Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)
Demonstrations Track

AILA: A Question Answering System in the Legal Domain

Weiyi Huang[1,2,3], **Jiahao Jiang**[1], **Qiang Qu**[1*] and **Min Yang**[1,3*]
[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences
[3]SIAT-DELI AI and Law Joint Lab
{wy.huang1, qiang, min.yang}@siat.ac.cn, jahojiang@gmail.com

Abstract

Question answering (QA) in the legal domain has gained increasing popularity for people to seek legal advice. However, existing QA systems struggle to comprehend the legal context and provide jurisdictionally relevant answers due to the lack of domain expertise. In this paper, we develop an **A**rtificial **I**ntelligence **L**aw **A**ssistant (AILA) for question answering in the domain of Chinese laws. AILA system automatically comprehends users' natural language queries with the help of the legal knowledge graph (KG) and provides the best-matching answers for given queries. In addition, AILA provides visual cues to interpret the input queries and candidate answers based on the legal KG. Experimental results on a large-scale legal QA corpus show the effectiveness of AILA. To the best of our knowledge, AILA is the first Chinese legal QA system which integrates the domain knowledge from legal KG to comprehend the questions and answers for ranking QA pairs. AILA is available at http://bmilab.ticp.io:48478/.

1 Introduction

Existing systems for gaining access to legal resources are either commercial search engines (i.e. Google and Baidu) or legal search platforms, which return legal cases, reports or news [Kejriwal and Szekely, 2018; McElvain *et al.*, 2019]. However, these returned lengthy documents cannot provide an exact solution to the user's problem, and it may be time-consuming to review all of them, without having a guarantee of finding the desired answers. Therefore, developing automated methods to identify valuable answers in response to the user's natural language queries is of practical importance.

Community question answering (QA) websites offer a new opportunity to obtain the desired information in a more rapid and efficient way [Wang *et al.*, 2010; Sakai *et al.*, 2011; Yang *et al.*, 2019]. Inspired by the remarkable success of deep learning in text matching, great efforts have been devoted to applying deep neural networks [Yin *et al.*, 2016; Tan *et al.*, 2015; Yang *et al.*, 2018] to automatically select or generate answers for QA systems in the insurance [Feng *et al.*, 2015] and medical [Tian *et al.*, 2019] domains. However, these QA systems are not applicable in the legal domain since they cannot comprehend the legal context and provide jurisdictionally relevant answers due to the lack of usable QA pairs and domain expertise in the legal domain. For example, in Figure 1, the general-purpose QA systems may fail to identify the correct answers involving "copyright infringement disputes" without the legal knowledge. The words "plagiarize" and "article", which are related to "copyright infringement disputes", can route the input question to the right answer about "copyright".

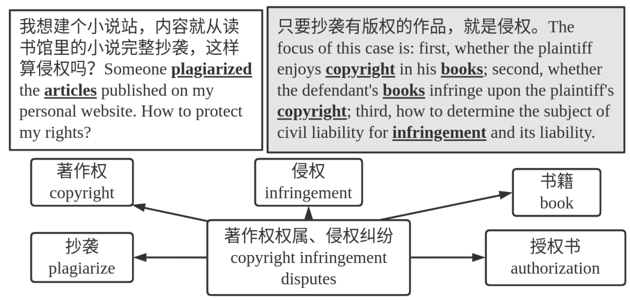

Figure 1: An example of a subgraph from legal knowledge graph, in which the entities in bold are mentioned in the question-answer pair.

In this paper, we develop AILA, a Chinese Artificial Intelligence Law Assistant system, which enables users to ask legal questions and obtain direct answers that are expected to carry useful legal advice. First, we employ a *data acquisition module* to collect a large-scale QA corpus from scratch, which is fundamental to building a high-quality QA system. We also manually construct a legal knowledge graph (KG) that contains legal concepts and their relations, providing legal knowledge to the AILA system. Second, we build a *question answering module*, which fully explores the knowledge from both question/answer and legal KG by interactively learning knowledge-based and context-based sentence representations for ranking answers. Third, an *interpretation module* is devised to provide visual cues to interpret the input queries and candidate answers based on the legal KG. AILA also actively

*Qiang Qu and Min Yang are corresponding authors. This work was partially supported by the National Science Foundation of China (No.61906185, No.61902385), Shenzhen Basic Research Fundation (No.JCYJ20180302145645821 and JCYJ20180302145633177) and Natural Science Fundation of Guangdong (No.2018A030313943).

records users' implicit feedback (e.g., clicked answers and similar questions), which is then utilized to improve the quality of our system. The demonstration video for the system is available in https://youtu.be/BxW7KnztP40.

