

# Legal Artificial Intelligence - Have You Lost a Piece from Jigsaw Puzzle?

Changlong Sun,<sup>1</sup> Yating Zhang,<sup>1</sup> Qiong Zhang,<sup>1</sup> Xiaozhong Liu,<sup>2</sup>

<sup>1</sup>Alibaba Group, Hangzhou, Zhejiang, China

<sup>2</sup>Indiana University Bloomington, Bloomington, Indiana, USA

changlong.scl@taobao.com, ranran.zyt@alibaba-inc.com, liu237@indiana.edu, qz.zhang@alibaba-inc.com

## Abstract

Legal artificial intelligence, as a special track in AI, is playing an increasingly important role to address different kinds of legal needs, and to provide vital potentials to help clients, lawyers and judges to access, understand, predict, and generate legal information in the context of legal domain knowledge. Legal AI, however, can be more challenging than other AI topics, and a comprehensive multi-view legal case representation, across different stages, can be essential for a number of downstream tasks, e.g., legal prediction, court debate mining, and legal QA/chatbot. In this paper, we explore the theoretical and methodological foundations, potentials and challenges to address this novel problem.

## Introduction

Over the past few years, a number of domains, like text mining, computer vision, and auto-drive, have reaped the benefits of embracing data-driven methods along with the emerging deep learning models. These approaches simplify systems while minimize the potential for humans to introduce their own biases. More importantly, such enabling technologies has been commercialized to satisfy various kinds of needs from massive users. Legal domain AI, however, can be more challenging and bewildering than other text mining/NLP disciplines, and some studies even expressed the concern that the exaggeration of AI in legal area backfired, and machine should not step into this serious domain (Mills 2016). In this context, legal AI investigation can be critical while such needs are both necessary and inevitable. For instance, based on New York Times report, “Trial judges are suffering from ‘daunting workload’<sup>1</sup> is becoming an increasingly critical issue, which challenges the efficiency of legal justice ecosystem in different nations. According to the report of statistics, the typical active federal district court judge closed around 250 cases in a year, therefore, applying novel artificial legal intelligence techniques to facilitate the lawsuit process so as to alleviate the overwhelmed workload of judges is of great significance (OECD 2013).

In this study, we investigating the opportunities and challenges in legal AI from case representation, learning model

and data availability perspectives. While a number of very recent studies are explored and strategically reviewed, a new learning framework, Joint Multi-Stage Case Representation Learning (JMCRL), is proposed to characterize the semantic, logic, and knowledge context of a legal case.

## Recent Legal Deep-AI Efforts with Different Stages

Unlike most existing topics in machine learning, a legal case, or to say a legal information need from user, may experience different stages. To the best of our knowledge, so far, few studies dynamically explore the case representation across different legal stages. Recently, however, deep learning has been successfully investigated to address a number of AI problems in each stage along, which can be summarized as the followings.

**Case prior knowledge stage** addresses the legal case/information-need context characterization problem, e.g., the relevant precedents, statutes and undisputed legal concepts retrieval from the legal databases and characterize the case context. The legal contextual information can provide essential information to represent the target case/need. For instance, (Li et al. 2018) locate the relevant statuses from DB, as the target case’s context, by using CNN plus Correlation Matrix that could cope with ambiguity and variability problems.

In a **pre-trial stage**, the case indictment and evidences, from plaintiff or defendant, can provide key information to predict the legal decisions. (Zhou et al. 2019), for instance, proposed a novel multitask learning model to represent the case by using plaintiff (buyer), defendant (seller), and indictment (dispute) information in an eCommece ecosystem. More importantly, authors found the legal knowledge (graph) can play an important role for case representation learning, e.g., ablation test showed legal knowledge enhanced case representation can improve the model performance by 5%.

In the **trial court stage**, different parties, like plaintiff, defendant, judge and lawyer, have the chance to change the sentence result and the associated case representations in a court debate context. While the debate representation learning can be more challenging, more recently, (Duan et al. 2019) proposed a novel deep debate representation learning

