

Interpretable Charge Predictions for Criminal Cases: Learning to Generate Court Views from Fact Descriptions

¹Hai Ye*, ¹Xin Jiang*, ²Zhunchen Luo* and ¹Wenhan Chao
(* indicates equal contribution)

¹School of Computer Science and Engineering, Beihang University, Beijing, China
²Information Research Center of Military Science, Beijing, China

{*yehai, xinjiang, chaowenhan*}@buaa.edu.cn
zhunchenluo@gmail.com

June 4, 2018

Charge Prediction: overview

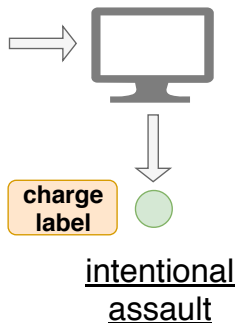
- **input:** *fact description* in a criminal case
- **output:** *charge label*, e.g. negligent homicide, drunk driving, intentional injury, etc.

Charge Prediction: overview

- **input:** *fact description* in a criminal case
- **output:** *charge label*, e.g. negligent homicide, drunk driving, intentional injury, etc.

Fact Description

After hearing, our court identified that at 23:00 on July 10,2009, the defendant Chen together with other eight or nine young men stopped Lee who was riding a motorcycle on street near the road in Xinliao town Xuwen County, after that the defendant Chen and the others beat Lee with steel pipe and knife. According to forensic identification, Lee suffered minor wound. ...



Charge Prediction: drawback

- lack of *interpretations* in charge determination
- we propose to study the problem of ***Court View Generation*** to relieve the above drawback

Court View Generation: overview

- *what is court view?*

Court View Generation: overview

- *what is court view?*
 - court view: the **written explanation** from judges to interpret the charge decision for certain criminal case and is also the core part in a legal document, consisting of **rationales** and **charge labels**.

Court View Generation: overview

- *what is court view?*
 - court view: the **written explanation** from judges to interpret the charge decision for certain criminal case and is also the core part in a legal document, consisting of **rationales** and **charge labels**.

court view

Our court hold that ***the defendant Chen ignored the state law and caused others minor wound with equipment together with others.*** His acts constituted the crime of ***intentional assault charge***

rationales

charge labels

Court View Generation: overview

- **input:** *fact description*
- **output:** *rationales* part of court view

Court View Generation: overview

- **input:** *fact description*
- **output:** *rationales* part of court view
 - charge labels can be obtained from charge prediction systems [Luo et al. 2017] or decided by human beings.

Court View Generation: overview

- **input:** *fact description*
- **output:** *rationales* part of court view
 - charge labels can be obtained from charge prediction systems [Luo et al. 2017] or decided by human beings.

Fact Description

After hearing, our court identified that at 23:00 on July 10,2009, the defendant Chen together with other eight or nine young men stopped Lee who was riding a motorcycle on street near the road in Xinliao townXuwen County, after that the defendant Chen and the others beat Lee with steel pipe and knife. According to forensic identification, Lee suffered minorwound. ...



Rationales

Our court hold that the defendant Chen ignored the state law and caused others minor wound with equipment together with others.

Court View Generation: overview

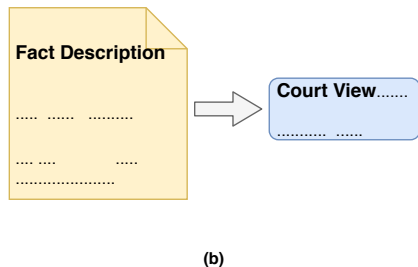
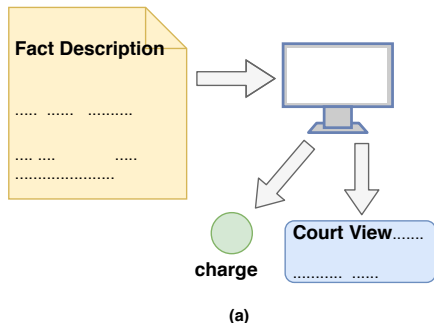
Court view refers to **rationales** in the rest of the presentation

Court View Generation: applications

- (a) **interpretability**: charge predictions can decide a charge for a case accompanying the rationales.
- (b) automatic **legal document generation**: court view part in a legal document.

Court View Generation: applications

- (a) **interpretability**: charge predictions can decide a charge for a case accompanying the rationales.
- (b) automatic **legal document generation**: court view part in a legal document.