2 System Overview

The proposed AILA system, as shown in Figure 2, consists of three main modules. (1) The *Data Acquisition* module collects question-answer (QA) pairs from a legal forum and stores the QA pairs in a database. (2) The *Question Answering* module selects top-n best matching answers for a given question. (3) The *Interpretation* module shows selected answers and visualizes the partial results, e.g., related KG entities and attention weights for the answer. Next, we will introduce each component of AILA system in detail.

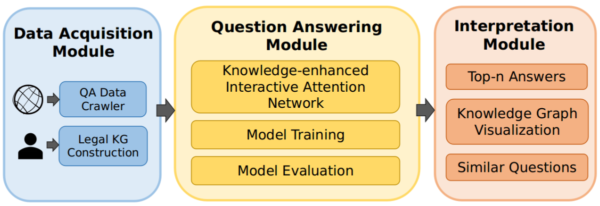

Figure 2: The architecture of the proposed AILA system.

Data Acquisition Module. The *Data Acquisition* module contains an active data crawler which collects QA pairs from the internet and a relational database that stores the collected QA pairs. In this work, we crawl QA pairs from an online Chinese legal forum, which contains a large number of questions asked by real users and corresponding answers provided by licensed lawyers. We also manually construct a legal knowledge graph that contains legal concepts and their relations by inviting professionals from a law firm for the annotation. In total, the legal KG contains 42,414 legal concepts belonging to 1,426 disputes.

Question Answering Module. The QA module is the core part of AILA, which is designed to pick out the best answers for a given query. We propose a knowledge-enhanced interactive attention network to implement the QA module. Specifically, we use bidirectional LSTM (Bi-LSTM) to compute semantic representations of the question and answer. To enable the QA pairs to be aware of background information in the legal domain, we also leverage domain knowledge from legal KG to enrich the representation learning of QA pairs. The concatenation of context and knowledge representations forms the question and answer representations. Inspired by a previous work [Min *et al.*, 2018], we design a coattention mechanism to learn the correlation between the question and answer representations, making use of the interactive information from the QA pairs to supervise the modeling of each other at different representation subspaces. Finally, the interactive question and answer representations are fed into a hidden layer and a softmax layer to obtain QA matching probability. The whole model is trained by minimizing cross-entropy between the ground-truth and the prediction.

Interpretation Module. Given a question, the *Interpretation* module demonstrates the top-n answers and provides partial interpretable results including related legal KG entities and attention weights of the selected answers. In particular, we support users to click on a particular answer to inspect the interpretable results in a user-friendly way.

Model	MAP	MRR
CNN	0.8013	0.8639
Bi-LSTM	0.7903	0.8538
AP-BiLSTM	0.8097	0.8652
BiMPM	0.8151	0.8697
KABLSTM	0.8223	0.8565
AILA	**0.8301**	**0.8866**

Table 1: Performance Comparison

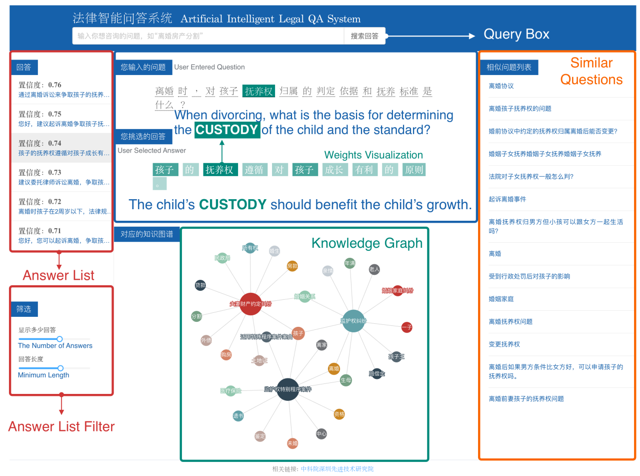

Figure 3: A screenshot of AILA.

3 Quantitative Evaluation

We conduct experiments on LegalQA dataset, which includes 139,468 QA pairs, to evaluate the effectiveness of AILA quantitatively. The evaluation metrics include Mean Average Precision (MAP) and Mean Reciprocal Rank (MRR), which are widely used in answer selection [Tay *et al.*, 2018; Sha *et al.*, 2018; Lai *et al.*, 2018]. We compare AILA with several strong baseline methods including CNN [Severyn and Moschitti, 2015], Bi-LSTM, AP-BiLSTM [Tan *et al.*, 2016], BiMPM [Wang *et al.*, 2017], KABLSTM [Shen *et al.*, 2018]. The experimental results, illustrated in Table 1, show that AILA outperforms the baselines significantly.

