

Abstract

Legal document summarization is an important problem since the documents are very long containing dense, complex text, and Law practitioners have to go through many legal documents in order to formulate their arguments for a given case. As a result, there has been decades of research on this problem, and a wide variety of summarization models have been tried on legal documents, starting from traditional extractive models to recent generative models. This thesis aims to advance the understanding and performance of legal document summarization by focusing on three key objectives. First, it conducts a detailed analysis of extractive summarization methods applied to legal case judgments, comparing algorithmic sentence selections with those made by legal experts. This includes an exploration of the positional and rhetorical characteristics of selected sentences and the use of both traditional metrics (e.g., ROUGE) and rhetorical role analysis to assess alignment. Second, the thesis investigates ensemble-based approaches to extractive summarization, exploring whether combining multiple algorithms through voting, ranking, or graph-based techniques can yield more effective summaries than individual methods. Finally, it evaluates the performance of generative summarization models – including both domain-specific abstractive summarization models and general-purpose large language models (LLMs) – in capturing the essence of legal judgments. This involves assessing summary quality, factual consistency, and hallucination rates to determine the practicality and limitations of LLMs for summarization in the legal domain. Together, these objectives contribute to a deeper understanding of both extractive and abstractive summarization models in a high-stakes domain. The thesis also proposes ways to improve legal document summarization, e.g., by ensembling extractive models, and reducing hallucinations in generative/abstractive models.