Court View Generation: non-trivial

- what are *high-quality* rationales?

Court View Generation: non-trivial

- what are *high-quality* rationales?
 - 1) should contain *fact details* from the *fact description*
 - 2) should be *charge-discriminative* (or can also be called *charge-relevant*)

Court View Generation: non-trivial

- what are *high-quality* rationales?
 - 1) should contain *fact details* from the *fact description*

Court View Generation: non-trivial

- what are *high-quality* rationales?
 - 1) should contain *fact details* from the *fact description*
 - **fact details**: *activities* from the defendant which *break the law* or *important basis for charge determination*.

Court View Generation: non-trivial

- what are *high-quality* rationales?
 - 1) should contain *fact details* from the *fact description*
 - **fact details**: *activities* from the defendant which *break the law* or *important basis for charge determination*.

Fact Description: After hearing, our court identified that at 23:00 on July 10,2009, the defendant Chen together with other eight or nine young men stopped Lee who was riding a motorcycle on street near the road in Xinliao town Xuwen County, after that the defendant Chen and the others *beat Lee with steel pipe and knife*. According to forensic identification, Lee *suffered minor wound* ...

Rationales: Our court hold that the defendant Chen ignored the state law and *caused others minor wound with equipment* together with others. ...

Court View Generation: non-trivial

- what are *high-quality* rationales?
 - 2) should be *charge-discriminative*: with *deduced information*, e.g. killing motivation, which does not appear in fact descriptions

Court View Generation: non-trivial

- what are *high-quality* rationales?
 - 2) should be *charge-discriminative*: with *deduced information*, e.g. killing motivation, which does not appear in fact descriptions

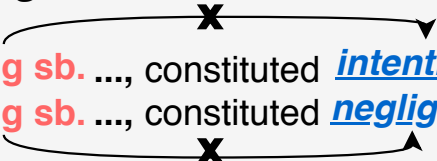
non-charge-discrimination:

-
- 1) ... **killing sb.** ..., constituted intentional homicide;
- 2) ... **killing sb.** ..., constituted negligent homicide;

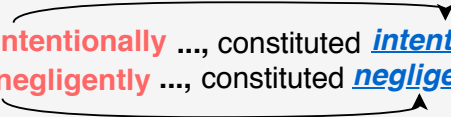
Court View Generation: non-trivial

- what are *high-quality* rationales?
 - 2) should be *charge-discriminative*: with *deduced information*, e.g. killing motivation, which does not appear in fact descriptions

non-charge-discrimination:

- 
- 1) ... **killing sb.** ..., constituted *intentional homicide*;
- 2) ... **killing sb.** ..., constituted *negligent homicide*;
- The diagram shows two curved arrows between the two sentences. The top arrow points from the first sentence to the second and has a large 'X' over it. The bottom arrow points from the second sentence to the first and also has a large 'X' over it, indicating that this is not charge-discriminative.

charge-discrimination:

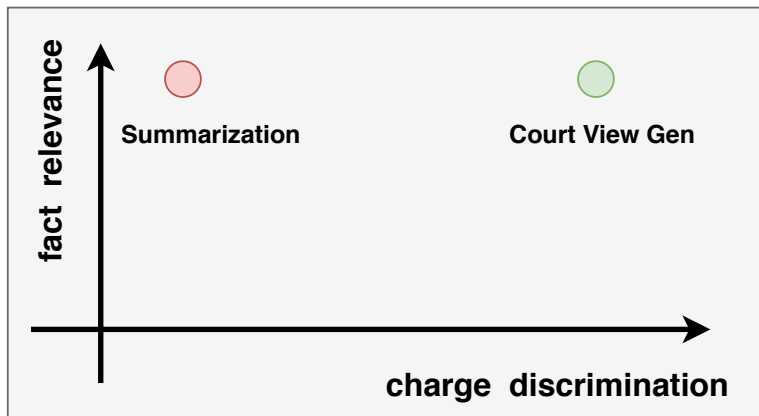
- 
- 1) ... **killing sb. intentionally** ..., constituted *intentional homicide*;
- 2) ... **killing sb. negligently** ..., constituted *negligent homicide*;
- The diagram shows two curved arrows between the two sentences. The top arrow points from the first sentence to the second. The bottom arrow points from the second sentence to the first. Both arrows are clear and do not have an 'X' over them, indicating that this is charge-discriminative.

