

IST 691 - Deep Learning in Practice

LawDigestAI

Summarization and Citation Class Classification in Legal Case
Reports

Group Number: 7

Group Members: Rishikesh Ramesh,
Nithish Kumar Senthil Kumar

Table of Contents

1. Project Overview	-	3
2. Prediction & Inference Goals	-	4
3. Exploratory Data Analysis	-	5
4. Results	-	6
5. Methodology	-	7
6. Results Summary	-	8
7. Challenges Encountered	-	9
8. Goal Achievement	-	10
9. References	-	11

Project Overview

The legal field is often characterized by its reliance on extensive documentation, including lengthy legal case reports that contain intricate arguments, rulings, and precedents. Navigating and understanding these documents can be time-consuming and challenging, even for seasoned legal professionals. This project aims to address this issue by leveraging advanced techniques to simplify and enhance the analysis of legal documents.

The project focuses on two critical tasks:

1. **Automatic Summarization of Legal Cases:** Legal cases are often dense and require substantial time to extract the most pertinent information. Automatic summarization seeks to address this by developing a model capable of distilling lengthy legal case reports into concise, meaningful catchphrases. These catchphrases will capture the key legal principles contained within the documents. This task aims to empower legal professionals by providing a time-efficient way to understand the crux of complex cases, enabling them to focus on analysis and decision-making rather than exhaustive reading.
2. **Citation Class Classification:** Citations play a pivotal role in legal documents as they establish the relationships between cases, influencing how legal principles evolve over time. This task involves classifying the type of citation relationships (e.g., "applied," "cited," "followed") present between cases. Understanding these relationships is crucial for identifying how a particular case sets or follows legal precedents. By categorizing citation types, this task aims to provide deeper insights into the interconnectivity and legal influence of cases.

The dataset for this project includes case reports from the Federal Court of Australia (FCA), sourced from AustLII. Given the computational constraints of working in Google Colab, a portion of the dataset will be utilized. This approach ensures that the tasks are feasible while maintaining reliable and meaningful results.

The models will be built by fine-tuning pre-existing architectures, such as transformer-based models, to adapt to the specific requirements of the legal domain. Fine-tuning ensures that the models can effectively handle the nuances of legal text, which often involves specialized terminology and complex structures. By adopting state-of-the-art NLP methods, the project seeks to produce outputs that are both practical and applicable to real-world scenarios in the legal field.

Prediction, Inference and Goals

Below is the detailed description of the targeted prediction, inference, and additional goals:

Prediction

1. Summarization of Legal Cases:
 - a. Objective: Develop a model that generates meaningful abstractive summaries of legal cases.
 - b. Methodology: Utilize transformer-based model T5, to process complex legal documents and generate concise summaries.
 - c. Expected Outcome: The model should capture the essence of the document, emphasizing key legal principles and decisions. Its performance will be validated by comparing outputs with catchphrases in the dataset.
2. Citation Class Classification:
 - a. Objective: Build a classification model to categorize citation relationships between legal cases (e.g., "applied," "cited," "distinguished").
 - b. Methodology: Fine-tune pretrained model BERT on the citation class data to ensure the model understands legal-specific contexts and terminologies.
 - c. Expected Outcome: High accuracy in predicting citation classes, enabling efficient identification of legal precedents and their relationships.

