AILA: A Question Answering System in the Legal Domain

Weiyi Huang¹,²,³, Jiahao Jiang¹, Qiang Qu¹* and Min Yang¹,³*
¹Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
²University of Chinese Academy of Sciences
³SIAT-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 Artificial Intelligence Law Assistant (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”.

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

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

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. 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.

Question Answering Module. Given a question, the Question Answering 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.

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.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAP</th>
<th>MRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>0.8013</td>
<td>0.8639</td>
</tr>
<tr>
<td>Bi-LSTM</td>
<td>0.7903</td>
<td>0.8538</td>
</tr>
<tr>
<td>AP-BiLSTM</td>
<td>0.8097</td>
<td>0.8652</td>
</tr>
<tr>
<td>BiMPM</td>
<td>0.8151</td>
<td>0.8697</td>
</tr>
<tr>
<td>KABLSTM</td>
<td>0.8223</td>
<td>0.8565</td>
</tr>
<tr>
<td>AILA</td>
<td>0.8301</td>
<td>0.8866</td>
</tr>
</tbody>
</table>

Table 1: Performance Comparison
References


