Few-Shot Legal Knowledge Question Answering System for COVID-19 Epidemic

Jiaye Wu
Jie Liu
Xudong Luo*
Guangxi Key Lab of Multi-Source Information Mining & Security
College of Computer Science and Information Engineering
Guangxi Normal University
Guilin, Guangxi, China
luoxd@mailbox.gxnu.edu.cn

ABSTRACT

Recently, the pandemic caused by COVID-19 is severe in the entire world. The prevention and control of crimes associated with COVID-19 are critical for controlling the pandemic. Therefore, to provide efficient and convenient intelligent legal knowledge services during the pandemic, in this paper, we develop an intelligent system for answering legal questions on the WeChat platform. The data source we used for training our system is "The typical cases of national procuratorial authorities handling crimes against the prevention and control of the new coronary pneumonia pandemic following the law", which is published online by the Supreme People's Procuratorate of the People's Republic of China. We base our system on BERT (a well-known pre-trained language model) and use the shared attention mechanism to capture the text information further. Then we train a model to minimise the contrastive loss. Finally, the system uses the trained model to identify the information entered by a user, and accordingly responds to the user with a reference case similar to the query case and give the reference legal gist applicable to the query case.

CCS CONCEPTS

• **Computing methodologies** → *Artificial intelligence*; **Natural language processing**; **Discourse, dialogue and pragmatics**;

KEYWORDS

question answering system, pre-trained model, neural network

ACM Reference Format:
Jiaye Wu, Jie Liu, and Xudong Luo. 2020. Few-Shot Legal Knowledge Question Answering System for COVID-19 Epidemic. In *2020 3rd International Conference on Algorithms, Computing and Artificial Intelligence (ACAI 2020), December 24–26, 2020, Sanya, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3446132.3446409

*This is corresponding author.

ACAI 2020, December 24–26, 2020, Sanya, China

ACM ISBN 978-1-4503-8811-5/20/12…$15.00
https://doi.org/10.1145/3446132.3446409

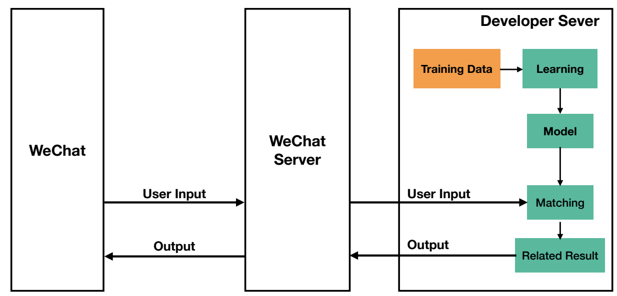


Figure 1: Flow chart of the system.

1 INTRODUCTION

A recent series of pneumonia outbreaks caused by a novel coronavirus have spread across China and the rest of the world. During the pandemic, many criminals took the opportunity to break the law and commit crimes, endangering people's health and safety, while also challenging to the prevention and control of the pandemic by the inspecting authorities. Given the unprecedented outbreak of the pandemic and its wide-ranging impact, the inspecting authorities need to be case-specific when confronted with cases against the prevention and control of the pandemic at the beginning. With the increasing number of criminal cases, this poses a significant challenge to the work of the authorities. On the other hand, people's knowledge of the laws related to the pandemic is also lacking. In the field of artificial intelligence, there has been a lot of research work that has explored the possibility of AI tools to aid legal judgments. Moreover, such tools can not only make legal work more efficient but can also help ordinary people with no background in legal.

For example, both[20] and [21] propose a case-based reasoning system on legal document, which can be used to solve real world problems. And Liu and Hsieh [14] use the K-nearest neighbour as a classifier to process Chinese judicial documents. With the success of deep learning in natural language processing, deep neural networks have also been used to predict guilties in legal cases [10]. Many researchers studied the application of these methods in the legal field with significant results. However, few attempts have been made to apply these methods to the legal cases involved in the COVID-19 pandemic and to the development of an operational system.

To date, the lack of open, high-quality data on legal cases involving the pandemic, the diversity of charges, and the complexity of cases make achieving a reliable question-and-answer model a challenging task. To the end, in this paper, we try to use the small amount of available data to implement a WeChat-based question-answering system on legal knowledge (see Figure 1). The system can identify similar cases against the pandemic, and then give the corresponding reference to similar cases and reference to the legal gist based on user input. The system evaluates the performance of the system through comments from users.