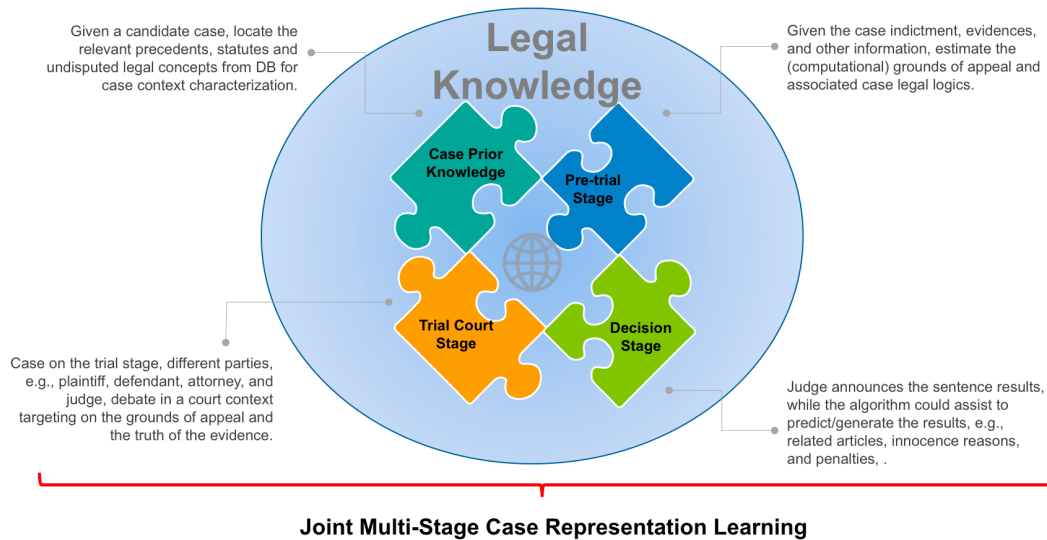


Figure 1: Example Dialog in Court Debate Dataset

framework. As the most interesting finding, authors proofed the role information can be more important than legal knowledge and case global information for debate mining. Various types of information can all contribute to the learning tasks, e.g., debate summarization.

### Joint Multi-Stage Case Representation Learning (JMCRL)

In the paper, we propose a novel legal case representation learning framework by comprehensively integrating four different stages: **Case Prior Knowledge (learning) stage**, **Pre-trial stage**, **Trial Court Stage**, and final **Decision stage**. Figure 1 depicts this model.

It is clear, legal case representation learning, comparing with other types of learning tasks, can be more challenging because of the following reasons.

First, the optimized legal data analytics/mining/prediction solutions may need to explore the understanding of their implications across different stages, and the case representation could change significantly when the same factor transferring from one stage to another. For instance, the ‘contract assessment result’ (of the target case) may change from pre-trial stage to trial court stage with the additional input from plaintiff, and the prediction result could change correspondingly.

Second, a multi-view learning should be used to encapsulate the heterogeneous information of the target case. Different factors, e.g., the information from different parties, can play different roles in the learning model, and more sophisticated representation learning algorithm should be applied to address this challenge.

Third, different kinds of information should be projected into two kinds of representation spaces, semantic space and legal knowledge space. For instance, both (Duan et al. 2019) and (Zhou et al. 2019) found legal knowledge (graph) can be nontrivial for legal case representation, and the linguistic features, like word, sentence, utterance (in debate), and sequential information should be projected into the legal knowledge space to enhance the representation accuracy.

Last but not least, we can further enhance the performance of the task(s) in the decision stage by leveraging *multi-task learning*. In a legal context, the final decision may highly likely associate different sub-tasks, e.g., related articles prediction, penalty calculation, and reason generation. A multi-task framework can help the model better optimize the representation parameters while enabling the communications among the tasks, which have been proofed as an effective means to enhance the legal AI jobs.

Unfortunately, data barrier restricts JMCRL investigations and implementations. While the case database, court debate corpus, and legal knowledge graph are increasingly available for legal AI research, no dataset can interconnect them for cross-stage case representation learning. In the next step, we will be working on this problem. Efforts will be made to create novel dataset to enable future legal AI studies.

### References

- Duan, X.; Zhang, Y.; Yuan, L.; Zhou, X.; Liu, X.; Wang, T.; Wang, R.; Zhang, Q.; Sun, C.; and Wu, F. 2019. Legal summarization for multi-role debate dialogue via controversy focus mining and multi-task learning. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 1361–1370. ACM.
- Li, C.; Ye, J.; Ge, J.; Kong, L.; Hu, H.; and Luo, B. 2018. A novel convolutional neural network for statutes recommendation. In *Pacific Rim International Conference on Artificial Intelligence*, 851–863. Springer.
- Mills, M. 2016. Artificial intelligence in law: The state of play 2016. *Thomson Reuters Legal executive Institute*.
- OECD. 2013. What makes civil justice effective? *OECD Economics Department Policy Notes* (18).
- Zhou, X.; Zhang, Y.; Liu, X.; Sun, C.; and Si, L. 2019. Legal intelligence for e-commerce: Multi-task learning by leveraging multiview dispute representation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 315–324. ACM.