4 System Demonstration

Figure 3 shows AILA consisting of 4 steps:

Step 1. The user can access the system by either a PC or a smartphone. After logging in to the system, the user can type in questions in the query box shown on the main page.

Step 2. After submitting the question, AILA offers top-n candidate answers with matching scores. Users can choose the number of returned answers and filter short answers.

Step 3. By clicking on a particular answer, the system visualizes the matching entities within the answer and displays the related entities. In addition, users can inspect the attention weights of answer terms when selecting a question word.

Step 4. Users can also refer to similar questions to the given query, which are listed on the right side of the page.

References

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

[Kejriwal and Szekely, 2018] Mayank Kejriwal and Pedro Szekely. Constructing domain-specific search engines with no programming. In *AAAI*, 2018.

[Lai *et al.*, 2018] Tuan Lai, Trung Bui, and Sheng Li. A review on deep learning techniques applied to answer selection. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2132–2144, 2018.

[McElvain *et al.*, 2019] Gayle McElvain, George Sanchez, Sean Matthews, Don Teo, Filippo Pompili, and Tonya Custis. Westsearch plus: A non-factoid question-answering system for the legal domain. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1361–1364, 2019.

[Min *et al.*, 2018] Sewon Min, Victor Zhong, Richard Socher, and Caiming Xiong. Efficient and robust question answering from minimal context over documents. In *ACL*, pages 1725–1735, 2018.

[Sakai *et al.*, 2011] Tetsuya Sakai, Daisuke Ishikawa, Noriko Kando, Yohei Seki, Kazuko Kuriyama, and Chin-Yew Lin. Using graded-relevance metrics for evaluating community qa answer selection. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 187–196, 2011.

[Severyn and Moschitti, 2015] Aliaksei Severyn and Alessandro Moschitti. Learning to rank short text pairs with convolutional deep neural networks. In *SIGIR*. ACM, 2015.

[Sha *et al.*, 2018] Lei Sha, Xiaodong Zhang, Feng Qian, Baobao Chang, and Zhifang Sui. A multi-view fusion neural network for answer selection. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[Shen *et al.*, 2018] Ying Shen, Yang Deng, Min Yang, Yaliang Li, Nan Du, Wei Fan, and Kai Lei. Knowledge-aware attentive neural network for ranking question answer pairs. In *SIGIR*, pages 901–904. ACM, 2018.

[Tan *et al.*, 2015] Ming Tan, Cicero dos Santos, Bing Xiang, and Bowen Zhou. Lstm-based deep learning models for non-factoid answer selection. *arXiv preprint arXiv:1511.04108*, 2015.

[Tan *et al.*, 2016] Ming Tan, Cicero Dos Santos, Bing Xiang, and Bowen Zhou. Improved representation learning for question answer matching. In *ACL*, pages 464–473, 2016.

[Tay *et al.*, 2018] Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. Hyperbolic representation learning for fast and efficient neural question answering. In *WSDM*, pages 583–591. ACM, 2018.

[Tian *et al.*, 2019] Yuanhe Tian, Weicheng Ma, Fei Xia, and Yan Song. Chimed: A chinese medical corpus for question answering. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 250–260, 2019.

[Wang *et al.*, 2010] Baoxun Wang, Xiaolong Wang, Chengjie Sun, Bingquan Liu, and Lin Sun. Modeling semantic relevance for question-answer pairs in web social communities. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1230–1238. Association for Computational Linguistics, 2010.

[Wang *et al.*, 2017] Zhiguo Wang, Wael Hamza, and Radu Florian. Bilateral multi-perspective matching for natural language sentences. In *IJCAI*, pages 4144–4150, 2017.

[Yang *et al.*, 2018] Min Yang, Qiang Qu, Kai Lei, Jia Zhu, Zhou Zhao, Xiaojun Chen, and Joshua Z Huang. Investigating deep reinforcement learning techniques in personalized dialogue generation. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pages 630–638. SIAM, 2018.

[Yang *et al.*, 2019] Min Yang, Lei Chen, Xiaojun Chen, Qingyao Wu, Wei Zhou, and Ying Shen. Knowledge-enhanced hierarchical attention for community question answering with multi-task and adaptive learning. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 5349–5355. AAAI Press, 2019.

[Yin *et al.*, 2016] Wenpeng Yin, Hinrich Schütze, Bing Xiang, and Bowen Zhou. Abcnn: Attention-based convolutional neural network for modeling sentence pairs. *Transactions of the Association for Computational Linguistics*, 4:259–272, 2016.