Court View Generation: non-trivial

- what are *high-quality* rationales?
 - 2) should be *charge-discriminative*
 - **Court-View-Gen** \neq **Document Summarization**
 - **Rationales** = fact details + deduced information
 - **Summarization** = fact details

Court View Generation: non-trivial

- what are *high-quality* rationales?
 - 2) should be *charge-discriminative*
 - **Court-View-Gen** \neq **Document Summarization**
 - **Rationales** = fact details + deduced information
 - **Summarization** = fact details

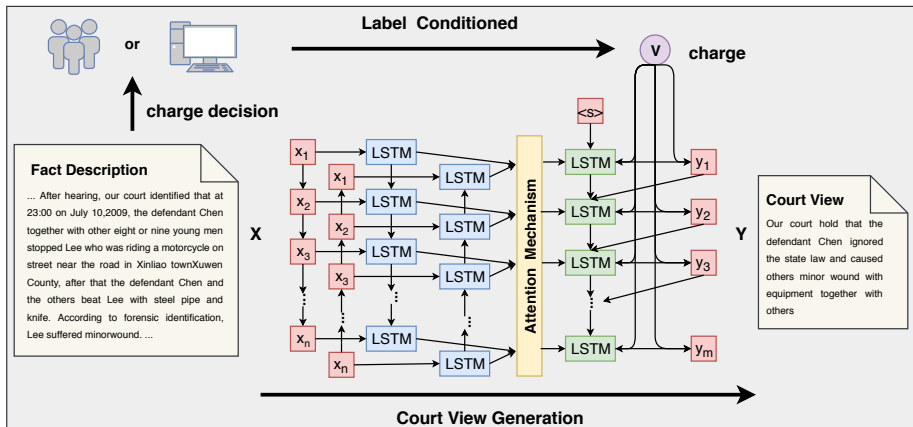


Model: Label-conditioned Sequence-to-Sequence Model

- *how* to generate *charge-discriminative* rationales with accurate *fact details*: enforce model to focus more charge-related information by encoding charge labels.

Model: Label-conditioned Sequence-to-Sequence Model

- how to generate *charge-discriminative* rationales with accurate *fact details*: enforce model to focus more charge-related information by encoding charge labels.



Model: Label-conditioned Sequence-to-Sequence Model

- The task of COURT-VIEW-GEN is to find rationale $\hat{\mathbf{y}}$ given fact description \mathbf{x} conditioned charge label v :

$$\hat{\mathbf{y}} = \arg \max_{\mathbf{y}} p(\mathbf{y} | \mathbf{x}, v) \quad (1)$$

$$p(\mathbf{y} | \mathbf{x}, v) = \prod_{i=1}^{|\mathbf{y}|} p(y_i | \mathbf{y}_{<i}, \mathbf{x}, v) \quad (2)$$

where $\mathbf{y}_i = y_1, \dots, y_{|y|}$. At time t , in the decoder, the probability to predict y_t is:

$$p(y_t | \mathbf{y}_{<t}, \mathbf{c}_t, v) = \text{softmax}(\mathbf{W}_1 \tanh(\mathbf{W}_0 [\mathbf{s}_t; \mathbf{c}_t; \mathbf{E}_{[v]}^v])) \quad (3)$$

where \mathbf{c}_t is the context vector merged by global attention mechanism; \mathbf{W}_1 and \mathbf{W}_0 are learnable parameters; \mathbf{s}_t is the hidden vector; \mathbf{E}^v is the charge label embedding matrix.

Model: Label-conditioned Sequence-to-Sequence Model

- We further embed the charge label v to highlight the computing of hidden state \mathbf{s}_t in the decoder:

$$\begin{aligned}\mathbf{s}_t &= \text{LSTM}_d(y_{t-1}, \mathbf{s}_{t-1}^v) \\ \mathbf{s}_{t-1}^v &= f_v(\mathbf{s}_{t-1}, v) \\ f_v &= \tanh(\mathbf{W}^v[\mathbf{s}_{t-1}; \mathbf{E}_{[v]}^v] + \mathbf{b}^v)\end{aligned}\tag{4}$$

Experiments: Data Preparation

- Following Luo et al. (2017), we construct dataset from the **published legal documents** in China Judgements Online.
 - The paragraph started with “our court identified that” is regarded as the **fact description**.
 - The part between “our court hold that” and the “charge” are regarded as the **rationales**.

Experiments: Data Preparation

- Following Luo et al. (2017), we construct dataset from the **published legal documents** in China Judgements Online.
 - The paragraph started with “our court identified that” is regarded as the **fact description**.
 - The part between “our court hold that” and the “charge” are regarded as the **rationales**.