For the implementation of the function of providing a reference to the legal gist, given the small number of available datasets and the diversity of crimes, instead of using the idea of case classification, we use text matching to provide the legal gist applied by similar cases as the target output.

The main contribution of this paper are as follows. (1) We design and implement a system to assist legal practitioners and ordinary people during the pandemic. (2) The system implemented is based on the WeChat public platform for deployment, with a low threshold of access and user-friendly use. Both legal professionals and citizen can benefit from it.

The rest of this paper is organised as follows. Section 2 present the definition the problem we are going to solve in this paper. Section 3 give presents the details of our system. Section 4 experimentally evaluate our systems. Section 5 discusses the related work. Finally, Section 6 concludes this paper with future work.

2 PROBLEM DEFINITION

For each case entered by the user, we treat the factual description as a sequence of words

$$X = \{x_1, ..., x_N\},$$

where N is the number of words. The input of our system is a sequence of words, and its goal is to give the reference case that is most similar to the input case, or the reference legal gist applicable to that case. We implement the two functions based on similar case matching. Details of the implementation will be described in later sections.

3 SYSTEM PRINCIPLE

As shown in Figure 2, the implementation of the system is supported by the case matching model, which consists of a text encoder and an output module.

3.1 Text Encoder

We use BERT [4] as an encoder to obtain a vector representation of the input case. An illustration of the specific encoding process is shown in Figure 3. Then, we use the case description vector obtained by the encoder for case similarity calculations. Finally, we use the matching result to calculate the loss and train our model.

As shown in Figure 2, the text encoder is designed to convert the input word sequence into a vector representation sequence. Given a pair of case descriptions

$$X^{f_1} = \{x_1^{f_1}, ..., x_n^{f_1}\},$$
$$X^{f_2} = \{x_1^{f_2}, ..., x_t^{f_2}\},$$

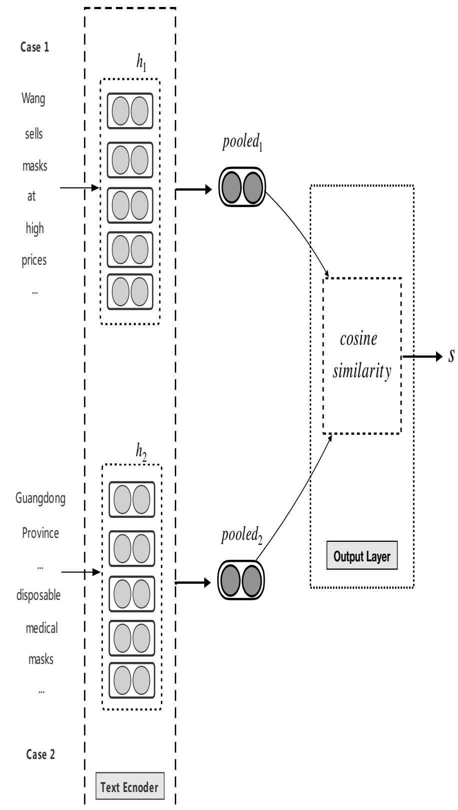


Figure 2: The framework of case matching model.

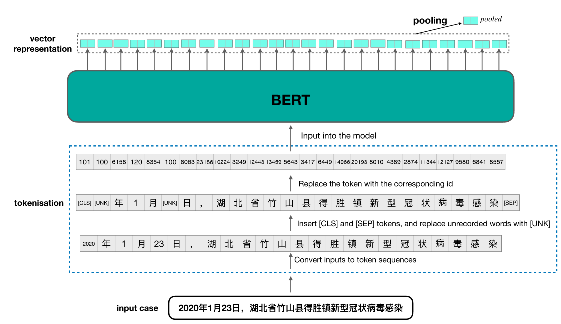


Figure 3: An example of text encoding.

where x_i indicates the i-th word in the sequence, n, t is the length of the sequence, inputting the word sequences into the character-level Bert Chinese model yields the high-level semantic representation vectors:

$$pooled_1 = Encoder(X^{f_1}), \tag{1}$$
$$pooled_2 = Encoder(X^{f_2}), \tag{2}$$

where $pooled_1, pooled_2 \in \mathbb{R}^d$ are the sentence-level vector representations, and d is the dimensionality of the hidden vector.

