

IFlyLegal: A Chinese Legal System for Consultation, Law Searching, and Document Analysis

Ziyue Wang[†], Baoxin Wang[†], Xingyi Duan[†], Dayong Wu[†], Shijin Wang^{†§}, Guoping Hu[†], Ting liu[‡]

[†]State Key Laboratory of Cognitive Intelligence, iFLYTEK Research, Beijing, China

[‡]Research Center for Social Computing and Information Retrieval,
Harbin Institute of Technology, Harbin, China

[§]iFLYTEK AI Research (Hebei), LangFang, China

[†]{zywang27, bxwang2, xyduan, dywu2, sjwang3, gphu}@iflytek.com

[‡]tliu@ir.hit.edu.cn

Abstract

Legal Tech is developed to help people with legal services and solve legal problems via machines. To achieve this, one of the key requirements for machines is to utilize legal knowledge and comprehend legal context. This can be fulfilled by natural language processing (NLP) techniques, for instance, text representation, text categorization, question answering (QA) and natural language inference, etc. To this end, we introduce a freely available Chinese Legal Tech system (IFlyLegal) that benefits from multiple NLP tasks. It is an integrated system that performs legal consulting, multi-way law searching, and legal document analysis by exploiting techniques such as deep contextual representations and various attention mechanisms. To our knowledge, IFlyLegal is the first Chinese legal system that employs up-to-date NLP techniques and caters for needs of different user groups, such as lawyers, judges, procurators, and clients. Since Jan, 2019, we have gathered 2,349 users and 28,238 page views (till June, 23, 2019).

1 Introduction

The term Legal Tech refers to legal technologies that apply computer technologies to legal services, such as legal consultation and judicial document analysis. Such techniques are able to ease the load of legal workers and provide easily accessible services for clients. Recently, researchers are concentrating on enhancing Legal Tech with NLP techniques (e.g., named entity recognition (Yin et al., 2018), sequence labeling (Yan et al., 2018)). Studies have proven that NLP techniques are markedly effective regarding several legal tasks, for instance, charge prediction (Hu et al., 2018) and law area classification (Sulea et al., 2017).

In industry, the majority of legal consultation products are merely platforms that redirect users to actual lawyers other than solving problems with a

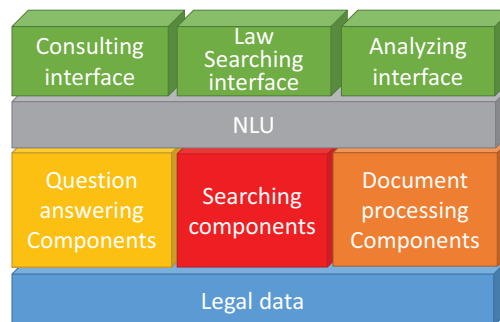


Figure 1: Architecture of the legal system

compact and efficient intelligent system. The partially automated ones, however, tend to have single functionality, either document analysis like case description analysis, or information acquiring like law searching and legal consultancy. Despite of plentiful advances in NLP, most of industrial applications remain constructed in conventional information retrieval (IR) manner or highly rely on hand-crafted responses. They do not take advantage of the most advanced algorithms in NLP and deep learning and fail to cater for the following needs: (i) flexibility: being flexible to comprehend various forms of questions or queries; (ii) diversity: being able to generate different and customized replies according to slight changes of input queries; (iii) accuracy: being able to deliver responses that correctly answer the input questions.

In this paper, we present an integrated system that adapts functionalities such as consultation and document analysis to legal context. Taking advantage of recent advances in NLP, the legal system can act as an artificial lawyer for clients and as an assistant for legal workers. It is a practical system for answering legal questions, performing law searching in multiple modes, and analyzing case descriptions. First, by coupling question-answering and scoring models with external legal knowledge such as statutes and legal commonsense, we design an architecture especially for

