Towards Reducing the Pendency of Cases at Court: Automated Case Analysis of Supreme Court Judgments in India

Shubham PANDEY [a], Ayan CHANDRA [a], Sudeshna SARKAR [a] and
Uday SHANKAR [a]

[a] *Indian Institute of Technology Kharagpur, Kharagpur, West Bengal 721302, India*

Abstract. The Indian court system generates huge amounts of data relating to administration, pleadings, litigant behaviour, and court decisions on a regular basis. But the existing Judiciary is incapable of managing these vast troves of data efficiently that causes delays and pendency of a large volume of cases in the courts. Some of these time-consuming tasks involve case briefing, examining the legal issues, facts, legal principles, observations, and other significant aspects submitted by the contending parties in the court. In other words, computational methods to understand the underlying structure of a case document will directly aid the lawyers to perform these tasks efficiently and improve the overall efficiency of the Justice delivery system. Application of Computational techniques (such as *Natural Language Processing*) can help to gather and sift through these vast troves of information, identify patterns, extract the document structure, draft documents and make the information available online.

Traditionally lawyers are trained to examine cases using the *Case Law Analysis* approach for case briefing. In this article, the authors aim to establish the importance and relevance of the automated case analysis problem in the legal domain. They introduce a novel case analysis structure for the supreme court judgment documents and define twelve different case law labels that are used by legal professionals to identify the structure. Finally the authors propose a method for automated case analysis, which will directly aid the lawyers to prepare speedy and efficient case briefs and drastically reduce the time taken by them in litigation.

Keywords. Law and Technology, AI in Law, Natural Language Processing (NLP), Legal Document Analysis, Case Analysis

1. Introduction

Case Analysis refers to the study of court cases and drawing essential conclusions on how courts have applied certain norms and how they interpret them in accordance with law. This is one of the most crucial parts of the training of all legal professionals to understand a judgment given by a court. Naturally, to analyse the cases heard at the Indian Courts, Case Law Analysis (CLA) technique is used by the lawyers. Thus, an automated approach to find different case analysis roles from a judgment document can directly help the lawyers in their study and preparing the case briefs. Furthermore, identifying the case analysis roles of the sentences in a judgment can also aid in many downstream tasks

such as semantic search, summarization etc. However, there is a set of reasons which make this task computationally difficult such as the length of the judgment text, lack of understanding of the domain knowledge incorporated in the computational approaches etc. We introduce and define different case analysis roles with examples in the upcoming sections.

In this work, we aim to achieve the following objectives: (a) *Establishing case analysis as an important problem in the legal domain*. (b) *Defining a case law analysis structure for the supreme court judgment documents*. (c) *Proposing an example automated case analysis system for the legal community*.

1.1. Importance of the Automated Case Law Analysis

The pendency of a large number of cases is a long ailing problem of the Indian Judiciary. The Indian court system consists of the Apex Court or the *Supreme Court of India*, the *High Courts* in the individual states and union territories, and the *lower courts* at the district level. The recent data estimates (as projected by National Judicial Data Grid [1] on August 21, 2021) show that a total of 36,780,460 cases are pending in the various courts. Even though the lower courts dispose of more than half of the new cases filed (56%) within a year, a significant number of these cases are either transferred to the higher courts, or the verdicts are challenged in the higher courts. Thus, it increases the overall volume of pending cases in the High courts and the Supreme court. Furthermore, the ongoing COVID-19 pandemic has impacted the judicial system severely causing further delays in the Justice delivery process.

Researchers, lawyers, government, and administration are working tirelessly to find ways to reduce the pendency of cases and improve the overall efficiency of the legal system. For every case, the advocates of the contending parties have to submit/put forward arguments on behalf of the respective parties, after going through all similar prior documents and the legal principles. Based on the available information on the current case and a set of similar prior documents (also known as *"precedents"*), they prepare a case briefing. Often, these case documents are thousands of pages long and it takes a considerable amount of time for them to prepare the briefing. Again, when the judge gives the verdict of the presented case, all the legal issues, facts, arguments provided by the advocates are examined by the court. For both processes, case analysis is one of the most used techniques or methods available to legal practitioners. Therefore, proposing approaches to solve the task of automated case analysis will directly improve the efficiency of the courts and help in mitigating the bigger problem by reducing the pendency of cases.

