CSECU-DSG at SemEval-2023 Task 6: Segmenting Legal Documents into Rhetorical Roles via Fine-tuned Transformer Architecture

Fareen Tasneem, Tashin Hossain,
Jannatun Naim, and Abu Nowshed Chy
Department of Computer Science and Engineering
University of Chittagong, Chattogram-4331, Bangladesh
{fareen.tasneem,tashin.hossain.cu,
jannatun.naim.cu}@gmail.com, and nowshed@cu.ac.bd

Abstract

Automated processing of legal documents is essential to manage the enormous volume of legal corpus and to make it easily accessible to a broad spectrum of people. But due to the amorphous and variable nature of legal documents, it is very challenging to directly proceed with complicated processes such as summarization, analysis, and query. Segmenting the documents as per the rhetorical roles can aid and accelerate such procedures. This paper describes our participation in SemEval-2023 task 6: Sub-task A: Rhetorical Roles Prediction. We utilize a finetuned Legal-BERT to address this task. We also conduct an error analysis to illustrate the shortcomings of our deployed approach.

1 Introduction

Analyzing and extracting critical information from legal documents poses a great challenge for researchers. Legal documents can vary substantially in content, format, and structure depending on the jurisdiction, domain, and nature of documents. The corpus of legal documents is expanding rapidly, resulting in the need for automated systems for the effective retrieval of consequential information. Moreover, to ensure that these legal documents are utilized efficiently and justly, they must be accessible to a wide range of people. The task benefiting from the automation of legal documents are case law analysis, summarization, semantic search, and so on. Identifying the rhetorical role of each segment in the legal documents can aid in these tasks. Rhetorical role identification refers to partitioning the document into semantically meaningful segments with each segment having specific roles and purposes. These rhetorical roles can be used to extract significant information for further processing and analysis. But due to legal documents being unstructured, varying from domain to domain, and containing domain-specific literature, the identification of rhetorical roles is a challenging task. To address this challenge, SemEval-2023 task 6 (Modi et al., 2023) incorporated Sub-task A: Rhetorical Roles Prediction (RR). The task is to segment a given legal document by predicting the rhetorical role label for each sentence. An instance of a document segmented according to the rhetorical roles is illustrated in Figure 1.

Hachey and Grover (2004) employed a feature set consisting of the location of the sentence in the document, thematic words, sentence length, and cue phrases. They utilized decision tree, naive bayes, winnow algorithm, and support vector machine (SVM) as the classifiers. On the other hand, Saravanan et al. (2008) investigated rule-based learning algorithms (such as SLIPPER) and conditional random fields (CRF) along with features like indicator features, local features, layout features, state transition features, and legal vocabulary features. A multitask learning (MTL) framework incorporating Bi-LSTM and CRF was developed by Malik et al. (2021). They used document rhetorical role label shift as an auxiliary task for segmenting a legal document. They experimented with modeling the label shift using both SBERT-Shift (Reimers and Gurevych, 2019) and BERT-SC. Moreover, experiments were conducted on whether generalization can be implemented so that model trained on one sort of legal domain works well on different legal domains while classifying rhetorical roles. Savelka et al. (2021) employed RoBERTa (Liu et al., 2019) and SVM as their classification model for investigating this notion. Their findings portray that a model trained on a certain legal domain can work well on other legal domains, in fact sometimes better than training with the same legal domain. Moreover, using multiple domains together to train a model yields good performance.

In this paper, we present our deployed framework to address the challenge of predicting rhetor-

The first three authors have equal contributions.

Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), pages 2112–2117
July 13-14, 2023 ©2023 Association for Computational Linguistics