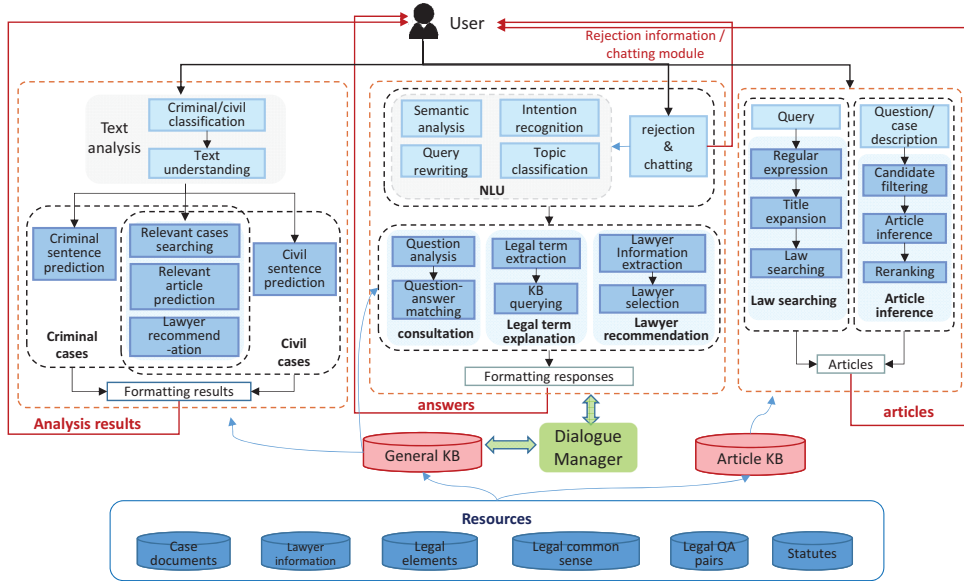


Figure 2: An illustration of the legal system

providing professional solutions to legal issues. Second, inspired by natural language inference task, we build a dedicated module to solving problems with statutes only, namely natural language article inference. Last but not least, we construct an analysis module to comprehend cases, predict judicial sentences, and retrieve similar cases.

The main contributions of this work lie in the followings: 1. It integrates the multiple legal services, consultation, law searching and document analysis, into a single application; 2. We propose a new task, natural language article inference, which replies to legal questions with a sets of possible articles of law only; 3. All the utilized models achieve practicable results.

2 Related Work

The current automated legal consultation applications usually rely on retrieving relevant text information from pre-constructed database containing legal question-answer pairs using text features such as TF-IDF and bag-of-words (BoW). Do et al. (2017) proposed QA models for legal consultation, and Hang (2017) preformed legal question classification with deep convolutional neural network trained in multi-task manner.

Law searching is an unneglectable demand of legal workers such as lawyers and procurators, as they need to support their views with sufficient articles. In industry, article retrieval applications simply parse inputs into phrases and adopt common IR approaches. Academically, Zhang et al. (2017) built a Chinese legal consultation sys-

tem to improve the precision of retrieving articles and predicting sentences by exploiting legal precedents when performing logical reasoning.

Legal document analysis is frequently viewed as text representation and classification task. Hu et al. (2018) introduces few-shot attributes to enrich the information of mapping from case descriptions to charges, and Sulea et al. (2017) used multiple SVM classifiers as an ensemble to perform law area classification.

3 Chinese Legal-tech System

The presented system consists of the following blocks: consultation, law searching and case analysis. We will go into these blocks at length in this section. Throughout the system, the LTP toolkit (Che et al., 2010) is employed for Chinese word segmentation, named entity recognition and semantic parsing tasks. The overall system architecture is depicted in Figure 1. As this paper focuses on legal context processing, we will provide details for principal modules related to legal services and omit the others like chit-chat. For better understanding, we will discuss experiments, user studies and use cases in the following sections.

3.1 Consultation Block

This block is responsible for legal QA. It contains four modules, one of which is called intention recognition and natural language understanding (NLU) module that filters out chit-chats, recognizes intentions of inputs, and analyzes the queries. The other three, namely general le-

gal consulting, legal term explaining, and lawyer recommending, are subsequent modules that response to legal consultation concerning the recognized intention. This block accepts short questions appealing for legal support and outputs literal replies to the questions. We extract text containing legal issues from public corpus (e.g., Wikipedia), and automatically collect web text from Chinese legal forums and online communities for pre-training language models.

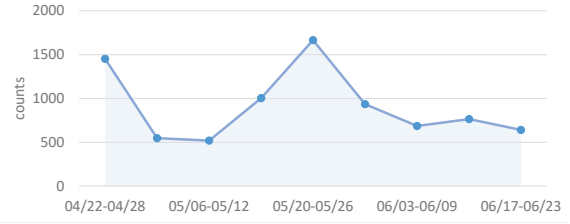
Intention Recognition and NLU module is the basis of consultation block. It functions as the combination of a gate and a converter. As we focus on legal scope, this module only admits legal related content, meanwhile it rejects and redirects the rest to an external chat module. This is achieved by a binary classifier that assigns label “chat” or “legal” to each input. Then, the admitted inputs are analyzed and rewritten for consultation using pre-trained models and predefined features.