In this work, we define a case analysis structure for the supreme court judgment documents with the help of the legal professionals and propose a rule-driven system by which one can perform the task. This is a significant contribution towards the legal community to reduce the pendency and backlogs in the court-rooms.

2. Limitations of Prior Work: Comparison between Case Analysis and Rhetorical role labelling

Earlier works [2] [3] assumed that case analysis can be performed using Rhetorical Role labeling of the legal document. Rhetorical role labeling of a sentence from a legal text is

a means to understand the semantic function of the sentence (such as arguments of the parties, the final judgment, background, statute and so on).

When we compare the rhetorical roles and the case law analysis information that a legal professional requires, we observe various limitations. We note that: **(i)** Certain key information labels such as "Legal issue" and "Observation" are neither present in the existing rhetorical structure nor can be derived from any of those. **(ii)** The information whether a sentence is an argument is not sufficient to a lawyer, but whether the argument is submitted by the appellant, respondent, or amicus is important. Furthermore, the party or advocate is mentioned once in the text and thereafter referred to by appropriate pronouns in consecutive sentences. Hence, both the semantic and case-related auxiliary information (named entities) and their anaphora and coreferences [5] are equally important in case analysis. **(iii)** Any mention of established law (e,g, Acts, Sections, Articles, Rules, Notices, etc.) is treated as a statute. However, not all the mentions of established laws in a judgment text are relevant to the analysis of the given case. In particular, the set of legal principles which are discussed for the current judgment are deemed important. Legal principles may include not only the established laws but also the mention of enumerated rights and unenumerated rights. e.g. "Right to clean air", "Right to education" are not statutes but these are some of the examples of enumerated rights which can be applied in a judgment by the court, and mentioned with its reason for application. Hence, the understanding of legal principle is different from that of the rhetorical role "statute". **(iv)** For long-running cases, the court often provides various interim orders to provide relief to one of the contending parties. These relief prayers asked by the parties, interim orders given by the court from time to time, and whether parties have complied with earlier orders become a crucial subject of examination in later judgments. Thus, relief prayer, interim order, and compliance also become part of the reasoning, and therefore, demands a separate label in the case analysis process.

Above observations eventually lead us to establish the automated case analysis of judgment documents as an important task in the legal domain different from the rhetorical role labeling.

3. Case Analysis Labels

After going through a large volume of court documents, we have come up with the following twelve case analysis roles. We describe the roles in the following:

1. **Legal Issues**: The legal issue or question that the Court is adjudicating upon in the current case. e.g. *"The issue for our consideration today is fixing standards for 34 industries with regard to the SO2, NOX and SOX emissions."*
2. **Argument by Appellant:** Different arguments submitted by the appellants or the advocates on their behalf in the court. e.g. *"Mr. Sundaram finally submitted that since none of the grounds given by the High Court in the impugned judgment for directing closure of the plant of the Appellants are well-founded, ..."*
3. **Argument by Respondent:** Different arguments submitted by the respondents or the advocates on their behalf in the court. e.g. *"It was pointed out by Shri Mukul Rohatgi, learned senior Counsel appearing on behalf of Government of NCT of Delhi that by applying odd-even scheme with respect to cars alone cannot be said to be a wholesome*

solution."