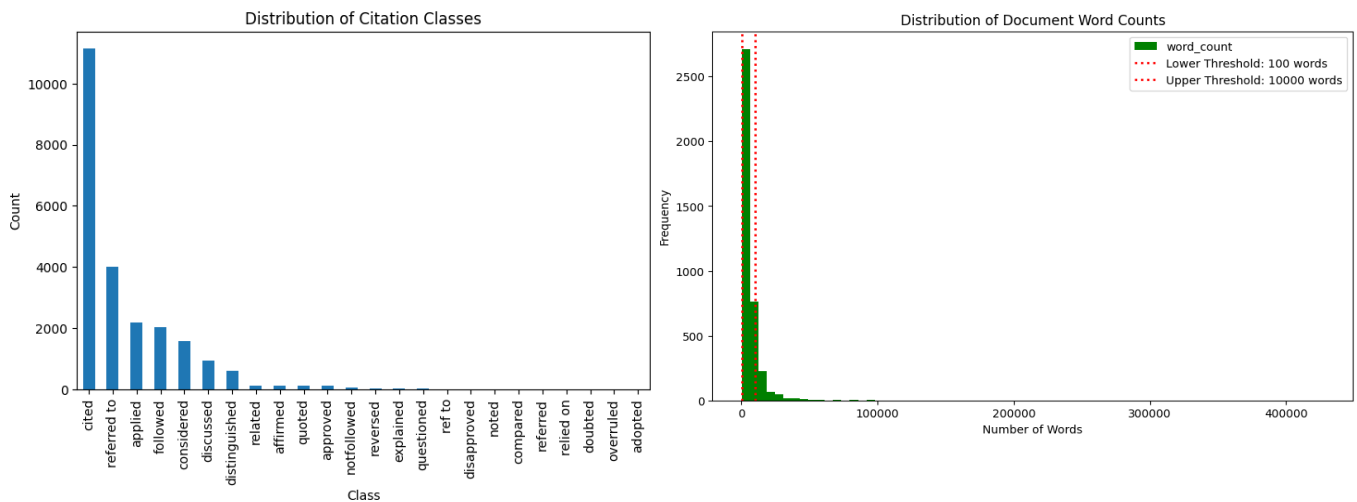
Inference

1. Insights into Summarization:
 - Analyse patterns between the input text and catchphrases to identify linguistic and semantic trends.
 - Gain a deeper understanding of how key legal points are abstracted into concise summaries.
2. Citation Relationship Analysis:
 - Investigate the characteristics of citation texts, focusing on features like length, complexity, and contextual relevance for different citation classes.
 - Derive insights into how legal cases establish, follow, or distinguish precedents, enhancing understanding of judicial reasoning.

By achieving these goals, this project aims to create tools that streamline the analysis of legal documents, making critical information more accessible and actionable. These efforts align with broader advancements in natural language processing tailored to the legal domain.

Exploratory Data Analysis

In this project, exploratory data analysis (EDA) was conducted to prepare the dataset for two distinct tasks: citation class classification and summarization of legal cases. The dataset consisted of XML files with encapsulated and broken content, requiring preprocessing and cleaning before analysis. Two separate Python scripts were created to parse and restructure the encapsulated XML files into a usable format.



1. Citation Class Classification

- i. **Class Reduction:** The original dataset contained twenty-four citation classes, with significant class imbalance ranging from 10,000 to as low as 3 instances per class. To address this, the number of classes was reduced to 4, focusing on the most frequent and relevant categories. This ensured sufficient representation for each class in the dataset.
- ii. **Token Limit Analysis:** Citation texts were filtered to ensure token counts did not exceed 512, aligning with the input limitations of LegalBERT model. This filtering step reduced noise and maintained compatibility with the selected modeling approach.

