

Circumstances enhanced Criminal Court View Generation

Linan Yue¹, Qi Liu^{1,*}, Han Wu¹, Yanqing An¹, Li Wang¹, Senchao Yuan¹, Dayong Wu²

¹Anhui Province Key Laboratory of Big Data Analysis and Application, School of Data Science & School of Computer Science and Technology, University of Science and Technology of China,

{lnyue, wuhanhan, anyq, wl063, yuansc}@mail.ustc.edu.cn; {qiliuqj}@ustc.edu.cn;

²IFLYTEK, {dywu2}@iflytek.com

ABSTRACT

Criminal Court View Generation is an essential task in legal intelligence, which aims to automatically generate sentences interpreting judgment results. The court view could be seen as the summary of crime circumstances in a case, including ADjudging Circumstance (ADC) and SEntencing Circumstance (SEC). However, different circumstances vary widely, and adopting them to generate court views directly may limit the generation performance. Therefore, it is necessary to identify the ADC and SEC related sentences in case facts and enhance them into the court view generation, respectively. To this end, in this paper, we propose a novel Circumstances enhanced Criminal Court View Generation (C3VG) method, consisting of the extraction and generation stage. Specifically, in the extraction stage, we design a *Circumstances Selector* to select ADC and SEC related sentences. After that, we apply them to two generators to generate the circumstances enhanced court views, respectively. After merging the two types of court views, we could obtain the final court views. We evaluate C3VG by conducting extensive experiments on a real-world dataset and experimental results clearly validate the effectiveness of our proposed model.

CCS CONCEPTS

• Applied computing → Law.

KEYWORDS

Court views; Crime circumstances; Text generation

ACM Reference Format:

Linan Yue, Qi Liu, Han Wu, Yanqing An, Li Wang, Senchao Yuan and Dayong Wu. 2021. Circumstances enhanced Criminal Court View Generation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), July 11–15, 2021, Virtual Event, Canada*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3404835.3462984>

1 INTRODUCTION

Court views in criminal cases are an important portion of legal judgment, which are written to interpret verdicts and sentencing by judges. Generating court views automatically could assist judges

* Corresponding Author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGIR '21, July 11–15, 2021, Virtual Event, Canada

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8037-9/21/07...\$15.00

<https://doi.org/10.1145/3404835.3462984>

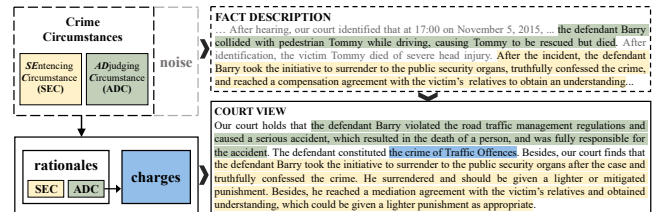


Figure 1: An example of fact description and the court view from a legal document in a case.

in reducing stress and releasing much human labor when they deal with vast cases [13].

In the judicial scenarios, the criminal court views consist of rationales and charges, where charges are derived and explained by the rationales [11, 13]. Among them, rationales could be regarded as the summary of case facts, or more accurately, of the crime circumstances. Crime circumstances are the key elements in cases facts for determining judgment results [14], which consist of ADjudging Circumstance (ADC) and SEntencing Circumstance (SEC). ADC is a part of fact description which indicates that a certain act meets the requirements of the charge. SEC is the partial fact that should be considered when sentencing. An example of fact description and the court view is shown in Figure 1. Specifically, the fact description consists of ADC, SEC, and redundant circumstances (i.e., noise) which do not affect the determination of court views such as “time”. Rationales and the corresponding charge in court views are both related to the crime circumstances. Among them, the green highlighted sentences in **FACT DESCRIPTION** and **COURT VIEW** represent the ADC related sentences in fact and rationales of ADC, respectively. And the yellow highlighted are relevant to SEC. The blue sentence in **COURT VIEW** represents the corresponding charge (i.e., the crime of Traffic Offences) which is supported by rationales of ADC. From this example, we could observe that criminal court view generation should be based on both ADC and SEC.

In the literature, a great amount of progress has been made in this area [4, 12, 13]. Ye et al. [13] investigated the problem of court view generation for the first time and explored the charge to generate views. Huang et al. [4] combined the language model and question answering based method to generate reasonable court views. However, these methods generated court views on the basis of the whole case facts. Actually, from the previous analysis, we could find the court view is generated based on the crime circumstances. Therefore, in this paper, we aim to generate court views by enhancing crime circumstances into the process of generation.