Table 1: The statistics of the dataset. The table shows number of cases for each crime type.

Crime Type	Number
Hindering the Prevention and Treatment of Infectious Diseases	9
Disrupting Public Service	10
Defiance and Affray	10
Selling Fake and Inferior Products	7
Illegal Business	3
Fraud	10
Robbery	14
Rumour-mongering	29
Illegal Hunting and Killing of Rare and Endangered Wildlife Protected By The State	5
Illegal Acquisition and Sale of Rare and Endangered Wildlife	6

Given the limited amount of training data available in the current system and the uneven distribution of data, we do not use Google's open-source pre-trained Bert model. Instead, we use another open-source pre-trained $Bert_{base}$ model [24] from Tsinghua University, which trains on large-scale criminal instruments, and thus can improve the overall performance of the system.

3.2 Output Module

After passing the text encoder, the original input sequence pairs $< X^{f_1}, X^{f_2} >$ are transformed into new representation, $pooled_1$ and $pooled_2$. In order to better model the similarity between texts, the goal of our model is to minimise the contrastive loss:

$$L = \frac{1}{N}\sum_{i=1}^{N} y_i d^2 + (1 - y_i)\max(margin - d, 0)^2, \quad (3)$$

where N is the size of the batch, y_i is the true label value, $d = ||u_1 - u_2||$, representing the Euclidean distance between two vectors u_1 and u_2, and the margin is a threshold value. Optimising for contrastive loss allows the distance between similar texts to be reduced, while non-similar texts increase their distance to the margin.

Besides, we use cosine similarity to calculate the similarity between text vectors:

$$\text{simarlity}(pooled_1, pooled_2) = \frac{pooled_1 \cdot pooled_2}{||pooled_1|| \cdot ||pooled_2||}. \quad (4)$$

To compare with the true label and calculate the accuracy, we record the cosine similarity greater than 0.5 as 1 and less than and equal to 0.5 as 0.

4 EXPERIMENT

This section will present our experiments.

4.1 Data Preparation and Experimental Setups

We collect data from "Typical cases of crimes against the prevention and control of the COVID-19 pandemic handled by national procuratorates by the law". It is published on the official website of the Supreme People's Procuratorate of the People's Republic of China.[1] By the time the experiment is completed, the statistics of collected data is shown in Table 1. The data collection program is implemented in Python using the BeautifulSoup and Request tools.

Before constituting the final training dataset, we use regular expressions and string processing to extract the facts, the relevant legal gist, and the crime from the raw text data. The constructed dataset consists of 103 items, which is stored in the format of "Crime", "Law Gist" and "Fact" (see Table 3).

Since the implementation of the system is based on similar case matching, we need to process the above data into a format that the model can be trained. Here, we use the crime as a benchmark for judging whether the two cases are similar or not. If the crime of two cases is the same, we view these two cases as similar and set the label to 1; otherwise, we treat these two cases are not similar and set the label to 0. As shown in Table 4, the final training data is a sequence of 3-tuples: $\{y, X^{f_1}, X^{f_2}\}$, where y is a label, X^{f_1} is a description of one case, and X^{f_2} is a description of another case.

The model is implemented with PyTorch framework. We use BertAdam optimiser with learning rate 4e-5 to minimise the loss during training and set the epoch to 8, batch size to 16.

Table 2: Experimental evaluation result of performance in the training dataset.

	Accuracy	Precision	F1
Bi-LSTM	23.39	12.66	22.27
Our Model	11.88	11.88	21.23

4.2 Results and Analysis

We employ Accuracy, Precision, and F1-score as our evaluation metrics. We choose Bi-LSTM as baseline. The results on the final training dataset are shown in Table 2. From the table, we can see that even with BERT the powerful pre-trained language model, the result of our system are still not good enough. It performs even worse than Bi-LSTM. Besides, both models do not perform well enough. These may imply:

(1) To apply deep learning models to real-world scenarios, we need to have a sufficient amount of training data.
(2) In the real-world context of the system, there are too many crime types and too few cases of each type, which poses a considerable challenge to train reliable models.
(3) Since we are using the crime type as a benchmark to determine whether two cases are similar, this can lead to an unbalanced distribution of data between similar and dissimilar cases and a high degree of variability in the within-class sample, which also poses a challenge for fitting the model.