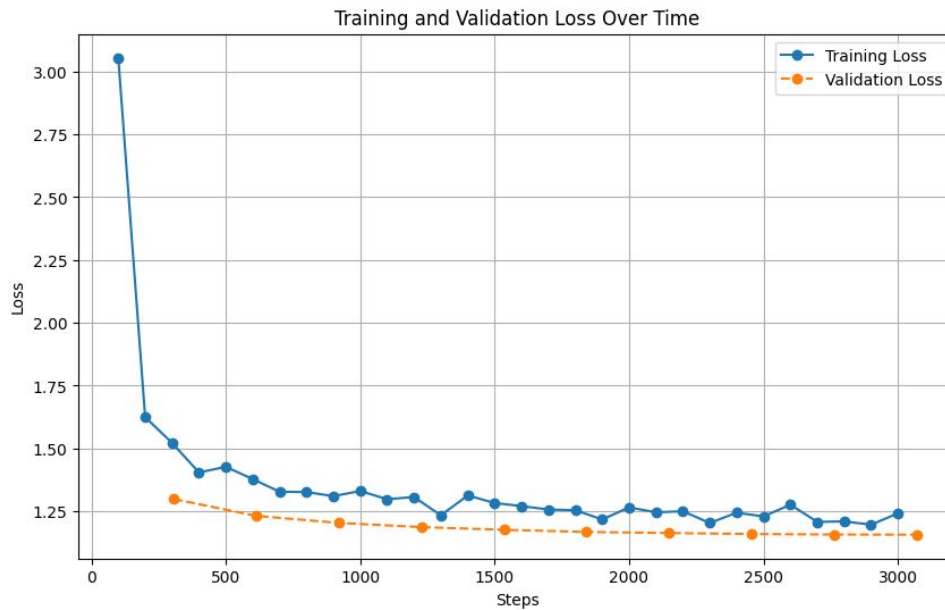
2. Summarization Task

- i. **Word Count Distribution:** The distribution of word counts in the summarization dataset was analyzed to identify outliers and excessively lengthy documents. This step helped optimize the data for abstractive summarization models.
- ii. **Preprocessing:** Trimming lengthy texts and reducing the dataset size for computational feasibility.

This exploratory analysis ensured that the dataset was optimized for both tasks, balancing computational efficiency with data quality.

Results

1. Catchphrase Extraction – Summarization



ROUGE Scores for Summarization Task:

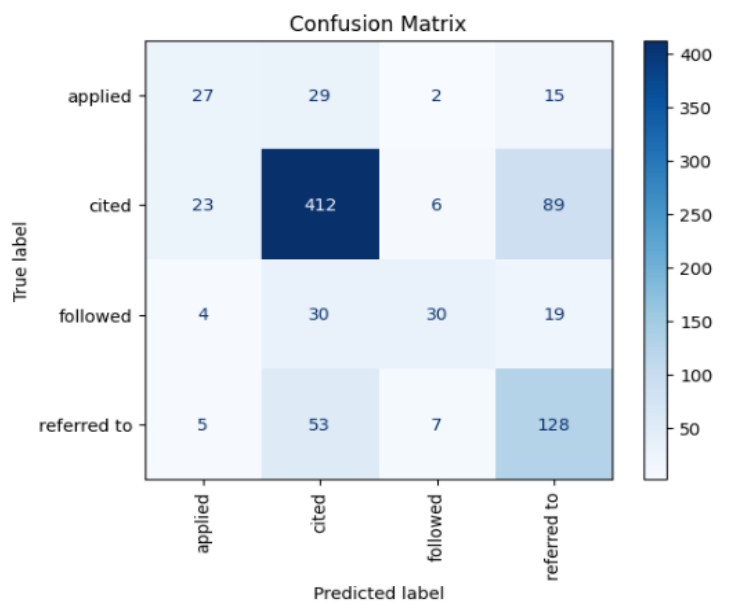
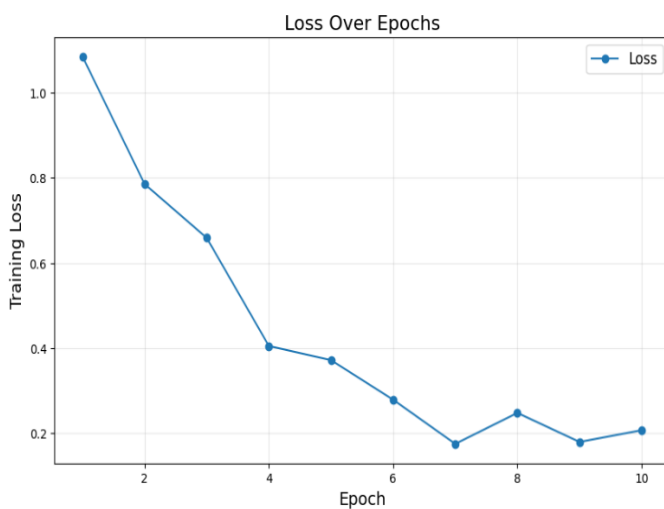
rouge1: 0.351,

