absolute constraint
      - the trigger must appear exactly once and certain roles are compulsory
    - Event-based consistency constraints
      - If a trigger type is detected, all and only its related roles should be detected

Our model: EPM

Model structure

- Hierarchical Event Extraction
- Incorporating law article semantics
- Legal judgment prediction layer



南京大學 NANJING UNIVERSITY

Dataset

Public dataset CAIL

- a large-scale publicly available Chinese legal document dataset that has been widely used.

Dataset	CAIL-small	CAIL-big
#Training Set Cases	96,540	1,489,932
#Validation Set Cases	12,903	–
#Testing Set Cases	24,848	185,647
#Law Articles	101	127
#Charges	117	140
#Term of Penalty	11	11

南京大學
NANJING UNIVERSITY

Evaluation Setting

- **Training**
  - Pre-train EPM without event components on CAIL, and then fine-tune EPM on LJP-E
- **Testing**
  - use the pretrained version of EPM to predict samples that do not belong to the 15 types
  - use the fine-tuned version of EPM to predict samples that belong to one of the 15 types
- **Baselines**
  - SOTA models: MLAC, TOPJUDGE, MBPFN, LADAN, NeurJudge
- **Metrics**
  - Accuracy (Acc), Macro-Precision (MP), Macro-Recall (MR) and Macro-F1 (F1)

Results

		Law Article				Charge				Term of Penalty			
		Acc%	MP%	MR%	F1%	Acc%	MP%	MR%	F1%	Acc%	MP%	MR%	F1%
1	MLAC	94.90	79.06	66.91	69.41	94.72	83.42	72.38	75.62	56.43	46.87	40.43	41.89
2	TOPJUDGE	95.83	82.10	71.94	74.32	95.77	85.95	77.11	79.58	58.09	47.73	42.47	44.07
3	MBPFN	95.67	84.00	74.40	76.44	94.37	85.60	75.86	77.98	55.48	47.27	38.26	40.01
4	LADAN	95.78	84.93	75.88	78.79	94.58	85.52	77.36	80.04	56.34	47.76	40.48	42.02
5	NeurJudge	95.59	84.01	75.54	77.06	94.12	85.48	77.21	79.83	55.52	47.25	40.76	42.03
6	EPM	96.63	85.93	77.60	79.72	95.88	88.67	79.49	81.99	58.19	51.50	43.25	44.99
7	EPM@G	96.72	85.79	79.68	81.77	96.45	88.78	81.93	82.84	58.67	53.93	45.86	46.58
8	MLAC+EPM	95.50	79.71	70.29	72.81	95.45	84.18	73.14	75.86	57.39	47.08	41.53	43.07
9	TOPJUDGE+EPM	96.01	83.68	74.77	77.26	95.86	86.21	78.67	81.23	58.11	48.20	44.30	45.07
10	MPBFN+EPM	95.81	83.36	74.61	76.39	95.62	86.34	77.34	79.35	57.53	50.04	40.46	42.01
11	LADAN+EPM	96.15	84.90	76.54	79.26	95.96	88.07	78.98	81.79	58.40	50.36	42.71	44.17
12	NeurJudge+EPM	96.20	85.16	77.83	78.21	94.77	89.75	77.46	80.19	57.81	49.36	41.77	43.79
13	TOPJUDGE+Event	95.93	83.55	73.03	75.86	95.82	86.34	77.20	80.29	58.21	47.73	44.36	45.00

Table 4: Comparisons with the SOTA models on CAIL-big.

