CLASS: A Novel Method for Chinese Legal Judgments Summarization

Dongjin Li
Department of Computer Science,
Sichuan University, Chengdu,
Sichuan, China
2019223040074@stu.scu.edu.cn

Ke Yang*
China National Tobacco Corporation
Sichuan Province Co., Chengdu,
Sichuan, China
5779293@qq.com

Lijun Zhang
Chengdu Ruibei Yingte Information
Technology Limited Company,
Chengdu, Sichuan, China
630452696@qq.com

Dawei Yin
Sichuan Zhiqian Technology Co., Ltd,
Chengdu, Sichuan,, China,
42453592@qq.com

Dezhong Peng
Department of Computer Science,
Sichuan University, Chengdu,
Sichuan, China
pengdz@scu.edu.cn

ABSTRACT

We propose a novel method to generate abstractive summarization of Chinese legal judgments named CLASS (Chinese LegAl judgmentS Summarization) which exploits the element structure of the legal judgments. Firstly, we extract sentences with high importance from the legal judgments. Secondly, the extracted sentences along with its summaries are split into different source-target element pairs that are used for training an abstractive model to summarize different elements of the judgments separately. Finally, a complete summary is generated by combining the summaries of each element. We conduct comparative experiments on Chinese legal judgments dataset and the results show that CLASS can generate more faithful summaries with less information lost, which shows the effectiveness of CLASS on capturing the deep contextualized information.

CCS CONCEPTS

• **Computing methodologies**; • **Artificial intelligence**; • **Natural language processing**; • **Information systems**; • **Information systems applications**; • **Networks**; • **Network architectures**;

KEYWORDS

Text summarization, Chinese legal judgments, Pretrained model

ACM Reference Format:
Dongjin Li, Ke Yang*, Lijun Zhang, Dawei Yin, and Dezhong Peng. 2021. CLASS: A Novel Method for Chinese Legal Judgments Summarization. In *The 5th International Conference on Computer Science and Application Engineering (CSAE 2021), October 19–21, 2021, Sanya, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3487075.3487161

CSAE 2021, October 19–21, 2021, Sanya, China

ACM ISBN 978-1-4503-8985-3/21/10…$15.00
https://doi.org/10.1145/3487075.3487161

Text summarization aims to produce condensed summaries covering salient and factual information from the original input text. The most popular approaches for summarization can be divided into two categories: extractive and abstractive. Extractive methods usually select sentences directly from the input document based on their importance [1-3]. Liu et al. [4] design a variant of BERT and use it for extractive summarization. Zhong et al. [5] provide a new thought which formulates the extractive summarization task as a semantic text matching problem. In contrast, abstractive summarization methods [6-8] attempt to generate more novel words and sentences that may not appear in the original input, which is much closer to the way humans make a summary. See et al. [9] propose pointer-generator network with coverage mechanism to reduce inaccuracies and repetition of the abstractive summaries. Zhang et al. [10] focus on abstractive sentence summarization, they present a framework named FAR-ASS to improve factual correctness and readability of generated summaries. Hu et al. [11] construct a large scale Chinese short text summarization dataset from website and provide a baseline for further research.

Although text summarization technologies have achieved great success in domains like news articles, emails and scientific papers, there are only few works in the legal domain due to large scale dataset is very rare and hard to be constructed manually. The situation is even worse for Chinese. Besides, the long input and distinctive discourse structure of legal judgments make it easy to lose key information in the generated summaries and much more time consuming.

In order to draw more people's attention to automatic Chinese legal judgments summarization, in this paper, we propose a new method named CLASS which contains three steps: 1) extracting important sentences from the input, 2) splitting the legal judgments into six rhetorical roles including claims for the plaintiff, defense of defendant, court view, relevant law articles and results of judgment document, 3) generating abstractive summaries of each rhetorical role. An example of legal judgment document and rhetorical roles is shown in table 1. We conduct a series of comparative experiments on Chinese legal judgments and the results show that the summaries generated by CLASS is more accurate with less factual error.

