Applying Reinforcement Learning to Legal Document Summarisation

Vipul Mishra

August 8, 2024

1 Introduction

Legal documents are often lengthy and complex, making it challenging for legal professionals, researchers, and laypeople to extract the most pertinent information quickly. Automatic text summarization can help, but traditional methods may struggle with the nuances and importance of certain legal terms and contexts. In this report, we explore how Reinforcement Learning (RL) can be leveraged to create a more context-aware summarization system that learns to prioritize the most relevant information, ensuring the summary maintains the document's legal integrity.

2 Problem Description

Legal document summarization is the process of creating a concise and accurate representation of the most critical information in a legal text. The primary challenge lies in maintaining the legal integrity and contextual relevance of the document while reducing its length.

3 Why Reinforcement Learning?

Reinforcement Learning is suitable for this task because it allows the model to learn from feedback and improve over time. Traditional supervised learning models require large labeled datasets, which can be difficult to obtain in legal contexts. With RL, the model can learn to optimize the summary based on feedback, such as how well the summary preserves key legal information, instead of relying solely on predefined summaries. RL's trial-and-error approach enables the model to explore various strategies for summarization, potentially leading to more efficient and legally accurate results.

4 Environment Description

The environment in this RL setup consists of legal documents of varying lengths and complexities. The agent interacts with these documents, aiming to generate summaries that capture the most critical information while adhering to legal standards.

- **Reward Signal:** The reward can be based on multiple factors, including similarity to expert summaries (using ROUGE scores), preservation of critical legal points, and feedback from legal professionals.
- **Episodes:** Each episode could be a single document that the agent needs to summarize.

5 State Space Description

The state space represents the current state of the document being summarized. Each state could be a combination of:

- The current position in the document.
- The sentences or clauses already selected for the summary.
- Metadata such as sentence length, frequency of legal terms, or importance scores from a pre-trained model.

The state space would be dynamic as the agent progresses through the document, adjusting based on previous actions.

6 Action Space Description

The action space consists of:

- Selecting the next sentence or clause to include in the summary.
- Skipping a sentence.
- Terminating the summarization process when the agent believes the summary is complete.

Each action leads to a new state, impacting the document's summary.

7 Example of Agent Interaction

Consider a legal document with several clauses. The agent starts at the beginning, analyzing the first clause:

- **State:** Start of the document, no sentences selected.
- **Action:** The agent decides to include the first clause in the summary.
- **New State:** First clause selected, the agent now considers the second clause.

The agent continues this process, receiving rewards based on the quality and completeness of the summary.

8 Learning Algorithm: Proximal Policy Optimization (PPO)

Proximal Policy Optimization (PPO) is a robust RL algorithm well-suited for this task because it balances exploration and exploitation, making it less likely to diverge during training. The agent can learn to optimize the summary by gradually improving its policy based on feedback without making overly large updates, which is crucial when dealing with the sensitive and nuanced information in legal documents.

8.1 Algorithm Outline

- **Initialize** policy and value function networks.
- **For each episode** (document):
  - Collect trajectories by interacting with the document.
  - Compute rewards based on the quality of the summary.
  - Update the policy using PPO's objective, ensuring updates are within a safe range.
- **Repeat** until the policy converges to an optimal summarization strategy.

9 Alternative Approach: Transformer-based Summarisation

Beyond RL, a Transformer-based model like BART or PEGASUS can be fine-tuned on a dataset of legal documents to perform summarization. These models have shown state-of-the-art results in various NLP tasks and could serve as a strong baseline or complement to the RL approach. The Transformer model would handle the summarization in a more deterministic way, relying on learned patterns from the training data.

10 Conclusion

This report outlines an innovative application of RL to legal document summarization, emphasizing the importance of preserving legal integrity while efficiently condensing information. By leveraging the trial-and-error nature of RL, the model can adapt to the unique challenges of legal texts, potentially offering a more effective solution than traditional summarization methods.