General Legal Consulting is an indispensable module for a legal aid system. Following the idea of general QA system (Quaresma and Rodrigues, 2005), we trained an end-to-end QA model especially for legal consulting using data collected from online forums and communities. Note that a legal consultation system should give neutral replies, biased comments are thoroughly removed. This module roles as a virtual lawyer who analyze queries and response with appropriate answers.

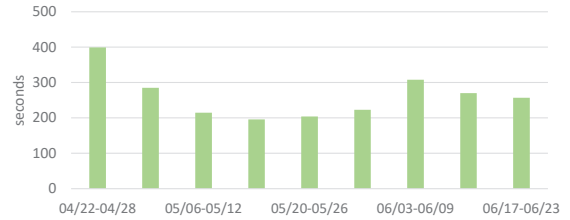
Furthermore, the consultation block has two complementary modules called legal Term explanation and lawyer recommendation that compensate the general legal consulting module with extra useful information. If a legal term appears by itself as a query, it will be detected by intention recognition module and fed into legal term explanation module for detail descriptions. Lawyer recommendation module is personalized with regards to users’ preferred features, for example, lawyer’s location and statistical winning percentage.

3.2 Law Searching Block

Aiming to work as a legal assistant, this block performs law searching in three manners. One is query-based law searching that follows the idea of standard IR approaches. Another is document-based law searching that reads long documents and retrieves applicable articles. The last is a novel task, natural language based law searching. We will discuss it in detail in the next paragraph.



(a) Total Page Views (per week).



(b) Average Viewing Time (per week).

Figure 3: User studies on cumulative page views (a) and average viewing time (b) for the recent two months (from April 22 to June 23)

Article inference is a natural language based law searching approach without formatting the queries. It is a challenging task on account that the number of related statutes to a given question sometimes remains unclear even if answered by experienced lawyers, and that there are different laws and regulations stating the same fact but focusing on different aspects. Undoubtedly, we can simply pair up the input question with all the statutes and perform one-way sentence matching, i.e. deducing articles from a question. Nevertheless, this will result in tremendous amount of sentence pairs due to countless articles of various levels and from different provinces and cities, which makes it impossible to response immediately.

To address this issue, we need to generate a set of candidate statutes to narrow down the searching space and improve the predicting accuracy. We part the article inference task into two consecutive phases. Firstly, we train a law-level classifier that categorizes inputs into 267 in-force Chinese laws and extract articles from the top 3 resulting laws as a coarse candidate set. The total number of this set could go beyond a thousand since some laws contain over 400 articles. So far, the candidate space is still too large to achieve instant response. We use a bidirectional LSTM architecture employing four perspective features from (Wang et al., 2017) as an intermediate article inference model in article-level to obtain a fine-grained candidate set. Secondly, we slightly adapt the BERT model (Devlin et al., 2018) so that it becomes more sensitive

Metrics	p@1(%)	p@3(%)
Automatic	42.67	23.56
Human (Partial)	90.00	75.56
Human (Exact)	76.67	52.22

Table 1: The final test results of article inference task. Human evaluation is contains three level, not match, partially match and exact match.

Models	p%	r%	f%
vanilla RNN	80.11	80.81	80.64
BiMPM	87.05	84.82	85.92
BERT	90.57	87.76	89.14

Table 2: The results of article inference (Test set positive:negative=1:4). **BiMPM** uses 4 perspective features from Wang et al. (2017). **BERT** here is adapted from Devlin et al. (2018).

to legal context. More concretely, the BERT-base Chinese model is trained for another 300,000 steps with 20% of all the question-article pairs. Then, the model is fine-tuned for final article inference task with the training data that has 1 correct article out of 5 on average. In practice, we feed the fine-grained candidate set obtained from phase one to the fine-tuned BERT and get the inference results. Eventually, 3 most probable articles are displayed and mapped to the names of laws accordingly.