4. **Argument by Amicus Curiae:** Sometimes, the court invites third party members or organisations to provide insights to a case that is being discussed. Such third parties are legally known as "Amicus Curiae". Amicus curiae may also submit its arguments through its advocates. e.g. *"Mr. Harish N. Salve, learned senior Counsel appearing as amicus, argued that imposition of ECC and the directions issued by this Court regarding diversion of commercial vehicles/trucks to alternative routes has made some difference.."*

5. **Relief Prayer:** If a case is continuing for long duration, one of the parties may seek for immediate relief from certain inconveniences, before the final verdict comes out. e.g. *"The Petitioners have approached this Court seeking emergent reliefs in relation to the extreme air pollution in the National Capital Region (hereinafter 'NCR')"*

6. **Observation Findings:** Observations made by the court while assessing the facts. There is no immediate relation between the observation and the conclusion of a case. But these observations may be important in understanding how the court weighs the arguments submitted by different parties. e.g. *"We must note that there has been no response from the States within the NCR giving the impression that air pollution is not a problem for the State Governments despite the ill-effects and health hazards of bursting fireworks"*

7. **Legal Principles:** In the judgment, the court may state different Enumerated and Unenumerated rights, established laws (statutes, acts, constitution) etc. These principles are crucial for establishing the conclusion of the judgment. Enumerated rights are the rights given by the established laws and constitution (e.g. fundamental rights). Unenumerated rights are the rights that are coming from case laws [earlier verdicts from different important cases.] , and may become part of established laws later. In general, a sentence that represents a legal principle does not include any mention of appellant, respondent or any parties involved. e.g. *"The polluter pays principle demands that the financial costs of preventing or remedying damage caused by pollution should lie with the undertakings which cause the pollution, or produce the goods which cause the pollution"*

8. **Fact:** Fact refers to the chronology of events that led to filing the case, and how the case evolved over time in the court system. Facts are the sentences that have pieces of information about appellant/respondent/other parties involved or the current court but it does not include any argument, legal principle, verdict, direction or causation in itself. e.g. *"The Appellants are the owners of Hotels, Beach Resorts and Beach Bungalows in Goa who have been facing the prospect of demolition of their properties for the last several decades"*

9. **Rationale:** Rationale is the reasoning on how legal principle and facts can have a causation relationship. A rationale has two parts in it: One part includes a Legal Principle, and the other part includes a Fact. e.g. *"Under the principle of 'delegatus non potest delegare', the delegatee (the Chairman of the Board) could not have further delegated the authority vested in him, except by a clear mandate of law."*

10. **Conclusion Verdict:** The decision or verdict, conclusion of the Court, or, texts that can be considered as part of the verdict. If the document is of type Order, then orders or directions mentioned in the text can be considered as parts of the final verdict. e.g. *"we direct the States of Punjab, Haryana and Uttar Pradesh to disburse the money and they should not wait for or write letters to the Central Government to give certain funds for this purpose."*

11. **Interim Order:** If a case is continuing for long duration, the court may issue interim

orders to grant relief to the parties temporarily if the parties seek relief. In general, interim order can be found along with relief prayer in an Order document. e.g. *"We accordingly direct the Government of NCT of Delhi to take immediate steps for repair of pavements and make pavements wherever the same are missing and also to take immediate steps for ..."*

12. **Compliance:** Compliance signifies if one of the parties involved in the case has complied to the directions given by court or any earlier order or direction given by other organisations. A non-abiding of compliance may become a part of the Observation which may influence the verdict. e.g. *"Unit has complied with the conditions and the consent order issued to the Unit."*

4. Supreme Court of India corpus of Air Pollution cases

We create a corpus of 28 Supreme Court of India corpus of Air Pollution cases annotated with the case analysis structure. We consider the legal judgments that are related to Air Pollution cases under the Environment category from the Supreme court of India cases gathered from Manupatra [6], a reputed online journal for the period January 2010-January 2021. We take the cases which are "Against the government" i.e. Union of India is one of the respondents in those cases. No proprietary information is taken in this process.

Initially one annotator with domain understanding annotated the documents. Each sentence from the judgment documents is labelled with a single argument label from the set of 12 case analysis roles and "Other" (if the sentence is not useful in the case analysis process). Then a panel of legal experts reviewed the annotations and rectified it wherever necessary based on consensus decision. As the review of annotation was conducted based on the unanimous decision of a panel of legal professionals from *Rajiv Gandhi School of Intellectual Property Law, IIT Kharagpur* [1] with varied academic and court experiences, the quality of the annotation may be considered to be high.