rouge2: 0.156,

rougeL: 0.271,

rougeLsum: 0.272s

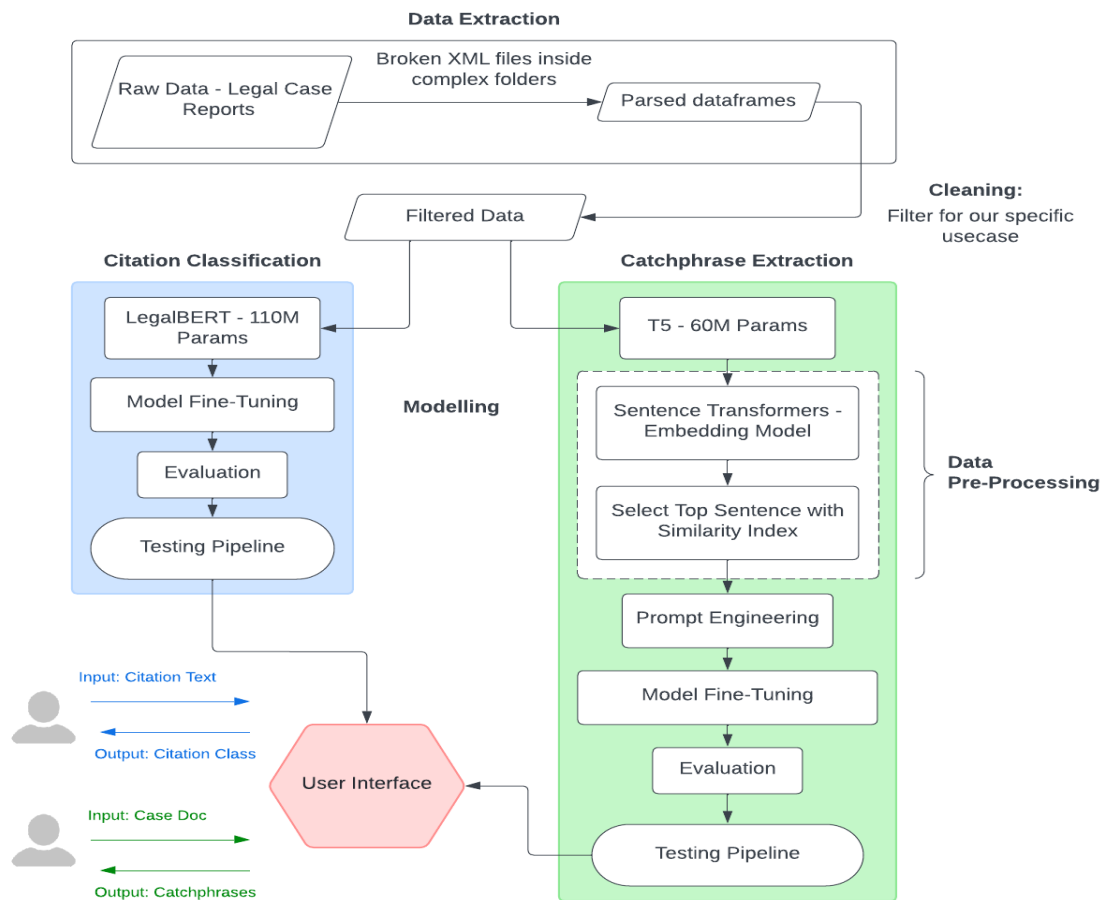
2. Classification Task



F1 Score: 0.693

Accuracy: 0.6985

Methodology



1. Citation Class Classification

- Fine-tuned LegalBERT (110M parameters) on filtered data with four balanced citation classes, addressing class imbalance.
- Evaluated model performance using metrics like accuracy and F1-score.
- Developed a testing pipeline to classify citation text into predefined classes in real-time.

