Chapter 11
Indian Legal Documents Corpus for Court Judgment Prediction and Summarization: CJPS

Ayushi Gautam
Vellore Institute of Technology, Chennai, India

Vishali Sharma
Vellore Institute of Technology, Chennai, India

Piyali Saha
https://orcid.org/0009-0001-1794-3334
Vellore Institute of Technology, Chennai, India

Atul R. Patel
Vellore Institute of Technology, Chennai, India

Sujithra Kanmani
Vellore Institute of Technology, Chennai, India

ABSTRACT

We propose the development of an automated system aimed at assisting judges in predicting case outcomes to streamline the judicial process. To ensure practical utility, this system should provide explanations for its predictions. To support research in this area, we introduce the Indian Legal Documents Corpus (ILDC), a substantial collection comprising 35,000 Indian Supreme Court cases, annotated

DOI: 10.4018/979-8-3693-6230-3.ch011

with their original court decisions. Additionally, a subset of this corpus serves as a test set, annotated with expert-generated explanations for reference. Building upon the ILDC, we present the Court Judgment Prediction (CJP) task, which involves the automated prediction of case outcomes. We conduct experiments using various baseline models for case prediction and introduce a hierarchical occlusion analysis of explanations generated by our proposed algorithm. This analysis reveals a notable disparity between the algorithm's perspective and that of legal experts when explaining judgments, suggesting promising avenues for future research in this domain.

1. INTRODUCTION

Drawing from the Indian Legal Documents Corpus

(ILDC), we propose a novel task called COURT JUDGMENT PREDICTION AND SUMMARIZATION (CJPS). This task is designed to predict the final decision reached in a legal case, considering all the presented facts and arguments, and simultaneously providing a concise summary that justifies the predicted decision. The possible decisions are either "allowed," indicating a ruling favouring the appellant/petitioner, or "dismissed," indicating a ruling favouring the respondent. The CJPS task presents unique challenges that set it apart from conventional text classification tasks. Firstly, legal court case documents, particularly in the Indian context, are characterized by their unstructured nature, verbosity, and noise. Extracting and directly utilizing the relevant facts and arguments from these documents is far from straightforward. Secondly, the domain-specific vocabulary and terminology used in legal cases render models pre-trained on general text ineffective when applied to such documents.

Consequently, standard models must be adapted to the legal domain to excel in the judgment prediction aspect of the proposed task. Thirdly, providing concise and accurate summarizations in legal documents is notably challenging, as it demands a deep understanding of the presented facts, logical progression of arguments, and the application of legal rules and principles to arrive at the final decision.

We introduce the CJPS task, comprising two interrelated sub-tasks: (a) Court Judgment Prediction (CJP) and (b) Summarization of the Prediction. While CJP is not entirely novel in isolation, the synergy between judgment prediction and summarization within the CJPS task presents a fresh and complex challenge. Furthermore, the requirement for automated summarization places constraints on the methodologies applicable to CJP. In the CJPS task, we intentionally withhold gold summarizations in the training set, expecting that the trained algorithms should generate concise and relevant summaries without the need for additional annotated data during their training.