Segments	Rhetorical Roles
PETITIONER: THE COMMISSIONER OF INCOME-TAXNEW DELHI Vs. RESPONDENT: M/s. CHUNI LAL MOONGA RAM ……………………	PREAMBLE
JUDGMENT: CIVIL APPELLATE JURISDICTION .- Civil Appeals Nos. 39 and 40 of 1960. ……………………	ANALYSIS
Appeals from the judgment and order dated January 23, 1957, of the Punjab High ……………………	NONE
These two appeals have been brought to this Court on a certificates of fitness granted by the High, Court of Punjab under s. 66A(2) of the Indian Income-tax Act, 1922 ……………..	FAC
The High Court was then moved under s. 66 (2) of the Indian income-tax Act, 1922 and the High, Court heard the two applications together and directed the Tribunal to state ...………	RLC
Thereafter the Commissioner of Incometax, Delhi, asked for and obtained a certificate under s.66A(2) ……………………	FAC
As to the first question the learned Additional Solicitor- General, appearing on behalf of the appellant, ……………………	ARG_PETITIONER
In Civil Appeal No. 40 of 1960 the second question falls for decision. ……………………	ANALYSIS
We may here read s.5 and the third proviso thereto : .lm15 " s. 5. ……………………	STA
In Commissioner of Income-tax v. Karamchand Premchand Ltd.(1). ……………………	PRE_RELIED
If that part of the business has to be treated as a separate business for the purposes of the Excess Profits Tax Act, ……….	ANALYSIS
We are of the view that on the facts found, the answer to the second question must be in favour of the appellant ……………	RPC

Figure 1: Instance of segmenting a legal document as per rhetorical roles

ical roles as elucidated in SemEval-2023 task 6: Subtask A. We employ Legal-BERT (Zheng et al., 2021) and further finetune it to obtain a better performance.

We organize the remaining contents of the paper as follows: we illustrate our proposed framework in Section 2. Section 3 elucidates the experimental specification. We depict an analysis of erroneous predictions of rhetorical roles in Section 4. In Section 5, we conclude this paper with some strategies that we intend to adopt in future.

2 Proposed Framework

In this section, we elucidate our presented framework for the task of rhetorical role predictions. We exploit fine-tuned Legal-BERT (Zheng et al., 2021) for classifying legal sentences into 13 rhetorical roles. Given a legal judgment document, we segment it into separate sentences to cast this challenge as a multi-class sequence classification task. The overview of our proposed framework is depicted in Figure 2.

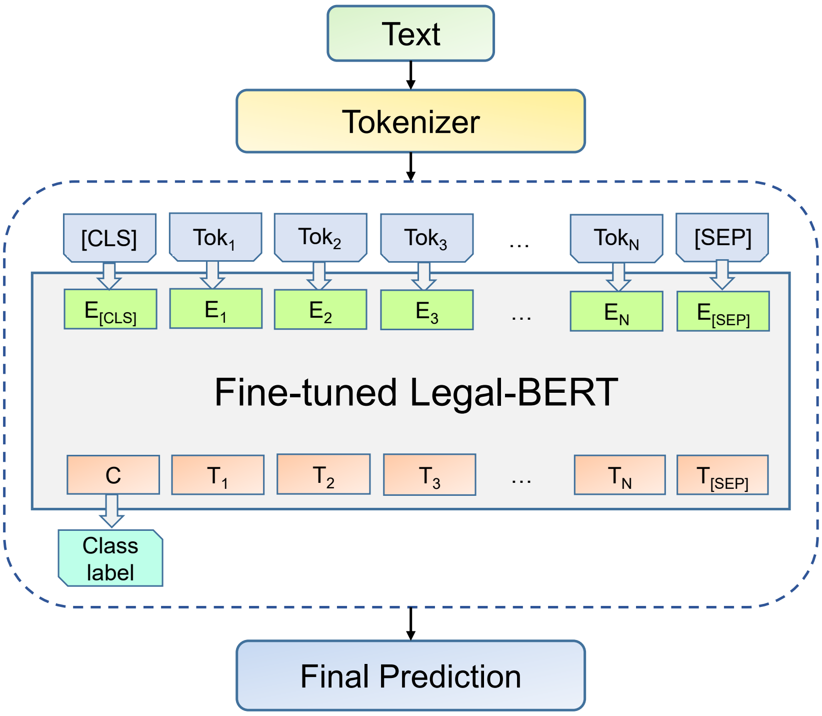

Figure 2: Proposed rhetorical roles classification framework.