4.3 User Evaluation

As shown in Figure 4(a), we customised the "Followed Reply" function of WeChat Official Account, when new users enter the dialogue window, the system automatically gives a brief introduction

[1]https://www.spp.gov.cn/spp/xwfbh/wsfbt/

Table 3: Examples of formatted data after our processing.

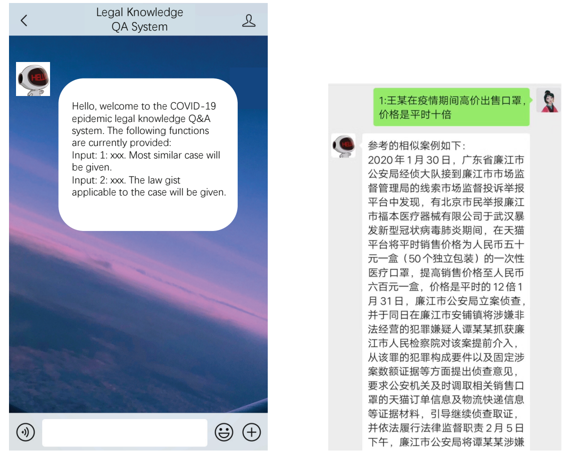
Crime	Law Gist	Fact
Hindering the prevention and treatment of infectious diseases	...does not comply with the provisions of articles 114 and 115, paragraph 1, of the Penal Code...	January 20, 2020, a hospital in Wuhan City, Hubei Province, ...
Disrupting public service	...with the law to prevent and control the pandemic shall be convicted and punished for obstructing official duties in accordance with the provisions of article 277, paragraphs 1 and 3, of the Penal Law.	...the suspect Liu Mou disregarded the government ban, from Zhushan County, Zhushan Town...
Selling fake and inferior products	...is in compliance with the provisions of article 140 of the Criminal Code, the person is convicted and punished for the crime of...	Suspect Shao is a foreign trade practitioners, learned that the COVID-19 outbreak market urgent need mask...
Fraud	...shall be convicted and punished by fraud in accordance with the provisions of article 266 of the Penal Code.	The suspect Cai through the news media was informed of the recent outbreak of new coronavirus infected pneumonia in Wuhan...
...	...	...

Table 4: An example of final training data.

Training data
{“label”: 0, “A”: “January 20,2020, a hospital in Wuhan City, Hubei...”, “B”: “January 23,2020, Desheng Town, Zhushan County, Hubei Province...”}
......

of system functions and usage examples. Users can enter content according to the system’s prompts and get the appropriate responses.

Based on the observations in Section 4.2, the effectiveness of the model is not evaluated using performance metrics on the validation dataset. Instead, real user experience studies are conducted to assess the performance of the system and help improve the design of model architecture. Specifically, we invite our teachers and students to use the system. Based on users’ feedback, the matching results may be less informative when the input case information is briefer or



(a) Tips for new followers translated from Chinese.

(b) An example that a real user uses our system.

Figure 4: (a) When new users follow our WeChat Official Account, the system will give relevant instructions on how to use the system. (b) A real user entered a case about selling masks at an excessive price, and the system responded with a similar case for that user.

when the input case does not fall into any of the training data types. Figure 4(b) shows an example of using the system. In practice, we use the whole saved formatted data (see Table 3) as a matching database.

5 RELATED WORK

This section will discuss the related work to show how our work advances the state of the art in relevant research areas.

5.1 Text Matching

As a basis for similar case matching, the goal of text matching is to measure the similarity between two texts. As an application of text semantic matching, text matching is essential for many tasks of natural language processing, such as reading comprehension, question and answer, natural language reasoning, and summarisation. As a result, this task has received increasing attention from researchers.