Table 1: An Example of Segmented Legal Judgment and Their Rhetorical Roles

Input Document	Summary	Rhetorical role
原告台前县城关镇徐岭东村民委员会诉被告徐某贤侵权责任纠纷一案, ...公开开庭进行了审理。 The Plaintiff Xulingdong Village Committee of Chengguan Town, Taiqian County v. Defendant Xu Mouxian for tort liability dispute...and **held an open trial.**	原告与被告侵权责任纠纷一案。 The plaintiff v. the defendant for the tort liability dispute **case.**	**Case Type**
原告向本院提出诉讼请求**:**1.判令被告立即返还侵占的土地补偿款2万...。 **Plaintiff files claims to the court:** The defendant returns the compensation of 20,000 yuan for the occupied land immediately.	原告提出诉求**:**判令被告立即返还侵占的土地补偿款。 **Plaintiff files claims:** The defendant returns the compensation of 20,000 yuan for the occupied land immediately.	**Plaintiff Claims**
被告徐某贤辩称,... The defendant Xu **contends that**...	被告辩称...。 The defendant **contends that**...	**Defendant Defense**
本案中, 政府对梁苗沟治理的补偿款... **In this case,** the government's compensation for the treatment of Liangmiaogou	经查明政府对梁苗沟治理的补偿款...。 **The court identifies that** the government's compensation for the treatment of Liangmiaogou	**Court View**
依照《中华人民共和国侵权责任法》第十五条之规定 **In accordance with** Article 15 of *The Tort Law of the People's Republic of China*	根据《侵权责任法》第十五条之规定, **According to** Article 15 of the Tort Law	**Relevant Law**
判决如下**:**被告徐某贤将对梁庙沟治理的补偿款2万元返还给原告台前县城关镇徐岭东村民委员会。案件受理费150元, 由被告徐某贤承担。 **The decisions are as follows:** The defendant Xu returned the compensation of 20,000 yuan for the treatment of Liangmiaogou to the plaintiff Xulingdong Village Committee, Chengguan Town, Taiqian County.The cost of this lawsuit is 150 yuan, which shall be borne by the defendant Xu	判决:被告将对梁庙沟治理的补偿款2万元返还给原告。 **The decisions are as follows:**The defendant Xu returned the compensation of 20,000 yuan for the treatment of Liangmiaogou to the plaintiff Xulingdong Village Committee, Chengguan Town, Taiqian County.	**Judgment Results**

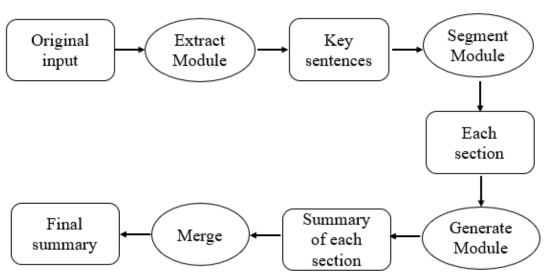

Figure 1: An Overview of CLASS Framework.

1 METHOD

In this section, we will introduce our summarization method which is consisted of three modules: extraction, segmentation and generation. Firstly, the extraction module extracts important sentences from the original judgment. Then, the segmentation module splits these sentences into six rhetorical roles. Finally, the generation module generates abstractive summaries and fuses into one. An overview of CLASS is shown in figure 1

1.1 Extraction

The extraction module aims to select sentences with high importance from the input judgment. It is usually regarded as a sequence labeling task which visit each sentence in the original document sequentially and tag each sentence with label 1 or 0. Let $D = (S1, S2, \ldots, SN)$ be the input where N represents the number of sentences in a legal judgment. Firstly, we get the embedding of each sentence by BERT[12] that pretrained on the legal domain. After that, a document level bi-directional LSTM is used to encode the sequence of sentences and get their hidden representations $d1, d2, \ldots, dN$. Finally, the importance score of each sentence can be calculated by a linear layer with sigmoid function:

$$Y_i = \sigma (W_e d_i + b_e) \tag{1}$$

where W_e and b_e are learnable parameters and σ is the sigmoid activation function. We train the model to minimize the binary cross-entropy function between the output probability Y_i and the golden label y_i. The architecture of extraction network is shown in figure 2

1.2 Segmentation

The output length of extraction module is still so long. We decide to split key sentences by their rhetorical roles for the following two advantages: 1) the length of the key sentences is shortened so that the generation model can be trained faster and more effective. 2) information of each rhetorical role will be included in the final summary so that the final summaries can be more accurate.