Query-based law searching provides retrievals to queries asking for certain laws or as detailed as certain statutes, which is the basic function for a law searching application. It is achieved by retrieving a set of articles with IR system and weighting and ranking by multiple features like textual similarity, priority and validity of laws. **Document-based law searching** returns articles with respect to a piece of case description. It reads and analyzes case description and returns related articles ordered by relevance to the description. In practice, it can act as a fast candidate article pool for a particular case in court trials. The difference between this module and article inference lies in the inputs where document-based searching deals with massive formal text written by law experts while article inference tackles arbitrary short oral questions coming from daily life.

3.3 Case Analysis Block

Case description is an essential component of a judicial document. It states the facts involved in a case, including sequence of events, people presenting during the event, consequences, etc. This block reads and comprehends case descriptions

Functional Modules	Accuracy%
Civil+criminal article prediction	78.0
Civil/criminal classification	95.5
Criminal accusation prediction	98.2
Civil cause prediction	84.3

Table 3: The results of analysis modules.

and reports the analysis results in different formats including statistical graphs of sentences, similar cases, relevant statutes and recommended lawyers.

Civil/Criminal Classification is a preliminary task of case analyzing, since the judging criteria and sentences varies with categories of cases. Cases are generally divided into two classes, civil and criminal. A civil case happens between citizens and the penalty usually excludes the term of imprisonment, while in a criminal case the defendant is prosecuted by public prosecution organ and would be sent to jail if proven guilty. This is a binary classification task fulfilled by pre-trained word vectors and 1-layer convolutional neural network using case description data.

Case Analysis involves civil case analysis and criminal case analysis. The results of both categories contains lists of similar cases, relevant articles and recommended lawyers together with their professional history. For sentences prediction, civil case analysis outputs the possibilities of accuser/defendant winning the lawsuit. The prediction section of criminal case is also known as automatic sentencing that models the sentencing results given a paragraph of criminal case description. The sentences contain the predicted accusation, term of imprisonment and the legal grounds, i.e. the articles of criminal law. We adopt the disconnected recurrent neural network described in Wang (2018) for automatic sentencing task.

4 Experiments and User Study

We conduct experiments on all modules and will report some important results in this section. User studies on cumulative page views and average viewing times are presented in figure 3. For the consultation block, the testset contains 2000 questions with varying legal topics. The corresponding replies are automatically obtained by the system. The retrievals are manually scored between 1 to 5, where 1 represents irrelevant answers and 5 stands for the best matches. The system achieves an average precision of 80% for the top 1 retrievals.

We compare the results of article inference task



Figure 4: Use case 1: use case of consultation block.



Figure 5: Use case 2: use case of law searching block.

regarding different models and list them in Table 2. Statistics show that the vanilla RNN is able to properly predict the article from the given candidates, but is much inferior to the other complicated models. Yet, it is not the final result for the task in practice due to the inconsistencies between the distributions of experimental and real data. The manually created dataset has one correct article out of 5 candidates on average. In reality, surprisingly, it almost equals to pick up the only one correct answer out of 600, making it an extremely challenging task. We evaluate the results of 200 arbitrary questions by precision@n in two ways, automatic and manual, and report in Table 1. Regardless of disappointing automatically evaluated results, manual evaluation reveals the practicability of the models and proves that they are able to figure out most of the answers.

The testset for the analysis block is automatically extracted from the original court documents, where categories are indicated in the titles and information like the relevant articles and sentences are always listed at the end. We evaluate the performance of analysis block with respects to the following tasks: (i) civil and criminal article prediction; (ii) binary classification of civil and criminal cases; (iii) accusation prediction for criminal cases and (iv) cause prediction for civil cases, which is the counterpart of accusation as to criminal cases. The results are listed in Table 3.

5 Use Cases

In this section, we will present cases illustrating the three blocks, consultation, law searching and

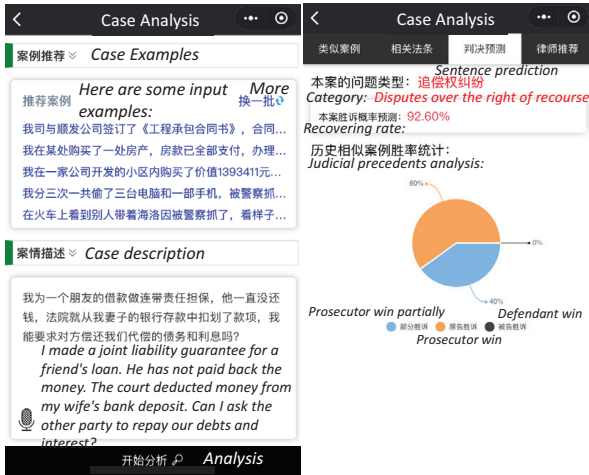
case analysis. Also, we encourage readers to try iFlyLegal via scanning the QR code in figure 7.