As a typical example of earlier work, Huang et al. [11] designed a deep structured potential semantic model for large-scale web search applications that performed well in the web text sorting tasks of the time. Shen et al. [19] improved on previous work by replacing the deep structure with a convolutional pooling structure that captures the rich contextual structure of documents and thus more effectively captures the semantic information in user queries and documents.

Recently, with the successful application of deep learning methods in natural language processing, researchers have also begun to try neural networks for text matching tasks. For example, Feng et al. [6] used a CNN-based QA framework for QA tasks, benefiting from the sparse interaction, parameter sharing, and isotropic representation effects of the CNN network, and experimental results show that the authors' used method performs well beyond the traditional benchmarking approach. Mueller and Thyagarajan [18] used a twin version of the LSTM network to solve the problem of calculating the semantic similarity of pairs of variable-length sequences. Moreover, their experimental results show that LSTM is a powerful language model that can well handle tasks that require complex semantic understanding. The studies mentioned above are all based on the twin structure, which consists of two consistent subnetworks (*e.g.*, CNN and LSTM), and the two subnetworks extract features from two inputs respectively during training [2].

5.2 Legal Intelligent System

Benefiting from a large number of available high-quality textual datasets [3, 5, 15, 23] and web-based public data in the legal field, recently researchers started to study extensively how natural language processing can empower the legal field and solve legal problems. With the successful application of deep learning in the field of natural language processing, many researchers have dedicated their efforts to the problem of legal judgment prediction [9, 12, 16, 17].

Meanwhile, the automatic generations of legal opinions and legal text summaries [1, 13] are also two challenging and worthwhile issues in the area of legal intelligent systems. Ye et al. [22] generate legal opinions based on factual descriptions and use the generated legal opinions to improve the interpretability of the crime prediction system. Whereas, to address the issue of automatic generation of legal text summaries, Grover et al. [7] classify sentences in text based on the rhetorical role of sentences to extract specific types of sentences to form summaries. Another work by Grover et al. [8] differs from the former in that its approach to classifying sentences is based on the argumentative role of sentences.

Much previous work has focused on exploring task-specific improvements on large datasets and demonstrating the feasibility of methods under specific designs. However, there is a lack of empirical evaluation of the application of these methods in real systems. Specifically, if these methods cannot be integrated with specific workflows and provide with relevant usage information, we may not use them in real-world systems.

In this paper, we aim to explore the usability of the domain-specific pre-trained model in legal knowledge question answering system for COVID-19 pandemic. Specifically, using a small number of existing sets, we apply a pre-trained language model to encode the features of the input cases as the basis for training model to further aid in the implementation of the system. Our work explores the effectiveness of transforming the case identification task into the case-matching task in COVID-19 pandemic case. The realisation of this paper contributes to the increase of examples of relevant systems in the field of law and provides empirical knowledge for the practice of relevant systems.

6 CONCLUSION

In this paper, we design and implement a legal knowledge QA system for the COVID-19 pandemic. The system supports the following functions: (1) identify the crime cases entered by the user and find the most similar cases to be pushed to the user as answers; and (2) identify the crime cases entered by the user and give the reference legal gist applicable to the case. Meanwhile, the study could benefit the development of a more comprehensive legal knowledge QA system and other similar systems.

In the future, with the increasing availability of training data, we will introduce more experiment to analyse the effect of the various neural model in this dataset and explore a better model to improve the reliability of the system.

ACKNOWLEDGMENTS

We thank all anonymous reviewers for their valuable comments. This work was supported by the National Natural Science Foundation of China (No. 61762016) and Guangxi Key Lab of Multi-Source Information Mining & Security (No. 19-A-01-01).

REFERENCES

[1] Paheli Bhattacharya, Kaustubh Hiware, Subham Rajgaria, Nilay Pochhi, Kripabandhu Ghosh, and Saptarshi Ghosh. 2019. A comparative study of summarization algorithms applied to legal case judgments. In *European Conference on Information Retrieval*. Springer, 413–428.

[2] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. 1994. Signature verification using a" siamese" time delay neural network. In *Advances in neural information processing systems*. 737–744.

[3] Ilias Chalkidis, Emmanouil Fergadiotis, Prodromos Malakasiotis, and Ion Androutsopoulos. 2019. Large-Scale Multi-Label Text Classification on EU Legislation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 6314–6322.