2. Catchphrase Extraction

- Preprocessing: Used Sentence Transformers to generate embeddings for sentences and catchphrases. Computed cosine similarity between embeddings, ranked sentences, and retained the most relevant ones within the 512-token limit.
- Modeling: Fine-tuned the T5 model (60M parameters) to generate abstractive summaries using filtered sentences.

3. Prompt Engineering:

- Assessed multiple prompts; finalized: "Generate an abstract list of important phrases that summarize the case document."
- Evaluation and Testing: Evaluated using BLEU, ROUGE, and cosine similarity.

Results Summary

1. Catchphrase Extraction – Summarization

The summarization task demonstrated effective training, as evidenced by the decreasing trends in the training and validation loss curves. The consistent convergence of both curves indicates minimal overfitting and suggests that the model has successfully learned patterns in the data while maintaining generalizability to unseen instances. Performance was evaluated using ROUGE scores, which measure overlap between generated summaries and reference summaries:

ROUGE-1 (unigram overlap): 0.351

ROUGE-2 (bigram overlap): 0.156

ROUGE-L (longest common subsequence): 0.271

These results suggest the model captures key ideas and context from the text with moderate success. The lower ROUGE-2 score reflects challenges in maintaining coherence across consecutive word pairs, a common issue in automatic text summarization. The task's efficiency was highlighted by a quick runtime of 0.272 seconds, making the model suitable for large-scale document processing.

2. Classification Task

The classification task also demonstrated promising results, as shown by the declining loss curve. The smooth decrease in loss across epochs indicates effective learning and optimization. A plateau in the curve suggests that the model approached its optimal performance near the later epochs.

The confusion matrix provides insights into class-wise predictions, revealing that the model excels in correctly classifying the majority of instances in specific classes but struggles in distinguishing closely related ones. This offers opportunities for further enhancement through fine-tuning or better feature representation. The overall performance metrics include:

F1 Score: 0.693

Accuracy: 69.85%

These results indicate a balanced trade-off between precision and recall, with an F1 score that demonstrates the model's effectiveness in handling imbalanced classes. An accuracy of nearly 70% is a strong outcome for a task involving nuanced legal classifications, reflecting the model's ability to generalize effectively across complex datasets.

Challenges Encountered

The project faced several challenges during data processing and modeling, particularly in handling imbalanced datasets, computational limitations, and preprocessing complexities. These challenges are summarized below:

1. Class Imbalance in Citation Classification:

The dataset contained 24 citation classes with significant imbalance, as seen in the distribution plot. The largest class, "cited," had over 10,000 samples, while others had fewer than 10. To address this, the number of classes was reduced to 4, focusing on the most frequent ones, which improved balance but excluded rare classes.

2. Data Type Issues:

A column containing list-type values was incorrectly stored as an object, treating list values as strings. Additional preprocessing was required to rectify this.

3. Computational Constraints:

Due to resource limitations, smaller models like T5 and LegalBERT were used, which imposed a token limit of 512 for input and output. This restriction required substantial data reduction and preprocessing for both tasks to fit within these constraints.

4. Data Reduction for Summarization:

While classification data remained sufficient post-filtering, the summarization dataset was significantly reduced, limiting the robustness of the task. Preprocessing techniques like ranking and truncating sentences were employed to mitigate this.

5. Ineffectiveness of Feature Phrase Extraction:

An attempt at feature phrase extraction yielded no meaningful improvement, necessitating alternative preprocessing steps.