# Training set	153,706
# Dev set	9,152
# Test set	9,123
Avg. # tokens in fact desc.	219.9
Avg. # tokens in rationales	30.6
Num. of # charge labels	51
# Dict. size in fact desc.	222,482
# Dict. size in rationales	21,305

Table: Statistics of our dataset.

Experiments: Comparisons with Baselines

● Baselines:

- **Rand** is to randomly select rationales in court views from the training set (**Rand_{all}**). We also randomly choose rationales from pools with same charge labels (**Rand_{charge}**).
- **BM25** is to index the fact description matching to the input fact description with highest BM25 score (Robertson and Walker, 1994) from the training set, and use its rationales as the result (**BM25_{f2f}**). Fact descriptions from pools with same charges are also retrieved (**BM25_{f2f+charge}**),
- **MOSES+** (Koehn et al., 2007) is a phrase based statistical machine translation system mapping fact descriptions to rationales.
- **NN-S2S** is the basic Seq2Seq model without attention (Sutskever et al., 2014) for machine translation. We set one LSTM layer for encoder and another one LSTM layer for decoder.
- **RAS⁺** is an attention based abstract summarization model (Chopra et al., 2016)). To deal with the much longer fact descriptions, we exploit the more advanced bidirectional LSTM model for the encoder instead of the simple convolutional model. Another LSTM model is set as the decoder coherent to Chopra et al. (2016).

Experiments: Comparisons with Baselines

- **Results:** Automatic Evaluation

MODEL (%)	AUTOMATIC EVALUATION			
	B-4	R-1	R-2	R-L
Rand _{all}	6.4	26.5	6.2	25.1
Rand _{charge}	24.9	53.6	29.1	49.3
BM25 _{f2f}	40.1	63.5	43.7	60.3
BM25 _{f2f+charge}	42.8	67.1	47.4	63.8
MOSES+	6.2	39.8	20.8	18.6
NN-S2S	38.4	65.5	45.1	62.2
RAS [†]	44.1**	69.1**	50.3**	65.9**
Ours	45.8	70.9	52.5	67.7

Table: Results of automatic evaluation with BLEU-4 and full length of F1 scores of variant Rouges. Best results are labeled as boldface. Statistical significance is indicated with $** (p < 0.01)$ and $* (p < 0.05)$ comparing to our full model.

Experiments: Comparisons with Baselines

● **Results:** Human Evaluation

- 1) how *fluent* of the rationales in court view is
- 2) how *accurate* of the rationales are (how many fact details have been accurately expressed)
 - 5 scales for both fluent and accurate evaluation (5 is for the best)
- 3) whether rationales can be adopted for use in comprehensive evaluation (*adoptable*)
- three raters are asked to conduct evaluation

Experiments: Comparisons with Baselines

● Results: Human Evaluation

- 1) how *fluent* of the rationales in court view is
- 2) how *accurate* of the rationales are (how many fact details have been accurately expressed)
 - 5 scales for both fluent and accurate evaluation (5 is for the best)
- 3) whether rationales can be adopted for use in comprehensive evaluation (*adoptable*)
- three raters are asked to conduct evaluation

MODEL	HUMAN JUDGEMENT		
	FLUENT	ACC.	ADOPT.(%)
BM25 _{f2f}	4.95	3.66**	0.47**
BM25 _{f2f+charge}	4.94	3.90**	0.50**
MOSES+	1.39**	1.31**	0**
NN-S2S	4.97	4.07**	0.62*
RAS [†]	4.96	4.25*	0.64*
Ours	4.93	4.54	0.72

Further Analysis: Impact of Exploiting Charge Labels

• Charge-discriminations Analysis.

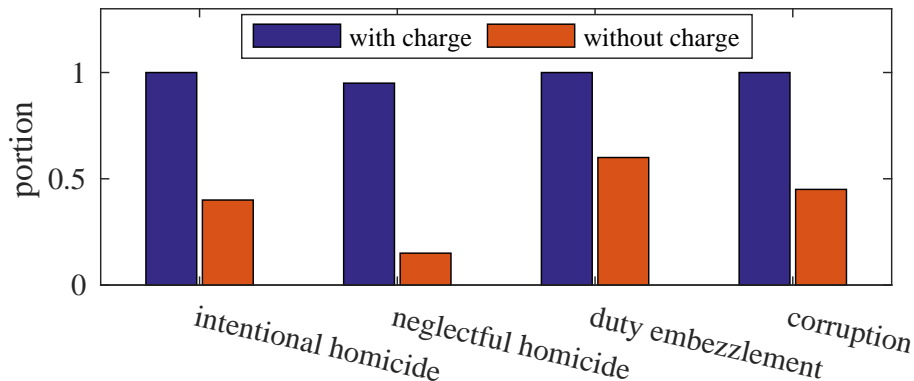


Figure: Portions of charge-discriminative rationales in court views for every charge with 20 candidates.