However, it is a non-trivial problem. According to the definition and function of ADC and SEC, the two circumstances vary widely [14]. Thus, directly applying these circumstances to the task of

court view generation is unsuitable, and may bring a lot of noise to generation. Take the case in Figure 1 as an example, as there exists no relationship between the green highlighted texts (i.e., ADC) and the yellow ones (i.e., SEC) in the case fact, ADC related facts may noise the rationales of SEC generation. Therefore, it is significant to design a strategy that could precisely identify ADC/SEC sentences in case facts and adopt them for generation, respectively.

To tackle the challenges mentioned above, we propose a Circumstance enhanced Criminal Court View Generation (C3VG) method, following a two-stage architecture which is from extraction to generation. Specifically, in the extraction stage, we explore a *Circumstances Selector* to select ADC related facts and SEC related ones. We define this problem as a sentence-level sequence tagging task. Next, in the generation stage, based on the two types of selected sentences, we implement two generators (i.e., the ADC enhanced generator and the SEC enhanced one) to generate corresponding charges and circumstance specific rationales, respectively. After merging views from the two generators, we could obtain the final court views. Finally, we validate the effectiveness of our method with extensive experiments on real legal documents.

2 CIRCUMSTANCE ENHANCED CRIMINAL COURT VIEW GENERATION

2.1 Problem Definition

Here, we formulate our problem with the following definition:

Fact Description is a descriptive document which consists of ADC, SEC and other redundant circumstances (denoted by noise). Here, we denote the case fact $f = (adc, sec, noise)$, where $adc = \{w_1^a, \dots, w_{l_a}^a\}$, $sec = \{w_1^s, \dots, w_{l_s}^s\}$, and $noise = \{w_1^n, \dots, w_{l_n}^n\}$, and supposing we already label which sentences are related to SEC. For example, as the case shown in Figure 1, the yellow highlighted texts in fact are labeled as SEC related sentences denoting by 1 and other sentences by 0.

Corresponding Charge is denoted as $c \in Y_c$, where Y_c represents a collection of charges (e.g., *robbery* and *traffic offences*). The corresponding charge is determined based on ADC related facts.

Rationales contain two components which are denoted by v_a and v_s , where $v_a = \{y_1^{v_a}, \dots, y_{l_{v_a}}^{v_a}\}$ represents rationales of ADC, and $v_s = \{y_1^{v_s}, \dots, y_{l_{v_s}}^{v_s}\}$ denotes the rationales of SEC.

Criminal Court View denoted by v consists of the corresponding charge and rationales. Besides, as the court view is a continuous sentence, we design a “template” function $F(\cdot)$ which maps v_a , v_s , and c into a fixed pattern (i.e., “Our court holds that [**Rationales of ADC**]. The defendant constituted [**Corresponding Charge**]. Besides, our court finds that [**Rationales of SEC**].”). The court view in Figure 1 is an exemplar.

Then, our task could be denoted as generating the court view $v = F(v_a, c, v_s)$ based on the whole case fact $f = (adc, sec, noise)$.

2.2 Model Assumption

In this section, we first explain the reason why we could extract different circumstances and adopt them to generate the court view. For a certain case, given the case fact f , using Bayes rule, we would

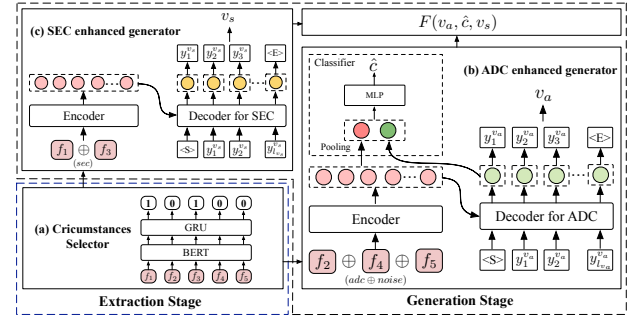


Figure 2: Architecture of C3VG which consists of Extraction Stage and Generation Stage.

train the model to generate the court view v as $P(v | f) = P(v | adc, sec, noise)$.