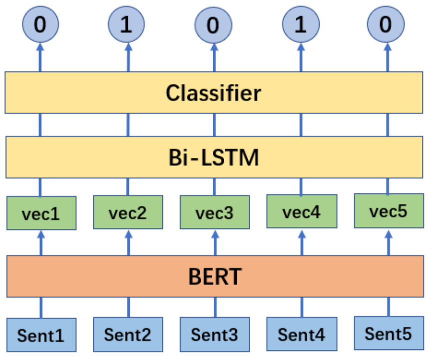

Figure 2: Architecture of Extraction Module.

First of all, we select the judgments that contain indicator clauses of all the six rhetorical roles and segment documents by these indicator clauses. Then we give a rhetorical label to each sentence in the document automatically. After that, we train a Bert-BiLSTM-CRF model with all the parameters of BERT frozen. Finally, we use the trained model to split the rest legal judgments to six rhetorical roles automatically. The summaries are split with the same procedure. An example of key sentence and summary pairs is showed in table 1

1.3 Generation

In the next, the key sentence and summary pairs will be fed into the generation module to generate the corresponding abstractive summaries of each rhetorical role. We select some representative models for comparison.

- Seq2Seq+attention [13]: We use a Seqence-to-Sequence model with attention[14] as our baseline model. The inputs are fed into a single layer Bi-LSTM encoder to get their hidden states and a single layer unidirectional LSTM is used to generate the summaries.
- PGN [9]: The pointer generator network is a novel architecture which is proposed to solve the factual error and repetition problem of Seq2Seq model by using pointing mechanism and coverage mechanism.
- SuperAE [15]: The SuperAE model use the annotated summaries to supervise the source content representation learning due to Seq2Seq model is difficult to learn accurate semantic representation. It achieves the state-of-the-art performance on a Chinese social media dataset.
- CGU [16]: The CGU model add a convolutional gated unit to perform global encoding to tackle the repetition and semantic irrelevance problem of Seq2Seq model.
- Unilm [17]: The Unilm model is based on Transformers and trained by three language model objectives which makes it suitable for both natural language understanding and generation tasks.

In our framework, we use Unilm model to generate the abstractive summary of the extracted key sentences.

2 EXPERIMENT

In the following, we will introduce the dataset and the automatic evaluation metric used in our experiments first. Then, we will compare the results of CLASS with other models.

2.1 Dataset

We use the legal judgement summarization dataset of the Chinese AI and Law challenge (CAIL2020). There are 13531 civil judgments of the first instance and their corresponding summaries. The input documents are split into sentences with importance label 0 or 1, all of which are manually labeled by legal professionals. For data preprocessing, we remove the head and tail sentences of the judgments. The dataset is divided into 80%, 10% and 10%, used for training, validation, and testing respectively. The detailed statistics of the dataset are shown in table 2. We can see that the original input is very long with an average length of 2123.7 words, which makes it challenging to generate abstractive summaries automatically.

2.2 Implementation Details

We use bert-base model to get the embeddings of the input, the dimensions of LSTM hidden states are always kept at 300. In extraction module and segmentation module, we use Adam optimizer with initial learning rate 1e-4 and the weight decay is 1e-5. We trained the model for 50 epochs with a mini-batch size of 16. As for generation module, we finetune Unilm for 20 epochs with a mini-batch size of 8. The masking rate is 0.7 and label smoothing rate is 0.1. Initial learning rate is set to 1e-5. We use beam search with a size of 5 during decoding.

2.3 Evaluation

We use recall score to evaluate the performance of the extraction module. As for the quality of the final generated summaries, we choose standard ROUGE metric [18] and report the macro-F1 scores for ROUGE-1, ROUGE-2 and ROUGE-L computed by pyrouge package.

2.4 Results

Table 3 shows the recall scores of different extraction models. We can see that deep models perform much better than KNN. Both CRF and BERT will improve the performance of Bi-LSTM model.