Further Analysis: Analysis through Cases

- Fake Charge Label Conditioned Study:
 - fake charge: the wrong charge label.
- Case Study.

MODEL	[CHARGE]	GENERATED COURT VIEWS CONDITIONED ON FAKE CHARGE LABEL
Gold	[故意伤害罪] PP 故意 伤害 他人 身体, 致 一人 轻伤。# [intentional injury] PP intentionally injured others body, caused one person slight injury .	
	[寻衅滋事罪] PP 随意 殴打 他人, 致 一人 轻伤, 情节 恶劣。# [defiance and affray crime] PP beat others at will, caused one person slight injury .	
Ours	[故意杀人罪] PP 故意 非法 剥夺 他人 生命, 致 一人 轻伤。# [intentional homicide] PP intentionally illegally deprived someone of life, caused one person slight injury .	
	[过失致人死亡罪] PP 过失 致 一人 轻伤。# [neglectful homicide] PP neglectfully caused one person slight injury .	
MODEL	[CHARGE]	GENERATED COURT VIEWS
Gold	[交通肇事罪] PP 违反 交通 运输 管理 法规, 造成 一人 死亡, 二人 受伤 的 交通 事故, 负 事故 的 全部 责任。# [traffic accident crime] PP violated traffic transportation management regulations, caused one person dead, two people injured, take accident's full responsibility .	
	[过失致人死亡罪] PP 在 驾驶 机动车 过程 中, 疏忽 大意, 致使 他人 被 碾压 致 死。# [negligent homicide] PP when driving car, being neglectful, caused one person dead by rolling .	
Ours	[盗窃罪] PP 以 非法 占有 为 目的, 伙同 他人 多次 秘密 窃取 公民 财物, 数额 较大。# [larceny] PP in intention of illegal possession, ganged up with others and stole goods secretly in relatively large amount for several times .	
	PP 违反 交通 运输 管理 法规, 发生 交通 事故, 致 一人 死亡, 二人 受伤, 负 事故 的 全部 责任。# PP violated traffic transportation management regulations, caused traffic accident, caused one person dead, two people injured, take accident's full responsibility . ✓	
Ours _c	PP 因 疏忽 大意 致 一人 死亡。# PP neglectfully caused one person dead . ✓	
	PP 以 非法 占有 为 目的, 结伙 他人 秘密 窃取 他人 财物, 数额 较大。# PP in intention of illegal possession, ganged up with others and stole goods secretly in relatively large amount . ✗	
BM25 _{r2f+c}	PP 违反 交通 运输 管理 法规, 发生 重大 交通 事故, 致 一人 死亡, 负 事故 的 全部 责任。# PP violated traffic transportation management regulations, caused severe traffic accident, caused one person dead, took accident's full responsibility ✗	
	PP 违反 交通 运输 管理 法规, 发生 重大 交通 事故, 致 一人 死亡, 负 事故 的 全部 责任。# PP violated traffic transportation management regulations, caused severe traffic accident, caused one person dead, took accident's full responsibility . ✗	
BM25 _{r2f+c}	PP 以 非法 占有 为 目的, 秘密 窃取 他人 财物, 数额 较大。# PP in intention of illegal possession, stole goods secretly in relatively large amount . ✗	
	PP 违反 道路 交通 运输 管理 法规, 致 一人 死亡 且 负 事故 主要 责任。# PP violated road traffic transportation management regulations, caused one person dead, took accident's main responsibility . ✗	
BM25 _{r2f+c}	PP 驾驶 车辆 过程 中 疏忽 大意, 过失 致 一人 死亡。# PP when driving, neglectfully caused one person dead . ✓	
	PP 以 非法 占有 为 目的, 秘密 窃取 公民 财物。# PP in intention of possession, stole goods secretly . ✗	

- Our dataset in our paper can be obtained from <https://github.com/oceanpypt/Court-View-Gen>.

Any questions ?

The End

Thank you all !