Long Legal Article Question Answering via Cascaded Key Segment Learning (Student Abstract)

Shugui Xie[1], Lin Li[1*], Jingling Yuan[1], Qing Xie[1], Xiaohui Tao[2]

[1] School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan, 430070, China
[2]School of Mathematics, Physics, and Computing, University of Southern Queensland, Toowoomba, Australia
{xieshugui,cathylilin,yjl,felixxq}@whut.edu.cn, xiaohui.tao@usq.edu.au

Abstract

Current sentence-level evidence extraction based methods may lose the discourse coherence of legal articles since they tend to make the extracted sentences scattered over the article. To solve the problem, this paper proposes a **C**ascaded **A**nswer-guided key segment learning framework for long **L**egal article **Q**uestion **A**nswering, namely **CALQA**. The framework consists of three cascaded modules: *Sifter*, *Reader*, and *Responder*. The *Sifter* transfers a long legal article into several segments and works in an answer-guided way by automatically sifting out key fact segments in a coarse-to-fine approach through multiple iterations. The *Reader* utilizes a set of attention mechanisms to obtain semantic representations of the question and key fact segments. Finally, considering it a multi-label classification task the *Responder* predicts final answers in a cascaded manner. CALQA outperforms state-of-the-art methods in CAIL 2021 Law dataset.

Introduction

Legal Question Answering (LQA) refers to finding answers to a given question by reading and understanding a set of legal articles. Legal articles are rigorously structured and logical. The answer span may needs to be derived through consecutive sentences. Meanwhile, sometimes, the answer spans are derived through distributed sentences since it is in the multi-span answer type. The derivation of certain types of questions (such as multi-span answers) relies on key fact sentences consecutive or scattered over the legal article. This introduces a challenge to pretrained language models with limited input embedding size.

Extracting evidence like key sentences to guide answer prediction is proven useful in long text question answering in the legal domain. We argue that preserving the contextual semantics (discourse coherence) of key sentences is effective to accurate answering. Tackling the problem, this paper proposes CALQA, a *Sifter*, *Reader*, *Responder* cascaded framework, as shown in Figure 1. The *Sifter* first segments a long legal article into several segments, each of which contains one or more consecutive sentences. Key segments are automatically labeled based on whether containing answer spans in the data processing phase. We then design an

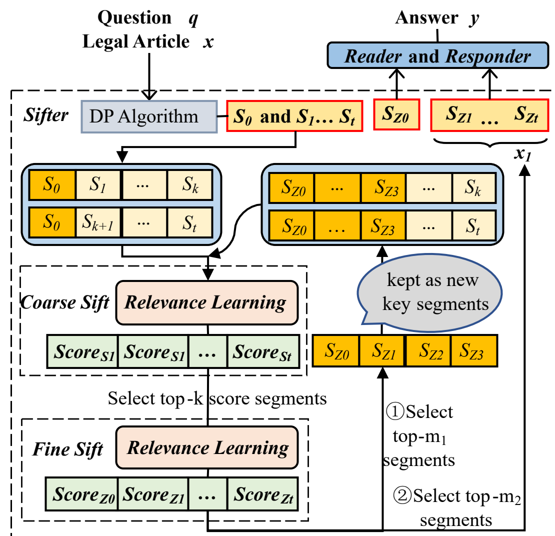

Figure 1: Overview of Our CALQA.

answer-guided supervised learning for *Sifter*, in a coarse-to-fine manner to select key segments and go through several iterations. Thirdly, the selected segments are fed into the semantic representation module *Reader* in their original order. The *Reader* utilizes co-attention and segment-level self-attention mechanisms to learn the question and key segments semantic representation, which will be used as input to a cascaded neural network called *Responder*. The *Responder* maps the semantic representation to the input of a multi-label classification task and obtains the final answer in a cascaded manner. The CALQA framework achieves 83.1 in Macro-F1 and 65.8 in EM and outperformed the state-of-the-art baseline model by 4.1% and 9.1%.

Methodology

Given the input: a long legal article x, a question q, our task predicts an output: y (span(s), yes, no, unknown).

Sifter: Key Segment Selection Based on the Dynamic Programming (DP) algorithm, the *Sifter* in a cascaded design aims to maintain the discourse coherence of the input legal article x and question q while summarizing them to

*Corresponding author: cathylilin@whut.edu.cn

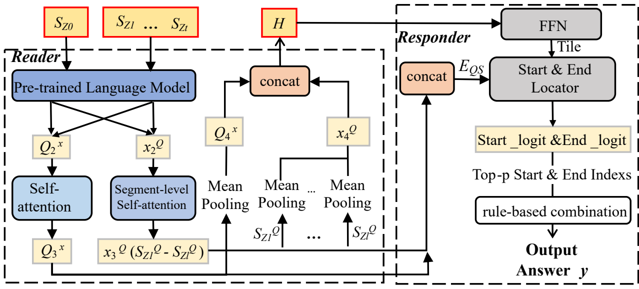

Figure 2: Details of Reader and Responder.

guide the answer prediction. Such a design will also satisfy the input size of downstream models. The *Sifter* first segments the legal article *x* and question *q* into several segments, each of which contains one or more consecutive sentences. A coarse-to-fine relevance calculation is then performed between legal segments and key fact segments as presented in next paragraph. We view the key fact segments as evidence to guide the answer prediction in downstream tasks. It's an iterative process based on an answer-guided training method. Segments with the highest relevance scores will be merged with existing key fact segments, forming new key fact segments. The evidence segments that sifted out through several iterations are fed into the semantic representation module, *Reader*, in initial order of the legal article.