Table 2: Statistics of the CAIL 2020 Summarization Dataset

	Original Input			Key sentence			Summary		
	Max	Min	Average	Max	Min	Average	Max	Min	Average
Words	12525	402	2123.79	3790	151	794.71	1594	66	280.33
Sentences	480	8	39.39	69	4	12.69	15	1	6.91

Table 3: Recall Scores of Different Extraction Modules

Model	Recall
KNN	84.96
Bi-LSTM	89.63
Bi-LSTM-CRF	91.42
BERT-BiLSTM-CRF	95.67

Table 4: Test Set Results on the CAIL2020 Summarization Dataset Using ROUGE Metric

Model	R-1	R-2	R-L
Seq2seq	66.24	52.23	64.27
PGN	67.86	53.24	65.92
SuperAE	68.96	54.21	66.77
CGU	68.12	53.86	66.12
CLASS	70.45	56.73	68.84

ROUGE scores on the CAIL 2020 dataset are shown in table 4, it is obvious that both SuperAE and CGU can improve the performance of conventional seq2seq model. The ROUGE score of SuperAE is slightly higher than CGU. Our proposed CLASS model performs best on the dataset, which shows the strong ability of the pretrained model to generate abstractive summaries. Besides, we think that CLASS benefits a lot from breaking the document into rhetorical roles, which increases the number of training data and loses less key information from the original legal judgment.

Table 5: Ablation Study Of Different Modules. '-' Denotes Removing the Module from CLASS

Model	R-1	R-2	R-L
CLASS	70.45	56.73	68.84
- Extraction module	44.36	17.79	41.96
- Segmentation module	67.21	53.01	65.47
- Generation module	52.04	34.14	50.68

2.5 Ablation Study

In order to have a better understanding of the influence of each module in our proposed CLASS, we also conduct several ablations including: (i) we remove the extraction module, (ii) we remove the segmentation module, (iii) we remove the generation module. The results are presented in table 5. Removing the extraction module or generation module will cause major performance degradation. Without extraction module, it is hard to split the long input accurately, which will propagate errors to the generation module. Without generation module, the irrelevant words will reduce the quality of the final summaries. The results prove that all the modules are absolutely necessary for CLASS.

2.6 Case Study

We perform case study to evaluate the qualities of the generated summaries. In table 6, we show an example of summaries generated by CLASS compared to the seq2seq model and the reference. Although there is little difference in the rouge scores between seq2seq model and CLASS, the seq2seq model is more likely to generate factual errors and miss important details, which verifies

Table 6: An Example of Summaries Generated by CLASS Compared with Seq2seq Model

Source(truncated): 原告陆凤祥及其委托诉讼代理人肖广玉、被告曹琴、丁国利到庭参加诉讼。本案现已审理终结。原告陆凤祥向本院提出诉讼请求:1、要求撤销联丰村按人头分配年终收益的决议;2、要求被告补偿原告从2012年至2016年未分配到的年终收益, 共计42,600元. . .判决:驳回原告的全部诉讼请求。案件受理费50元, 由原告陆凤祥负担。Fengxiang Lu and his litigation agent GuangYu Xiao, the defendants Qin Cao, Guoli Ding attended the court. The hearing of the case has now come to a close. Plaintiff files Lu Fengxiang claims to the court: 1. Request to revoke the resolution of LianFeng village that the end-of-year bonuses are distributed according to a capitation; 2. Request the defendants to compensate for the end-of-year bonuses from 2012 to 2016, about 42,600 yuan... The decisions are as follows: dismiss the complaint of Plaintiff files. The case acceptance fee of 50 yuan was paid by Fengxiang Lu.
Reference(truncated): 原告与被告侵权责任纠纷一案。原告提出诉求:1、要求撤销联丰村按人头分配年终收益的决议2、要求被告补偿原告从2012年至2016年未分配到的年终收益, 共计42,600元. . . The plaintiff v. the defendant for the tort liability dispute case. Plaintiff files claims to the court: 1. Request to revoke the resolution of LianFeng village that the end-of-year bonuses are distributed according to a capitation; 2. Request the defendants to compensate for the end-of-year bonuses from 2012 to 2016, about 42,600 yuan... **Seq2seq(truncated):** 原被告系侵权责任纠纷。原告提出诉讼请求:要求撤销按人头分配年终收益的决议;要求被告补偿原告从201年未分配到的年假收益, 共计42,600元... The plaintiff v. the defendant for the tort liability dispute case. Request to revoke the resolution that the end-of-year bonuses are distributed according to a capitation; 2. Request the defendants to compensate for the end-of-year bonuses from 201, about 42,600 yuan. . . **CLASS(truncated):** 原被告系侵权责任纠纷。原告提出诉讼请求:1、撤销联丰村按人头分配年终收益的决议;2、要求被告补偿原告未分配到的收益42600元... The plaintiff v. the defendant for the tort liability dispute case. Plaintiff files claims to the court: 1. Request to revoke the resolution of LianFeng village that the end-of-year bonuses are distributed according to a capitation; 2. Request the defendants to compensate for the end-of-year bonuses about 42,600 yuan. . .