5. System Overview

For the identification of the case analysis roles, the judgment copy is required as input. A judgment copy contains the judgment text, list of appellants, respondents, amicus if any, advocates of the contending parties, name of the judges, file number of the case, etc. Our proposed automated case analysis system consists of the following of steps.

Step 1: Extraction of the case metadata - The case metadata includes the date of the decision, the document type (e.g. Judgment, Order, etc.), the name of the judges, party names, and the name of the advocates of the parties. We segregate the case metadata from appropriate places in the judgment copy using position-based rules and store the segregated components in a configuration dictionary. We also consider various abbreviations and short names for party names and advocates in the dictionary.

Step 2: Entity Identification - In the judgment text, named entities such as the name

---

[1]RGSOIPL, IIT Kharagpur: http://www.iitkgp.ac.in/department/IP

of the appellants or respondents, name of the advocates, etc. are mentioned once and thereafter referred by appropriate pronouns in consecutive sentences or before the entities in the same compound sentence. In linguistics, this is known as co-reference and if the pronoun that refers to the entity comes after the mention of the entity, then it is called an anaphora. In order to understand which named entity that pronoun refers to, anaphora and co-reference resolution [5] are required. e.g. in the sentence *"He submits that the depending upon the improvement and the extent of collection of ECC, post installation of RFID, this Court could issue appropriate directions suitably balancing the equity among the State and the stakeholders."* , "He" refers to "Mr. Salve, learned Amicus" who was one of the advocates of the amicus in the given example. We use Neuralcoref [4] for the co-reference resolution as it performed well on our corpus. We replace each mention of the parties and their advocates, court, etc. in the resolved text with corresponding identifier tokens using the configuration dictionary. These identifier tokens will be used in different rules to identify the case analysis role.

Step 3: Syntactic Analysis - Syntactic structure of a sentence (such as the nature of the verb, and the relation of different noun phrases with the verb) plays a key role in constructing various rules for role identification. So, we perform parts of speech (POS) tagging and dependency parsing on the sentences from the text. We use Spacy [8] for both the tasks because it showed better results which agrees with [7].

Step 4: Rules Identification - Based on the annotated examples and the domain knowl-

Figure 1. Flow chart for determining the case analysis role of a sentence from a given judgment text.

edge of the legal experts, we attempt to find the rules for each of the case roles. In Figure

Table 1. Precision, Recall, and F-score for individual case analysis role from law documents of test set predicted by our method

Case analysis role	P	R	F1	Support
Observation Findings	1.00	0.61	0.76	18
Conclusion Verdict	0.95	0.80	0.87	45
Relief Prayer	1.00	1.00	1.00	6
Argument by Amicus	1.00	0.62	0.77	8
Argument by Appellant	1.00	0.77	0.87	22
Argument by Respondent	1.00	1.00	1.00	10
Fact	0.97	0.83	0.90	132
Interim Order	1.00	1.00	1.00	1
Legal Principle	0.90	1.00	0.95	9
Legal Issue	1.00	1.00	1.00	3
Rationale	0.40	1.00	0.57	2
Compliance	1.00	1.00	1.00	1

1, we describe the process by which the case analysis role of a sentence is determined. We use a combination of various regular expressions and syntactic analysis in this step.

We find the rules after extensive study of 24 documents from the corpus with the help of the legal experts and report the results on 4 additional documents. The rules for the labels are made available for the interested readers [2].

6. Results

We report the Precision (P), Recall (R), and F1-score (F1) for each of the case analysis roles on the test set of 4 documents in Table 1. For some of the case roles, precision and recall are quite high. We discuss about the issues with the existing rules in the following subsection. A list of examples are given in the supplementary material for the readers[3].