Figure 4 is a use case of the consultation block. The inputs can be arbitrary questions as long as they contain legal issues, such as “How to divide common property after divorce?” and “Is temporal worker a legal labor relation?”. A proper answer will be delivered along with a set of applicable statutes. Figure 5 displays part of the law searching results for the tested input “Criminal Law Article 200”. It will be recognized as “Criminal Law” plus a further restriction on the number of article, “Article 200”. Figure 6 demonstrates the results of case analysis block with a screen-shot of the predicted sentences. The other information can be found in the screencast.

6 Conclusion and Future works

We present a system called iFlyLegal for automated legal QA, multi-way law searching, and multi-perspective legal document analysis. The system is built upon a combination of classical text features and deep learning techniques in NLP. We conduct sufficient experiments and report important results in this paper. To help understand our system, we illustrate several use cases with snapshots and necessary literal explanation. These cases also prove that our system is capable to fulfill user demands.

Taking into consideration the need for ease of access, the system is demonstrated in form of WeChat Mini Program, which is compact, portable and freely available. Yet we will develop a web-based version for those who prefer access-



(a) Input Interface. (b) Predicted Sentences.

Figure 6: Use Case 3: use case of analysis block. (a) is the case description. (b) present the predicted sentences.



Figure 7: Mini program QR code of IFlyLegal (scanning via WeChat). We also provide a screencast for readers who do not have a Wechat.

ing via computers. During maintaining, the models and topology behind would be improved along with our researches. Although neural network is regarded as “black box”, we are currently working on the explainability of our models and trying to present users with evidences to the model-generated outputs so as to be convincing. IFlyLegal is an integrated and multi-functional system, whose build-in modules are complex and can be separated and adapted for different tasks such as text categorization and natural language inference. We intend to turn iFlyLegal into an NLP platform for legal AI research to cater for the boosting needs of NLP techniques in legal industrial and explore other valuable research topics of legal NLP. In the future, we will investigate into adapting the NLP research platform to other languages.

Acknowledgments

The authors would like to thank all the reviewers for their insight reviews. This paper is funded by National Key R&D Program of China

(No.2018YFC0807701).

References

- Wanxiang Che, Zhenghua Li, and Ting Liu. 2010. Ltp: A chinese language technology platform. In *Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations*, pages 13–16. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Phong-Khac Do, Huy-Tien Nguyen, Chien-Xuan Tran, Minh-Tien Nguyen, and Minh-Le Nguyen. 2017. Legal question answering using ranking svm and deep convolutional neural network. *arXiv preprint arXiv:1703.05320*.
- Nga Tran Anh Hang. 2017. Applying deep neural network to retrieve relevant civil law articles. In *Proceedings of the Student Research Workshop associated with RANLP*, pages 46–48.
- 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*, pages 487–498.
- Paulo Quaresma and Irene Pimenta Rodrigues. 2005. A question answer system for legal information retrieval. In *JURIX*, pages 91–100.
- Octavia-Maria Sulea, Marcos Zampieri, Shervin Malmasi, Mihaela Vela, Liviu P Dinu, and Josef van Genabith. 2017. Exploring the use of text classification in the legal domain. *arXiv preprint arXiv:1710.09306*.
- Baoxin Wang. 2018. Disconnected recurrent neural networks for text categorization. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 2311–2320.
- Zhiguo Wang, Wael Hamza, and Radu Florian. 2017. Bilateral multi-perspective matching for natural language sentences. *arXiv preprint arXiv:1702.03814*.
- Yukun Yan, Daqi Zheng, Zhengdong Lu, and Sen Song. 2018. Zooming network. *arXiv preprint arXiv:1810.02114*.
- Xiaoxiao Yin, Daqi Zheng, Zhengdong Lu, and Ruifang Liu. 2018. Neural entity reasoner for global consistency in ner. *arXiv preprint arXiv:1810.00347*.
- Ni Zhang, Yi-Fei Pu, Sui-Quan Yang, Ji-Liu Zhou, and Jin-Kang Gao. 2017. An ontological chinese legal consultation system. *IEEE Access*, 5:18250–18261.