[4] 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, Volume 1 (Long and Short Papers)*. 4171–4186.

[5] Xingyi Duan, Baoxin Wang, Ziyue Wang, Wentao Ma, Yiming Cui, Dayong Wu, Shijin Wang, Ting Liu, Tianxiang Huo, Zhen Hu, et al. 2019. Cjrc: A reliable human-annotated benchmark dataset for chinese judicial reading comprehension. In *China National Conference on Chinese Computational Linguistics*. Springer, 439–451.

[6] Minwei Feng, Bing Xiang, Michael R Glass, Lidan Wang, and Bowen Zhou. 2015. Applying deep learning to answer selection: A study and an open task. In *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. IEEE, 813–820.

[7] Claire Grover, Ben Hachey, Ian Hughson, and Chris Korycinski. 2003. Automatic summarisation of legal documents. In *Proceedings of the 9th international conference on Artificial intelligence and law*. 243–251.

[8] Claire Grover, Ben Hachey, and Chris Korycinski. 2003. Summarising legal texts: Sentential tense and argumentative roles. In *Proceedings of the HLT-NAACL 03 on Text summarization workshop-Volume 5*. Association for Computational Linguistics, 33–40.

[9] Congqing He, Li Peng, Yuquan Le, Jiawei He, and Xiangyu Zhu. 2019. SECaps: A Sequence Enhanced Capsule Model for Charge Prediction. In *International Conference on Artificial Neural Networks*. Springer, 227–239.

[10] Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, and Maosong Sun. 2018. Few-shot charge prediction with discriminative legal attributes. In *Proceedings of the 27th International Conference on Computational Linguistics*. 487–498.

[11] Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 2333–2338.

[12] Xin Jiang, Hai Ye, Zhunchen Luo, Wenhan Chao, and Wenjia Ma. 2018. Interpretable rationale augmented charge prediction system. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*. 146–151.

[13] Ravi Kumar and K Raghuveer. 2012. Legal document summarization using latent dirichlet allocation. *Int J Comput Sci Telecommun* 3 (2012), 114–117.

[14] Chao-Lin Liu and Chwen-Dar Hsieh. 2006. Exploring phrase-based classification of judicial documents for criminal charges in chinese. In *International Symposium on Methodologies for Intelligent Systems*. Springer, 681–690.

[15] Daniel Locke and Guido Zuccon. 2018. A test collection for evaluating legal case law search. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 1261–1264.

[16] Shangbang Long, Cunchao Tu, Zhiyuan Liu, and Maosong Sun. 2019. Automatic judgment prediction via legal reading comprehension. In *China National Conference on Chinese Computational Linguistics*. Springer, 558–572.

[17] 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*. 2727–2736.

[18] Jonas Mueller and Aditya Thyagarajan. 2016. Siamese recurrent architectures for learning sentence similarity. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. 2786–2792.

[19] Yelong Shen, Xiaodong He, Jianfeng Gao, Li Deng, and Grégoire Mesnil. 2014. A latent semantic model with convolutional-pooling structure for information retrieval. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*. 101–110.

[20] Rosina Weber, Alejandro Martins, and R Barcia. 1998. On legal texts and cases. In *Textual Case-Based Reasoning: Papers from the AAAI-98 Workshop*, Vol. 4050.

[21] Rosina Weber-Lee, Ricardo Miranda Barcia, Marcio C Da Costa, Ilson W Rodrigues Filho, Hugo C Hoeschl, Tania CD'Agostini Bueno, Alejandro Martins, and Roberto C Pacheco. 1997. A large case-based reasoner for legal cases. In *International Conference on Case-Based Reasoning*. Springer, 190–199.

[22] 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: Human Language Technologies, Volume 1 (Long Papers)*. 1854–1864.

[23] Haoxi Zhong, Chaojun Xiao, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. 2020. JEC-QA: A Legal-Domain Question Answering Dataset. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 05 (Apr. 2020), 9701–9708.

[24] Haoxi Zhong, Zhengyan Zhang, Zhiyuan Liu, and Maosong Sun. 2019. *Open Chinese Language Pre-trained Model Zoo*. Technical Report. https://github.com/thunlp/openclap