From the previous analysis, we could know that ADC, SEC and other redundant circumstances (i.e., noise) do not affect each other [14]. Therefore, we could assume that the three types circumstances are independent of each other and the probability of predicted v could be calculated as :

$$\begin{aligned} P(v | f) &= P(F(v_a, c, v_s) | adc, sec, noise) \\ &= P(v_a, c | adc) P(v_s | sec), \end{aligned} \quad (1)$$

where we remove *noise* because redundant circumstances do not affect court view generation. However, in the practice, it is hard to remove these circumstances. Therefore, the probability of v could be approximated as $P(v | f) \approx P(v_a, c | adc, noise) P(v_s | sec) = P(v_a, c | adc) P(v_s | sec, noise)$. Besides, as we have supposed which sentences are related to SEC, we train the generation model as $P(v | f) \approx P(v_a, c | adc, noise) P(v_s | sec)$. Based on this assumption, we design a model which is from extraction to generation.

2.3 Model Architecture

As shown in Figure 2, our proposed Circumstance enhanced Criminal Court View Generation (C3VG) method consists of the extraction and generation stage. Specifically, in the extraction stage, we extract the SEC related and other fact sentences through a *Circumstances Selector*. Then, based on the selected circumstances, we design two types of generators to obtain different components of court views in the generation stage, respectively. Finally, we merge the two components into $F(\cdot)$ to generate the final court view.

2.3.1 Extraction Stage. In this stage, we apply a *Circumstances Selector* to select SEC related facts from the fact descriptions. After selecting, we could divide the case fact into two components which consist of the SEC related sentences (*sec*) and other sentences ($adc \oplus noise$) including the ADC and redundant circumstances.

Our *Circumstances Selector* treats this extraction problem as a sentence-level sequence tagging task. In detail, given the case fact f , we first break the whole fact as sentences (i.e., f_1, \dots, f_n), where n is the number of sentences in the fact. Then, we adopt BERT [3] to learn good representations of these sentences (i.e., $h_i = \text{BERT}(f_i)$, where $h_i \in \mathbb{R}^{d_s}$ is the representation of the [CLS] token and could be seen as the hidden states of a certain sentence f_i). Next, we stack all h_i by column and obtain the fact representations $H = \{h_1, \dots, h_n\} \in \mathbb{R}^{n \times d_s}$. To further enhance the interaction among sentences, we pass H into a standard GRU [2] model

to obtain final sentences representations $H' = \{h'_1, \dots, h'_n\}$. We could then calculate the probability that the sentence is selected as $\sigma(W_s h'_i + b_s)$ with trainable parameters W_s and b_s , where $\sigma(\cdot)$ denotes an activation function. If the probability exceeds a threshold τ , we select it as the SEC related sentence. Finally, we connect selected SEC sentences together denoted by (*sec*), and concat other unselected sentences as (*adc* \oplus *noise*).

2.3.2 Generation Stage. After obtaining selected sentences, we adopt them to generate court views. First, to generate corresponding charges and rationales of ADC, we adopt an *ADC enhanced generator* which consists of *Encoder*, *Decoder for ADC* and *Classifier*, as shown in Figure 2(b). Specifically, we utilize a classical seq2seq model with attention to generate rationales of ADC. Besides, we design an auxiliary task (i.e., charge prediction) to alleviate the negative impact of noise and obtain the corresponding charge.

Encoder. Given the selected sentences (*adc* \oplus *noise*) which are sequences of words, the encoder transforms the words to embeddings. Then, the embedding sequences are fed to the Bi-GRU [2] to obtain the hidden states $H^a = \{h_1^a, \dots, h_{l_a+l_n}^a\}$.

Decoder for ADC. The decoder in our model is a GRU which is used to generate words, and at each step t , the probability to predict y_t^{va} is computed as follows:

$$P(y_t^{va} | \text{adc} \oplus \text{noise}) = \text{softmax}(W_1 \tanh(g_t^a)), \quad (2)$$

where $g_t^a = W_0[s_t^a; c_t^a]$. Among them, W_0 and W_1 are learnable parameters, s_t^a is the hidden state of decoder at time t , and c_t^a is the context vector generated from the encoder with an attention mechanism and could be calculated:

$$c_t^a = \sum_{j=1}^{l_a+l_n} \alpha_{tj} h_j^a, \quad \alpha_{tj} = \frac{\exp(s_t^{aT} W_2 h_j^a)}{\sum_{k=1}^{l_a+l_n} \exp(s_t^{aT} W_2 h_k^a)}. \quad (3)$$

