Developing a Legal Assistant System for Experts and the Common Man in India

Student Name

Department

Institute Name, India

Email id : xyz@gmail.com

Abstract

With the advancement of the Web and large number of legal documents being made available digitally, legal practitioners in India (as well as many other countries) are now facing certain new challenges. It is now intractable for legal practitioners to manually find relevant information (prior cases, related acts/statutes etc, law reports etc.) that would assist an ongoing case. Another serious problem in India and many other countries is the lack of awareness of law among the common people. Even for day-to-day issues, they are bound to seek legal help, the cost of which is prohibitively high. Additionally, the number of pending litigations in the Indian courts is increasing exponentially.

In this scenario, it is a need of the hour to design AI systems that can expedite the legal decision-making process in India. The objective of this proposal is to develop such a legal assistance system that can be used both by legal practitioners as well as by the common man in India. For the legal practitioner, the system will automate several tasks that are traditionally done manually, e.g., identifying relevant documents, summarizing legal text, predicting the outcome of arguments, and so on. For the common man, the system will attempt to guide him in legal situations and will play an important role in raising the general awareness of law.

Background and Motivation

The legal expert's side: In countries following the Common Law system (e.g., India, UK, Canada, Australia, and many others), there are two primary sources of law -- Statutes/Acts, which are the laws made by the legislature and Precedents, which contain solutions to similar legal problems not directly indicated in the law. When a new case comes to a legal practitioner, he has to study previous cases that are similar in terms of the legal facts and issues as the current case, in order to understand how the Court has discussed, argued and behaved in similar scenarios. Hence

lawyers have to go through hundreds of prior cases. There exists legal search systems like Manupatra and Westlaw India for such tasks. However, from our discussions with legal practitioners (from the Rajiv Gandhi School of Intellectual Property Law, IIT Kharagpur), we understand that all these systems charge very high subscription cost, and very few legal professionals (other than large law firms/institutes) can afford to regularly access these systems. Additionally, most of these systems perform a keyword-based search and the search results are not of much satisfaction to the legal experts. Besides, apart from a full-length case document in response to a legal query, a legal expert also wishes to see related statutes, catchwords of the documents, or certain segments of the document like facts/final judgment/arguments instead of the full document. These primary features are missing in existing legal search systems.

Additionally, the court case documents are long and unstructured with dense legal text. This makes reading and comprehending the full text of a case a difficult task, even for a legal expert. In scenarios like this, summaries of the case judgments prove to be beneficial. All popular legal retrieval systems provide summaries of case judgments manually written by legal attorneys [24]. Employing experts to write the summaries incurs high amount of cost (which in turn leads to high subscription costs for the commercial systems). This scenario calls for an automated legal document summarization system, that would assist both a legal expert and a common person for understanding numerous precedent cases in a short time.

Hence, from our discussions with legal practitioners, we understand -- when a case comes to a legal practitioner, he will benefit from an AI system that returns relevant documents and information, and about the different outcomes the case might have. Such a system will help the practitioner decide whether he will pursue the case or not, and if yes, how he/she should frame the arguments.

The common man's side: The common masses of India lack awareness of law. For even day-to-day issues like warranty of malfunctioning products, tenant-landlord issues, the common man in India is not aware of the relevant laws and the course of action to be taken. Additionally, the high cost of consulting a legal professional even for preliminary advice often prohibits the common man from taking legal course in getting his / her dues. The legal search systems available (stated above) are for use by legal professionals, and cannot be used by a common layman (who cannot search using legal keywords). In this scenario, there is an alarming need of the law to be made accessible in a way that is easily understandable in natural English language to the common people (in terms of the outcomes of similar prior instances, related statutes,

course of action to be taken). For instance, given the description of a scenario in *natural English*, the system will be able to suggest relevant prior cases, or give guidance as to what legal course of action can be taken. From our discussions with law experts, we understand that such a system will be very beneficial not only to common masses in India (who greatly lack knowledge of law) but also to law students and the legal academia in India.