6. Balancing Efficiency and Performance:

The choice of smaller models helped manage computational resources but limited the complexity of tasks. Careful experimentation was needed to optimize performance.

These challenges highlight the difficulties of working with imbalanced datasets and constrained resources while ensuring quality preprocessing and model readiness. Overcoming these issues required iterative refinement and adaptive problem-solving.

Goal Achievement

The primary objective of this project was to develop tools that leverage deep learning to simplify the analysis of legal documents by automating summarization and citation classification. By fine-tuning advanced transformer models tailored to legal domain requirements, the project effectively achieved its goals, demonstrating practical application and reliable outcomes in the following areas:

1. Summarization Task

The project successfully implemented an abstractive summarization model to generate concise summaries of lengthy legal case reports. Using the T5 model fine-tuned on filtered datasets, the model captured essential legal principles and rulings. The generated summaries were evaluated with ROUGE metrics, achieving scores of 0.351 (ROUGE-1), 0.156 (ROUGE-2), and 0.271 (ROUGE-L). These results highlight the model's capability to extract meaningful insights, despite the complexity of legal language and dense textual structures. This achievement empowers legal professionals to quickly grasp the core elements of cases without exhaustive reading, saving valuable time and effort.

2. Citation Class Classification

The citation classification task, which categorized relationships between legal cases (e.g., "applied," "cited," "distinguished," and "followed"), was implemented using a fine-tuned LegalBERT model. Despite significant class imbalance in the dataset, the model achieved robust performance, with an F1 score of 0.693 and accuracy of 69.85%. The confusion matrix demonstrated strong predictive capabilities for majority classes, with room for improvement in closely related categories. This milestone enhances the ability to analyze and understand case relationships, aiding in precedent research.

LawDigestAI

Catchphrase ExtractionCitation Classification

Catchphrase Extraction

Enter the case document text:

given on 4 September 2008 at Perth

whereby the Tribunal decided affirms the decision of the SSAT dated 23 November 2007

Run Catchphrase Extraction Model

Processing complete!

Extracted Catchphrases:

application for an adjournment of hearing of the administrative appeals tribunal

whether decision of authorised review officer of centrelink to cancel applicant's newstart allowance was correct.

social security

LawDigestAI

Catchphrase ExtractionCitation Classification

Citation Classification

Enter the citance text:

argument. I should note for completeness that at [11] Hely J relied upon what Adams J had said in Gargan v Scott (unreported, Supreme Court of New South Wales, 27 October 2003) and I respectfully adopt what Adams J said at [4]. I do not think that the argument based upon s 13 has any substance whatsoever.

Run Citation Classification Model

Processing complete!

Citation Class:

Followed

3. Real-Time Usability

A user-friendly interface was developed to integrate the summarization and classification tasks. This tool enables users to input legal texts and receive instant results for both catchphrase extraction and citation classification. The interface ensures the practical application of the models in real-world legal research scenarios, further supporting the goal of accessibility and efficiency.

In conclusion, this project successfully met its objectives, delivering a robust platform that simplifies legal document analysis. The tools developed are not only applicable in academic settings but also hold significant potential for practical deployment in the legal industry, paving the way for further advancements in NLP applications tailored to the legal domain.

References

1. Galgani, Filippo, Paul Compton, and Achim Hoffmann. "Combining Different Summarization Techniques for Legal Text." Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data (Hybrid2012), EACL 2012, 23 Apr. 2012, pp. 115–123. Association for Computational Linguistics, <https://aclanthology.org/W12-0515>.
2. Galgani, Filippo, and Achim Hoffmann. "LEXA: Towards Automatic Legal Citation Classification." Artificial Intelligence in Theory and Practice IV, edited by Max Bramer, Springer, 2010, pp. 445–454. Lecture Notes in Computer Science, vol. 6464, https://doi.org/10.1007/978-3-642-17432-2_45.