Classifier. As noise sentences in selected sentences (*adc* \oplus *noise*) are hard to recognize and may impact rationales of ADC generation, we design a charge prediction task to alleviate this problem. In practical, redundant circumstances would not distract the charge determination. Therefore, while generating rationales, the charge prediction could help the model focus on more information of ADC and guide the generation. Specifically, we first apply the encoder output H^a and the decoder output $G^a = \{g_1^a, \dots, g_{l_{va}}^a\}$ into a mean-pooling operation to obtain h^a and g^a , respectively. Besides, to connect the prediction and generation tasks together, we concat h^a and g^a , and adopt them to predict the charge \hat{c} :

$$\hat{c} = \arg \max_{i=1, \dots, |Y_c|} \hat{y}_{c_i}, \quad \hat{y}_c = \text{softmax}(W_c[h^a; g^a] + b_c). \quad (4)$$

In addition, to generate rationales of SEC (i.e., v_s), we utilize a *SEC enhanced generator* which follows the same seq2seq architecture as the *ADC enhanced generator*. Finally, we utilize the ‘‘template’’ function $F(\cdot)$ to generate final court views $F(v_a, \hat{c}, v_s)$.

2.3.3 Model Training. In the extraction stage, we adopt the binary cross-entropy as the loss:

$$\mathcal{L}_{ext} = \sum_{i=1}^n -y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i). \quad (5)$$

Table 1: The statistics of datasets.

| Models Stage | Training Set | Test Set | Avg. fact Length | Avg. rationale Length | Charge label Number |
|--------------|--------------|----------|------------------|-----------------------|---------------------|
| Extraction | 8,000 | 2,000 | 223.6 | — | — |
| Generation | 50,312 | 12,627 | 221.1 | 68.9 | 62 |

In the generation stage, for the classifier in the ADC enhanced generator, we use the cross-entropy as loss function (i.e., $\mathcal{L}_1 = -\sum_{k=1}^{|Y_c|} y_k \log(\hat{y}_k)$), where $|Y_c|$ denotes the number of charge labels. For decoders in both ADC and SEC enhanced generator, we adopt the common negative log-likelihood (NLL) as the loss function (i.e., $\mathcal{L}_2 = \frac{1}{T} \sum_{t=1}^T -\log P(y_t^*)$), where T is the length of real court rationales. Then, based on the two loss functions, the total loss in the generation stage could be calculated as $\mathcal{L}_{gen} = \mathcal{L}_1 + \mathcal{L}_2$.

3 EXPERIMENTS

3.1 Data Description

Although Ye et al. [13] released a real-world dataset for criminal court view generation, the ground truth is only about charges and rationales of ADC, but ignores the rationales of SEC. Therefore, we construct a new dataset CJO¹ from the published legal documents in China Judgments Online². As the documents are well-structured and human-annotated, we could easily extract fact descriptions, corresponding charges, and rationales of ADC and SEC using regular expressions. Besides, in this work, we only focus on the cases with one charge, leaving the complex cases which have multiple charges for future study. We collect 72,939 cases and divide them into two parts for the extraction and generation stage.

As there is no supervised data for *Circumstances Selector* in the extraction stage, we generate the corresponding training data by aligning the court view to the fact. Specifically, after deeply analyzing the relationship between facts and rationales, we find that SEC related facts have a high similarity with rationales of SEC while the ADC does not have this obvious relationship. Therefore, we break the whole fact and rationales of SEC into sentences, and calculate a matching score for a fact sentence and rationales of SEC. Each sentence corresponds to a group of matching scores, and if any one of them exceeds a predefined value 0.6, we recognize the sentence as a SEC related fact and denote it as 1, otherwise mark it as 0. The detailed statistics of the datasets are shown in Table 1.

3.2 Comparison methods

- **AttS2S** [1] is a seq2seq model with attention mechanisms. It is a classic model for the many generation tasks[5, 6].
- **PGN** [7] is designed to solve the out of vocabulary (OOV) problem in the text generation by adopting a pointer network.
- **Transformer** [8] is the state-of-the-art model in several text generation tasks.
- **Label-AttS2S** [13] is a seq2seq model which enhances the charge label into rationales of ADC generation. Here, we expand this method into both rationales of ADC and SEC generation.
- **C3VGw/oC** ablates the classifier for predicting charges and trains the remaining model as a text generation task.
- **C3VGw/oE** removes the extraction stage and adopt the whole case fact to generate rationales and predict charges.