The number of pending litigations in India is increasing exponentially, and the agonizingly slow judicial process in India affects millions of people[1]. This situation requires the intervention of an assistant system that can not only help to automate tasks that are currently done manually, but also help the common man and legal professionals to understand whether it is worthwhile to file a case at all (or to perform an out-of-court settlement) by understanding, e.g., the chances of winning the case based on decisions in similar prior cases. From our discussions with legal experts, we believe such a system can be useful to reduce the number of cases being filed daily. Additionally there is a huge scarcity of legal expertise in the country. Hence it will be useful to have a system for load balancing, that attempts to infer the complexity of a case, so that junior lawyers can be assigned routine cases, and senior lawyers can focus on non-trivial cases.

It can be noted that, though the focus of this proposal is on law in India, similar problems exist in several other countries of the world. For instance, from our discussions with legal experts in the UK (Dr. Adam Zachary Wyner, Associate Professor of Law and Computer Science, Swansea University, UK) and USA (Dr. Jack Conrad, Lead Research Scientist, Thomson Reuters), we understand that making law accessible to the common man is a serious challenge in the UK and USA as well. Hence the problems outlined in this proposal are of relevance not only in India, but also in several other countries of the world.

Problem Statement and Proposed Approach

We envisage a system that will encompass a large body of knowledge on the Indian Law, so that it can cater to many practical tasks / information needs of law experts as well as of the common masses having legal queries. The Web-based legal assistance system will be usable both by law experts as well as by the common masses to answer legal queries. Developing the system needs addressing certain challenges, which are described below.

---

[1] https://www.theguardian.com/world/2016/may/05/indias-long-wait-for-justice-27-million-court-cases-trapped-in-a-legal-logjam

1. <u>Modelling the Indian legal Judiciary:</u> In order to accomplish the wide variety of tasks mentioned above, it is necessary for the system to have a representation of the basic underlying structure of the Indian law. We are computationally modeling the legal system as a heterogeneous network (a knowledge graph). On this knowledge graph, it is then possible to apply computational techniques which will form the basic building block of the legal assistance system. Information extraction from legal documents is a challenging task. It has been shown in [24] that standard NLP tools do not work well on these documents.

2. <u>Legal Document Similarity and Prior Case Retrieval:</u> Since the Common Law system stands on the principle of precedents, it is an integral part of any legal data mining task to find out 'similar' cases based on the situations/premises of a current case. A 'similar' case is considered to be a precedent if it is authoritative by legal principles and has addressed some principle foundations of the existing law. The challenge here is to develop explainable AI models which can interpret as to why a document is similar to a current case, and what are the justifications for it to be considered as a precedent.

3. <u>Legal Document Segmentation and Summarization:</u> Legal case documents are usually very long (often spanning hundreds of pages) with dense legal text and complex domain-specific terminologies. This makes reading and comprehending the full text of a case, even by a legal expert, a difficult task. While case documents from other countries like Canada, Australia and UK contains section headings like "Facts of the case" , "Background" , "Relevant statutes", "Final Judgement" etc. Indian case documents are highly unstructured and do not contain any such section headings. A lawyer might intend to only read the facts and background of the case or the reasons for the final judgment rather than reading the full text. In such a scenario, segmentation of the case document turns out to be useful. This is a challenging task because there is little or no manual annotations available over which supervised models can be trained.

Although there has been methods developed for summarization of legal documents of other countries, it has been shown in our recent work [24] that these methods do not scale well to summarizing Indian court case documents. This motivates us to design new extractive and abstractive summarization algorithms which not only performs well in the Indian scenario but also generalizes well to documents across countries.

4. <u>Case Classification and Judge Assignment</u>: Classification of cases into categories is an important first step, that is presently done manually. Automating this step can speed up the process of judge and lawyer assignment. Also, based on judge profiles, a task allocation problem that can match the category of the case with the expertise of the judge is an interesting and socially useful problem to solve.

5. <u>Citation Network Analysis:</u> The legal system of a country is a large and growing system, with new cases citing other cases and statutes citing each other. As stated earlier, we are modeling the Indian legal system as a network. An in-depth analysis of this citation network can help to to determine the authority of cases, use citation context to understand the diversity of the cases citing a particular case, the citation profile of cases and statutes across years, distribution of the different types/categories of cases arising in different parts of the country and across timelines and many other interesting applications. While significant amount of work has been done in the area of citation analysis in scientific literature, only a very small amount of work has focused on studying case law citations [25].