Reader: Semantic Representation Learner As shown in Figure 2, The *Reader* is a cascaded structure with a range of co-attention and segment-level self-attention. After receiving the key fact segments from the *Sifter*, the *Reader* will represent the key fact segments $x_l \in \mathbb{R}^m$ and the question segment $S_{Z_0} \in \mathbb{R}^n$. It outputs the representation of a legal article and its question $E_{QS} \in \mathbb{R}^{m+n}$ and the segment-level representation $H \in \mathbb{R}^{(l+1)\times h}$, where E_{QS} is the concatenation of the context-aware question representation $Q_3^x \in \mathbb{R}^{n\times h}$ and the question-aware token-level representation $x_3^Q \in \mathbb{R}^{m\times h}$. The segment-level representation H is the concatenation of the question segment representation $Q_4^x \in \mathbb{R}^{1\times h}$ and the key segment representation $x_4^Q \in \mathbb{R}^{l\times h}$.

Responder: Cascaded Answer Predictor As shown on the right in Figure 2, the *Responder* takes $E_{QS} \in \mathbb{R}^{m+n}$, $H \in \mathbb{R}^{(l+1)\times h}$ as inputs to predict answers in a cascaded manner. A dense layer F^S with Tanh activation function is employed in segment learning layer.

The start and end logits of the answer are predicted by,

$$H_{start} = F_{start}([H_S^2; E_{QS}]) \in \mathbb{R}^{(m+n)\times h} \quad (1)$$

$$o_{start} = H_{start} W_{start} \in \mathbb{R}^{m+n} \quad (2)$$

$$H_{end} = F_{end}([H_{start}; E_{QS}]) \in \mathbb{R}^{(m+n)\times h} \quad (3)$$

$$o_{end} = H_{end} W_{end} \in \mathbb{R}^{m+n} \quad (4)$$

where o_{start} is the output logit vector of the start positions of the answer, F_{start} is a dense layers with Tanh activation function, and $W_{start} \in \mathbb{R}^{(h\times 1)}$ is a trainable parameter. H_{start} will be used in answer end locator. The training loss is computed over the aforementioned output logits, and jointly minimize these two cross-entropy losses.

	EM	F1
Sliding Window (Wang, Ng, and Ma 2019)	61.94	77.2
CogLTX (Ding, Zhou, and Yang 2020)	60.7	78.3
CogLTX+12 hidden layers attention	60.4	77.9
CogLTX+4,8,12 hidden layers attention	61.3	78.8
RikiNet (Liu and Gong 2020)	61.6	78.2
MacBERT (Cui, Che, and Liu 2021)	60.3	79.8
CALQA	**65.8**	**83.1**
CALQA w/o Sifter	63.3	80.2
CALQA w/o Reader	62.6	79.4
CALQA w/o Responder	63.7	80.8

Table 1: Overall Experimental Results.

Experiments

We follow the CAIL 2021 Law MRC contest setup to obtain the train, dev and test set for evaluating. In the data set, 56.38% articles are longer than 512. The commonly used sliding window results in hard truncated segments accounting for 97.1%, which can't guarantee discourse coherence since a sentence is cutoff directly. Our DP-based method works better with 6.9% hard truncated segments.

Our CALQA is a two-stage training framework. The accuracy of the Sifter sifting out key sentences exceeds 99.5%. In answer prediction, as listed in Table 1, it outperforms other six baselines on both dev set and test set. The result demonstrates that the cascaded key segment learning framework is capable of obtaining semantic information of legal documents and accurately answering multi-type questions.

Conclusions

This study proposes a new Legal Question Answering framework namely CALQA that reads legal articles to answer multi-type question. Experiments on the CAIL 2021 Law MRC dataset shows its promising performance.

Acknowledgments

The authors gratefully acknowledge financial support from Grant 62276196 from National Natural Science Foundation of China, Grant 2021BAA030 from Department of Science and Technology of Hubei Province, China, and Grant DP220101360 from the Australian Research Council.

References

Cui, Y.; Che, W.; and Liu, T. 2021. Pre-Training With Whole Word Masking for Chinese BERT. *IEEE ACM Trans. Audio Speech Lang.*, 29: 3504–3514.

Ding, M.; Zhou, C.; and Yang, H. 2020. CogLTX: Applying BERT to Long Texts. In *NeurIPS*.

Liu, D.; and Gong, Y. 2020. RikiNet:Reading Wikipedia Pages for Natural Question Answering. In *ACL*, 6762–6771.

Wang, Z.; Ng, P.; and Ma, X. 2019. Multi-passage BERT: A Globally Normalized BERT Model for Open-domain Question Answering. In *EMNLP-IJCNLP*, 5877–5881.