¹<https://github.com/bigdata-ustc/C3VG>

²<https://wenshu.court.gov.cn>

Table 2: Results of court view generation and the charge prediction.

| Models | Court View Generation | | | | | | Charge Prediction | |
|--------------|-----------------------|-------------|-------------|-------------|-------------|-------------|-------------------|-------------|
| | ROUGE | | | BLEU | | | Acc MF1 | |
| | R-1 | R-2 | R-L | B-1 | B-2 | B-N | | |
| AttS2S | 58.7 | 38.9 | 59.4 | 50.5 | 41.0 | 38.0 | — | — |
| PGN | 59.3 | 37.0 | 59.8 | 50.2 | 39.6 | 36.7 | — | — |
| Transformer | 59.9 | 39.6 | 60.9 | 50.8 | 41.3 | 38.1 | — | — |
| Label-AttS2S | 47.0 | 31.4 | 52.8 | 38.7 | 31.6 | 29.4 | — | — |
| C3VGw/oC | 59.4 | 37.2 | 59.5 | 50.3 | 39.8 | 36.9 | — | — |
| C3VGw/oE | 49.9 | 30.8 | 50.5 | 42.7 | 33.3 | 30.7 | 87.9 | 52.3 |
| C3VG | 60.1 | 40.5 | 62.5 | 52.1 | 43.5 | 40.6 | 90.6 | 71.2 |

3.3 Experimental Setup

In the extraction stage, we adopt the BERT to obtain sentences representations and set the maximum sentence length to 150. After that, we employ Macro-F1 and Micro-F1 as the evaluation metrics for the extraction model. In the generation stage, we set all GRU models with hidden state size to 150. For *Classifier*, we adopt accuracy (Acc) and Macro-F1 as evaluation metrics. Besides, to evaluate the effectiveness of generation, we report ROUGE F1 scores³ and BLEU scores⁴ in the following experiments. After that, we set the beam search size to 5 when generating court views in inference.

3.4 Experimental Results

3.4.1 Results of court view generation and the charge prediction. Table 2 reports the results on generating court views and predicting charges. From the court view generation side, there are several observations. First, compared with the classic text generation model AttS2S, PGN and Transformer, our C3VG performs better, which demonstrates adopting crime circumstances to generate court views is effective. Second, Label-AttS2S designed for court view generation performs not very well. This is probably because it utilizes the practical charge label to guide both rationales of ADC and SEC generation. And the charge label might noise the rationales of SEC generation and further impact the whole performance. Similarly, C3VGw/oE has a poor performance probably because the charge classifier takes the whole fact as inputs and has a bad impact on generating rationales of SEC. Simultaneously, the performance gap between C3VGw/oE and C3VG illustrates it is necessary to select crime circumstances before generation. Last, the gap between C3VGw/oC and C3VG demonstrates the effectiveness of our proposed charge classifier which could alleviate the negative impact of noise for rationales of ADC generation.

From the charge prediction side, C3VG outperforms C3VGw/oE which indicates utilizing ADC related facts to predict the corresponding charge is more effective than adopting the whole fact. This observation fits the practical judicial scenarios where the charge is determined on ADC related facts.

3.4.2 Results of Circumstances Selector. After statistics, the Macro-F1 and Micro-F1 in the extraction stage are 91.1% and 94.1%, respectively, which illustrates the *Circumstances Selector* could well extract the SEC related sentences. Besides, we design a variant denoted by C3VG_gold which ignores the *Circumstances Selector* and

³We adopt the results of ROUGE-1, ROUGE-2, and ROUGE-L.

⁴We calculate BLEU-1, BLEU-2 and BLEU-N as evaluation metrics. Among them, BLEU-N is the average of BLEU-1, BLEU2, BLEU-3 and BLEU-4.

adopts the labeled circumstances to the generation stage. The average results of ROUGE and BLEU are 56.4% and 42.4%, which only surpass C3VG by 2.0% and 1.8%, respectively. It also demonstrates the effectiveness of *Circumstances Selector*.

3.4.3 Performance by the length of court views. We investigate the model performance by the length of court views via comparing with C3VG and AttS2S on ROUGE-L and BLEU-4 scores. As shown in Figure 3, we find the model performance drops when the length of court views increase. Besides, C3VG performs better than AttS2S no matter how long the length, which indicates C3VG could generate both short and long court views well. This observation illustrates that adopting two types of generators to generate court views is effective.