6. <u>Predictive tasks and case-based reasoning:</u> Given the facts of a current situation, we plan that our legal assistant system will be able to retrieve the final decisions of cases with similar situations, and based on these cases, statistically predict the probable outcome of the current case. Such reasoning is known as case-based reasoning in legal literature. These predictive features would be especially effective in helping a legal practitioner to decide whether to take up a case, or in helping a common man decide whether to file a case at all (or go for an alternate settlement). From our discussions with legal experts, we understand that such predictive features can help in reducing the number of cases being filed, and thus help in reducing the backlog of pending litigations in India.

7. <u>Question Answering:</u> A common person having little or no knowledge of law may want to do a background study about the existing law and procedures relevant to his problem, before moving the Court. We envisage that our system, when given a description of the scenario in natural English, will be able to suggest relevant prior cases, or give guidance as to what legal course of action can be taken. This is a challenging problem to solve because the language/vocabulary of the law and the language of the common differs widely.

8. Extending the methodologies to legal documents of other countries / in other languages: Though initially we want to focus on Indian legal documents, we will explore later how the methods generalise to documents of other countries. Also, we intend to develop methodologies for non-English languages, e.g., other Indian languages. This is important in order to make law accessible to various populations.

Brief Literature Review

Some of the tasks described above have been addressed in the data mining domain in general, such as knowledge base creation for factual knowledge [27], summarization (of news articles) [28], citation network analysis (for scientific literatures) [29], open-domain question answering [30] etc. However, most of the methods have not been applied in the legal domain. Analysing legal text has several challenges that are domain-specific. For instance, it has been noted in [24] that standard NLP tools for pre-processing and Named Entity Recognizers do not work well in legal documents. Similarly, most state-of-the-art summarization methods are developed for news documents, which are a homogeneous body of text. Legal documents differ significantly in the way that there are different rhetorical categories/segments (facts of the case, background, arguments, reason for judgement etc.) and all these aspects need to be captured in the summary [24]. This calls for the need of legal domain-specific techniques.

In the legal domain, there has been prior work in developing ontologies [1,2,3], finding similarity of legal documents [6,7], legal document summarization [8,10,11], legal citation network analysis and understanding the authoritativeness of cases [13-18], question-answering [22,23], building legal recommender systems [19-21], and so on. It is to be noted that, most of these legal domain-specific systems or methodologies are at a very nascent stage, in comparison to systems being made for other open-domains (e.g., YAGO, DBPedia, Freebase for knowledge graph mining, deep learning models for summarization of news articles, etc.).

Also it can be noted that, most of these methods have been developed for legal documents of other countries (UK, USA, Canada, Australia), and there has been little work on the legal domain in India. However, in [24, 26] it has been seen that the methods developed for legal documents of one country do not generalise to Indian case documents. An important reason for this is that, unlike documents from other countries, Indian case documents are very less structured and there is a wide variation of legal terminologies across countries. Some initial work has been done on legal document summarization [9], ontology construction [2], document similarity [6,7] and

catchphrase extraction [26] from Indian legal documents; but state-of-the-art AI approaches such as neural networks have not been used in any of these works.

Apart from the problems where some work has been done, we aim to address some novel problems as well. To our knowledge, there has been no attempt towards designing systems that can aid a common man in answering questions in layman terms. While the problem can be cast in a machine translation framework, data required to train such systems may not be readily available. In order to make the law reach out to the common masses, it is essential to build systems that can provide information in non-English or regional languages. This presents a new direction in the field of multilinguality in law, for which there is no available literature, to the best of our knowledge. Classification of case documents and matching judge expertise for efficient allocation of cases is another important aspect to explore, keeping in mind the problem of pending litigation and scarcity of legal expertise in the country.

As stated earlier, from our discussion with legal experts from UK and USA, we understand these problems are not specific to India and developing generalized methods will greatly benefit the law community worldwide.

Ongoing Work

This section briefly describes the work that I have done in the first year of my PhD, which I am continuing at present.

1. Modelling the Indian Legal Judiciary and understanding Legal Document Similarity

   We are presently constructing a knowledge graph to learn models for understanding whether two legal case documents are similar or not.