- EPM (row 6) achieves the best results, outperforming the five SOTA models

NANJING UNIVERSITY

Results

		Law Article				Charge				Term of Penalty			
		Acc%	**MP%**	**MR%**	**F1%**	**Acc%**	**MP%**	**MR%**	**F1%**	**Acc%**	**MP%**	**MR%**	**F1%**
1	MLAC	94.90	79.06	66.91	69.41	94.72	83.42	72.38	75.62	56.43	46.87	40.43	41.89
2	TOPJUDGE	95.83	82.10	71.94	74.32	95.77	85.95	77.11	79.58	58.09	47.73	42.47	44.07
3	MBPFN	95.67	84.00	74.40	76.44	94.37	85.60	75.86	77.98	55.48	47.27	38.26	40.01
4	LADAN	95.78	84.93	75.88	78.79	94.58	85.52	77.36	80.04	56.34	47.76	40.48	42.02
5	NeurJudge	95.59	84.01	75.54	77.06	94.12	85.48	77.21	79.83	55.52	47.25	40.76	42.03
6	EPM	96.63	85.93	77.60	79.72	95.88	88.67	79.49	81.99	58.19	51.50	43.25	44.99
7	EPM@G	96.72	85.79	79.68	81.77	96.45	88.78	81.93	82.84	58.67	53.93	45.86	46.58
8	MLAC+EPM	95.50	79.71	70.29	72.81	95.45	84.18	73.14	75.86	57.39	47.08	41.53	43.07
9	TOPJUDGE+EPM	96.01	83.68	74.77	77.26	95.86	86.21	78.67	81.23	58.11	48.20	44.30	45.07
10	MPBFN+EPM	95.81	83.36	74.61	76.39	95.62	86.34	77.34	79.35	57.53	50.04	40.46	42.01
11	LADAN+EPM	96.15	84.90	76.54	79.26	95.96	88.07	78.98	81.79	58.40	50.36	42.71	44.17
12	NeurJudge+EPM	96.20	85.16	77.83	78.21	94.77	89.75	77.46	80.19	57.81	49.36	41.77	43.79
13	TOPJUDGE+Event	95.93	83.55	73.03	75.86	95.82	86.34	77.20	80.29	58.21	47.73	44.36	45.00

Table 4: Comparisons with the SOTA models on CAIL-big.

- EPM can improve the performance of the five SOTA models

南京大學 NANJING UNIVERSITY

Results

		Law Article				Charge				Term of Penalty			
		Acc%	**MP%**	**MR%**	**F1%**	**Acc%**	**MP%**	**MR%**	**F1%**	**Acc%**	**MP%**	**MR%**	**F1%**
1	MLAC	94.90	79.06	66.91	69.41	94.72	83.42	72.38	75.62	56.43	46.87	40.43	41.89
2	TOPJUDGE	95.83	82.10	71.94	74.32	95.77	85.95	77.11	79.58	58.09	47.73	42.47	44.07
3	MBPFN	95.67	84.00	74.40	76.44	94.37	85.60	75.86	77.98	55.48	47.27	38.26	40.01
4	LADAN	95.78	84.93	75.88	78.79	94.58	85.52	77.36	80.04	56.34	47.76	40.48	42.02
5	NeurJudge	95.59	84.01	75.54	77.06	94.12	85.48	77.21	79.83	55.52	47.25	40.76	42.03
6	EPM	96.63	85.93	77.60	79.72	95.88	88.67	79.49	81.99	58.19	51.50	43.25	44.99
7	EPM@G	96.72	85.79	79.68	81.77	96.45	88.78	81.93	82.84	58.67	53.93	45.86	46.58
8	MLAC+EPM	95.50	79.71	70.29	72.81	95.45	84.18	73.14	75.86	57.39	47.08	41.53	43.07
9	TOPJUDGE+EPM	96.01	83.68	74.77	77.26	95.86	86.21	78.67	81.23	58.11	48.20	44.30	45.07
10	MPBFN+EPM	95.81	83.36	74.61	76.39	95.62	86.34	77.34	79.35	57.53	50.04	40.46	42.01
11	LADAN+EPM	96.15	84.90	76.54	79.26	95.96	88.07	78.98	81.79	58.40	50.36	42.71	44.17
12	NeurJudge+EPM	96.20	85.16	77.83	78.21	94.77	89.75	77.46	80.19	57.81	49.36	41.77	43.79
13	TOPJUDGE+Event	95.93	83.55	73.03	75.86	95.82	86.34	77.20	80.29	58.21	47.73	44.36	45.00

Table 4: Comparisons with the SOTA models on CAIL-big.

- Better LJP results can be achieved by pre-train + fine-tune strategy rather than modifying the model to learn from event annotations.

Summary

- presented the first study on leveraging event extraction from case facts to solve LJP tasks
- defined a hierarchical event structure for legal cases
- collected a new LJP dataset with event annotations
- proposed a model that learns LJP and event extraction jointly subject to two kinds of constraints
- our model surpasses the existing SOTA models in performance

南京大學
NANJING UNIVERSITY

Thank you!

誠樸雄偉 勵學敦行