the effectiveness of CLASS on capturing the deep contextualized information.

3 CONCLUSION

We present a novel method called CLASS to generate abstractive summaries of Chinese legal judgments. To tackle the problem of the long input and the distinctive discourse structure, we extracting key sentences from the document at first and then splitting them to generate the final summaries. Experiments results on civil judgments dataset demonstrate the effectiveness of CLASS. In the future we will focus on exploring a data efficient model which can be trained end-to-end. How to ensure the factual consistency of the generated summaries is also worth to be explored.

ACKNOWLEDGMENTS

This work is supported by Sichuan Science and Technology Planning Project (Grants No. 2021YFG0317 2021YFG0301)

REFERENCES

[1] Cheng, J., & Lapata, M. (2016). Neural summarization by extracting sentences and words. *arXiv preprint arXiv:1603.07252*.

[2] Nallapati, R., Zhai, F., & Zhou, B. (2017, February). Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Thirty-First AAAI Conference on Artificial Intelligence*.

[3] Li, T., Kou, G., Peng, Y., & Shi, Y. (2017). Classifying with adaptive hyper-spheres: An incremental classifier based on competitive learning. *IEEE transactions on systems, man, and cybernetics: systems*, *50*(4), 1218-1229.

[4] Liu, Y. (2019). Fine-tune BERT for extractive summarization. *arXiv preprint arXiv: 1903.10318*.

[5] Zhong, M., Liu, P., Chen, Y., Wang, D., Qiu, X., & Huang, X. (2020). Extractive summarization as text matching. *arXiv preprint arXiv:2004.08795*.

[6] Rush, A. M., Chopra, S., & Weston, J. (2015). A neural attention model for abstractive sentence summarization. *arXiv preprint arXiv:1509.00685*.

[7] Paulus, R., Xiong, C., & Socher, R. (2017). A deep reinforced model for abstractive summarization. *arXiv preprint arXiv:1705.04304*.

[8] Deng, Z., Ma, F., Lan, R., Huang, W., & Luo, X. (2021). A two-stage Chinese text summarization algorithm using keyword information and adversarial learning. *Neurocomputing*, *425*, 117-126.

[9] See, A., Liu, P. J., & Manning, C. D. (2017). Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368*.

[10] Zhang, M., Zhou, G., Yu, W., & Liu, W. (2021). FAR-ASS: Fact-aware reinforced abstractive sentence summarization. *Information Processing & Management*, *58*(3), 102478.

[11] Hu, B., Chen, Q., & Zhu, F. (2015). Lcsts: A large scale chinese short text summarization dataset. *arXiv preprint arXiv:1506.05865*.

[12] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[13] Nallapati, R., Zhou, B., Gulcehre, C., & Xiang, B. (2016). Abstractive text summarization using sequence-to-sequence rnns and beyond. *arXiv preprint arXiv: 1602.06023*.

[14] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

[15] Ma, S., Sun, X., Lin, J., & Wang, H. (2018). Autoencoder as assistant supervisor: Improving text representation for chinese social media text summarization. *arXiv preprint arXiv:1805.04869*.

[16] Lin, J., Sun, X., Ma, S., & Su, Q. (2018). Global encoding for abstractive summarization. *arXiv preprint arXiv:1805.03989*.

[17] Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., ... & Hon, H. W. (2019). Unified language model pre-training for natural language understanding and generation. arXiv preprint arXiv:1905.03197.

[18] ROUGE, L. C. (2004, July). A package for automatic evaluation of summaries. In *Proceedings of Workshop on Text Summarization of ACL, Spain.*