2. Segmentation and Summarization of Indian case documents

   (paper published: A Comparative Study of Summarization Algorithms applied to Legal Case Judgments, in 41st European Conference on Information Retrieval (ECIR), 2019)

   Although some methods have been developed for summarization of legal documents of other countries, we find that these methods do not generalize well to summarizing Indian court case documents. Additionally there are neural network-based supervised methods and classical unsupervised methods for general text summarization, that can potentially be used to segment and summarize case documents We also explored whether these algorithms can be used for legal document summarization. Through this study, we understood that legal document summarization has some special needs, because it requires concise information of each

segment (facts, background, argument, final judgment etc) to be present in the summary, and so segmentation forms an integral part of the process. Having explored the limitations of existing algorithms, we now plan to develop better summarization algorithms for legal documents. Especially we aim to design approaches that can generalize across documents of several countries.

References

[1] Wyner et.al., 2008 : An ontology in OWL for legal case-based reasoning in Artificial Intelligence and Law

[2] Saravanan et.al., 2009 : Improving legal information retrieval using an ontological framework in Artificial Intelligence and Law

[3] Araujo et.al., 2017 : Ontology-based information extraction for juridical events with case studies in Brazilian legal realm in Artificial Intelligence and Law

[4] Al-Kofahi et.al., 2001: A Machine Learning Approach to Prior Case Retrieval in International Conference on AI and Law

[5] Kumar et.al., 2011: Similarity Analysis of Legal Judgments in ACM COMPUTE

[6] Kalyanasundaram et.al., 2016 : Analysis for Extracting Relevant Legal Judgments using Paragraph-level and Citation Information in AI4J – Artificial Intelligence for Justice, Workshop at the 22nd European Conference on Artificial Intelligence (ECAI 2016)

[7] Mandal et.al., 2017 : Measuring Similarity among Legal Court Case Documents in ACM COMPUTE

[8] Farzindar et.al., 2004 : Letsum, an automatic legal text summarizing system in JURIX

[9] Saravanan et.al., 2008 : Improving Legal Document Summarization Using Graphical Models in JURIX

[10] Polsley et.al., 2016 : CaseSummarizer: A System for Automated Summarization of Legal Texts in COLING (Demo)

[11] Hachey et.al., 2006 : Extractive summarisation of legal texts in Artificial Intelligence and Law

[12] Šavelka et.al., 2018 : Segmenting U.S. Court Decisions into Functional and Issue Specific Parts in JURIX

[13] Fowler et al., 2007 : Network analysis and the law: Measuring the legal importance of precedents at the US Supreme Court in Political Analysis 15.3 (2007): 324-346

[14] Winkels et.al., 2011 : Determining Authority of Dutch Case Law in JURIX

[15] Winkels et.al., 2012 : Survival of the Fittest: Network Analysis of Dutch Supreme Court

Cases in AICOL Workshops 2011

[16] Winkels et.al., 2013 : Creating Context Networks in Dutch Legislation in JURIX

[17] Zhang et.al., 2007 : Semantics based Legal Citation Network in ICAIL

[18] Mazzega et.al., 2009 : The Network of French Legal Codes in ICAIL

[19] Winkels et.al., 2014 : Towards a Legal Recommender System in JURIX

[20] Wang et.al., 2018 : Modeling Dynamic Pairwise Attention for Crime Classification over Legal Articles in SIGIR

[21] Bevan et.al., 2018 : Efficient and Effective Case Reject-Accept Filtering: A Study Using Machine Learning in JURIX

[22] Collarana et.al., 2018 : A Question Answering System on Regulatory Documents in JURIX

[23] Quaresma et.al., 2005 : A question-answering system for Portuguese juridical documents in ICAIL

[24] Bhattacharya et.al, 2019 : A Comparative Study of Summarization Algorithms applied to Legal Case Judgments, in 41st European Conference on Information Retrieval (ECIR), 2019

[25] Wyner et.al., 2017 : Recognizing Cited Facts and Principles in Legal Judgements in Artificial Intelligence and Law

[26] Mandal et.al., 2017 : Automatic Catchphrase Identification from Legal Court Case Documents in CIKM.

[27] Shi et.al., 2015 : A Survey of Heterogeneous Information Network Analysis in IEEE Transactions on Knowledge and Data Engineering

[28] Allahyari et.al., 2017 : Text summarization techniques: a brief survey in arXiv preprint arXiv:1707.02268

[29] Chakraborty et.al., 2015 : On the categorization of scientific citation profiles in computer science in Communications of the ACM

[30] Abujabal et.al., 2018 : Never-Ending Learning for Open-Domain Question Answering over Knowledge Bases in World Wide Web Conference (WWW)