6.1. Issues:

We went through the set of 24 legal documents and test set of additional 4 documents from the corpus and manually examined the case analysis roles of the sentences. Following issues are observed: **(i)** In some of the paragraphs, the third person pronouns are not properly co-referenced to the appropriate named entities. So, those sentences do not have any appropriate identifier token (e.g. appellants, etc.). As a result, the rules-identification module fails to identify these sentences properly. If we improve the co-reference resolution model, this issue can be mitigated for most of such examples. **(ii)** In case of interim orders, it was difficult to evaluate whether a sentence can be identified as an interim order or a conclusion, even though reliefs are asked. **(iii)** In certain documents, all the parties involved or heard in the case are not mentioned in the judgment copy beforehand. E.g. SDMC (South Delhi Municipal Corporation) and NDMC (North Delhi Municipal Corporation) are not mentioned as any party in the case MANU/SC/0609/2016 . When the rules identification module encounters the sentence *"These applications have been filed*

---

[2]Rules Description: https://github.com/legalArgMining/Case-Law-Analysis
[3]Same as above

by SDMC and NDMC seeking permission for registration of diesel vehicles used for collection and transportation of garbage on diesel based HCV and MCV vehicles.", it could not classify it into any existing roles and labels it as *"Other"*. However, a manual examination of the case conveys that both SDMC and NDMC are parties involved in the case. Therefore, this sentence should be treated as a "Fact". **(iv)** For certain sentences, "Fact" has been mis-predicted as "Rationale". The rules identification module rightly found the presence of both a legal principle and a Fact part in the sentence. However, the sentence conveys a generic statement and violation of certain legal principle and therefore, should be classified as a "Fact". For such cases, we need to incorporate specific exclusion rules. **(v)** In a few documents, the subject of the head word in the sentence is not explicitly mentioned. Such sentences are usually in passive voice. E.g. *"It is also submitted that.."* . In this sentence, the *agent* "by" of the head word verb "submitted" is not mentioned. As a result, the rules fail to identify these sentences properly.

7. Conclusion and Future Works

In this work, we introduce a case analysis structure for the Indian supreme court judgment documents with suggestions from the legal community, and show some initial reports for automated case analysis role identification to build the structure. Our methods of case analysis role identification are mostly a combination of different rules. However, many of these rules can be good indicators of possible features to be used in machine learning based methods. Exploring the task with more annotated examples using machine learning based models remains as a future direction of work. Also, our work centered around the Air Pollution cases. How the rules identification module performs across various categories of cases remains as a future task to be examined.

References

[1] National Judicial Data Grid (India) https://njdg.ecourts.gov.in/njdgnew/index.php

[2] Bhattacharya P, Paul S, Ghosh K, Ghosh S, Wyner A: *"Identification of Rhetorical Roles of Sentences in Indian Legal Judgments"* In: International Conference on Legal Knowledge and Information Systems (JURIX) 2019 ; Available from: https://arxiv.org/abs/1911.05405

[3] Walker VR., Pillaipakkamnatt K, Davidson AM., Linares M and Pesce DJ. *"Automatic Classification of Rhetorical Roles for Sentences: Comparing Rule-Based Scripts with Machine Learning"*. In: Proceedings of the Third Workshop on Automated Semantic Analysis of Information in Legal Texts, Montreal, QC, Canada, June 21, 2019 2019 . CEUR-WS.org.

[4] Neuralcoref, Coreference Resolution tool from Huggingface; Available from: https://github.com/huggingface/neuralcoref

[5] Gupta A, Verma D, Pawar S, Patil S, Hingmire S, Palshikar GK, Bhattacharyya P *"Identifying Participant Mentions and Resolving Their Coreferences in Legal Court Judgements."* In: Text, Speech, and Dialogue - 21st International Conference, TSD 2018, Brno, Czech Republic, September 11-14, 2018, Proceedings 2018 (pp. 153–162). Springer.

[6] Manupatra, online legal journal https://www.manupatrafast.com/

[7] Ma M, Podkopaev D, Campbell-Cousins A, Nicholas A. Deconstructing Legal Text: Object-Oriented Design in Legal Adjudication. MIT Computational Law Report [Internet]. 2020 Nov 20; Available from: https://law.mit.edu/pub/deconstructinglegaltext

[8] Spacy, Industrial-Strength Natural Language Processing; Available from: https://spacy.io/