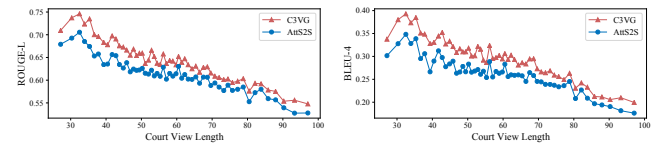


Figure 3: Model performance by the length of court views.

4 ETHICAL DISCUSSION

From the above observation on experiments, our C3VG could generate court views effectively following a two-stage architecture. However, there still exists an ethical consideration needed to be discussed. The goal of our method is to help judges reduce stress by generating court views automatically, which is not designed to replace judges. Judges need to proofread the generated views[11].

5 RELATED WORK

Court View Generation. In recent years, Legal Intelligence has addressed much attention and achieved great progress [9–11, 13, 15]. Among them, Court View Generation has been widely studied. From the civil cases side, Wu et al. [11] proposed counterfactual decoders to generate judgment-discriminative court views, which eliminated the bias in civil datasets. And from the criminal cases side, Ye et al. [13] first explored the court view generation in criminal cases and utilized the charge to generate charge-discriminative court views. Huang et al. [4] investigated a QA-based method to generate syntactically correct and logically reasonable court views.

6 CONCLUSIONS

In this paper, we proposed a Circumstance enhanced Criminal Court View Generation (C3VG) method which followed a two-stage architecture from extraction to generation. To be specific, C3VG adopted a *Circumstances Selector* to select ADC and SEC related facts in the extraction stage. Then, it employed ADC and SEC enhanced generators to generate the corresponding court views, respectively. Finally, it merged the two types of views together with a “template” function to obtain final criminal court views. Experimental results demonstrated the effectiveness of C3VG.

Acknowledgements. This research was partially supported by grants from the National Key Research and Development Program of China (No. 2018YFC0832101), and the National Natural Science Foundation of China (Grants No. 61922073 and U20A20229). Qi Liu acknowledges the support of the Youth Innovation Promotion Association of CAS (No. 2014299).

REFERENCES

- [1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *Proceedings of International Conference on Learning Representations (ICLR)* (2014).
- [2] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on EMNLP*.
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. [n.d.]. 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 (NAACL)*.
- [4] Weijing Huang, Xianfeng Liao, Zhiqiang Xie, Jiang Qian, Bojin Zhuang, Shaojun Wang, and Jing Xiao. 2020. Generating Reasonable Legal Text through the Combination of Language Modeling and Question Answering. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*.
- [5] Zhenya Huang, Qi Liu, Weibo Gao, Jinze Wu, Yu Yin, Hao Wang, and Enhong Chen. 2020. Neural mathematical solver with enhanced formula structure. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1729–1732.
- [6] Xin Lin, Zhenya Huang, Hongke Zhao, Enhong Chen, Qi Liu, Hao Wang, and Shijin Wang. 2021. HMS: A Hierarchical Solver with Dependency-Enhanced Understanding for Math Word Problem. (2021).
- [7] 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)*.
- [8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. [n.d.]. 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*.
- [9] Pengfei Wang, Yu Fan, Shuzi Niu, Ze Yang, Yongfeng Zhang, and Jiafeng Guo. 2019. Hierarchical matching network for crime classification. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*.
- [10] Pengfei Wang, Ze Yang, Shuzi Niu, Yongfeng Zhang, Lei Zhang, and ShaoZhang Niu. 2018. Modeling dynamic pairwise attention for crime classification over legal articles. In *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*. 485–494.
- [11] 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 EMNLP*.
- [12] Xiaotong Yang, Guozhen Shi, Jiapeng Lou, Shubei Wang, and Zichen Guo. 2019. Interpretable Charge Prediction with Multi-Perspective Jointly Learning Model. In *2019 IEEE 5th International Conference on Computer and Communications*.
- [13] Hai Ye, Xin Jiang, Zhunchen Luo, and Wenhan Chao. 2018. Interpretable Charge Predictions for Criminal Cases: Learning to Generate Court Views from Fact Descriptions. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*.
- [14] Mingkai Zhang et al. 2003. *Criminal Law*. Number 4. Law Press-China. 490–492,502–513 pages.
- [15] Xin Zhou, Yating Zhang, Xiaozhong Liu, Changlong Sun, and Luo Si. 2019. Legal Intelligence for E-commerce: Multi-task Learning by Leveraging Multiview Dispute Representation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 315–324.