2.1 Legal-BERT

To better capture the critical context of the legal sentences, we employ Legal-BERT (Zheng et al., 2021) for the rhetorical roles prediction task. Legal-BERT is a variant of BERT (Devlin et al., 2019) architecture that is one of the phenomena of harnessing transfer learning in language-related tasks such as sequence classification, question answering, and summarization. It is pre-trained on a corpus of Harvard Law legal documents from 1965 to 2021. It contains 3,446,187 case judgments of the entire federal and state courts in the US. The model was fine-tuned based on the pre-trained 'bert-base-uncased' model which contains 12 stacking encoder layers and utilizes the notion of understanding languages bidirectionally to capture the context of the text. Before fine-tuning, we segment the legal documents into separate sentences and then tokenize them using the BertTokenizer. Subsequently, we train the model through several iterations for the legal text classification task and get the final predicted label which is one of the 13 rhetorical roles.

3 Experiments and Evaluations

3.1 Dataset

The dataset used for this task is the Rhetorical Roles Corpus (Kalamkar et al., 2022) provided in the SemEval-2023 task 6 (Modi et al., 2023). The corpus comprises 354 English Indian legal judgment documents consisting of 40,305 sentences labeled with 13 rhetorical roles. The roles can be defined as preamble, facts, ruling by lower court, issues, argument by petitioner, argument by respondent, analysis, statute, precedent relied, precedent not relied, ratio of the decision, ruling by present court, and none. The judgment documents were collected from the Supreme Court, High Court, and some district-level courts of India. The documents are provided in JSON format where each sentence is annotated according to the rhetorical roles. Consecutive sentences with the same rhetorical role form a segment. The legal sentences were manually labeled by several selected law students. A brief summary of the dataset is presented in Table 1. We consider the micro F1-score as an evaluation measure for this task.

	Train	Dev	Test
Documents	247	30	50
Sentences	28986	2879	4158

Table 1: Summary of the dataset.

3.2 Experimental Setup

In Table 2, we illustrate the hyper-parameter setting of our proposed model. We utilized the Google Colaboratory platform for experimental purposes and implemented our model using Pytorch. We fine-tuned the pre-trained Legal-BERT model proposed by (Zheng et al., 2021) to handle the challenge of legal text classification. After assembling the training data into 32 batches, we train this model in 20 epochs with a learning rate of $1e-5$.

System	Settings
Fine-tuned Legal-BERT	1. *Max_seq_length*: 256 2. *Epochs*: 20 3. *Training batch size*: 32 4. *Optimizer*: AdamW 5. *Learning rate*: $1e-5$ 6. *Epsilon*: $1e-8$ 7. *Tokenizer*: 'bert-base-uncased' 8. *Model*: 'zlucia/legalbert'

Table 2: System settings for the proposed model.

3.3 Results and Analysis

In this section, we compare our system with other participants of the SemEval-2023 task 6. Table 3 depicts the accomplished results of other participants along with our submitted system. We can observe that our model attained a mediocre performance compared to the other top-performing teams. Though our model obtained a 64% F1-score on the test set of the judgment documents, it lacks by a huge margin compared to the top-performing team.

Team Name	F1-Score
AntContentTech (1st)	0.8593
DeepAI	0.8028
scholarly360	0.6902
CSECU-DSG	**0.6465**
ccidragons	0.5813
aminard	0.5681

Table 3: Comparative performance analysis on the test dataset.

4 Discussion

In this section, we further inspect the performance of our proposed system on the development set by presenting some of the predictions in Table 4.

Segments	Predicted Label	Gold Label
#1: Sub-clause (ix) which was inserted by the Finance Act, 1972 reads as follows: "(ix) any winnings from lotteries, crossword puzzles, races including horse races, card games and other games of any sort or from gambling or betting of any form or nature whatsoever;"	STA	STA
#2: The Income-tax authorities disallowed the claim on the ground that if the Bhatinda transactions had resulted in profits, such profits would have been exempt from tax in terms of s.14(2)(c) as it then stood and if the profits were exempt from tax	PREAMBLE	FAC
#3: As there was no incriminating evidence the statement of accused U/s.313 of Cr.P.C. is also dispensed with.	RPC	RATIO

Table 4: Performance analysis of the proposed model on the development dataset.

We can observe that in the first instance, our proposed method successfully predicted the text as a statute by capturing the context of describing an established law. However, in the second instance, the legal sentence is incorrectly predicted as a preamble (PREAMBLE) while it is actually a fact (FAC). This erroneous prediction can occur because of the mentioned names of the alleged parties which typi-

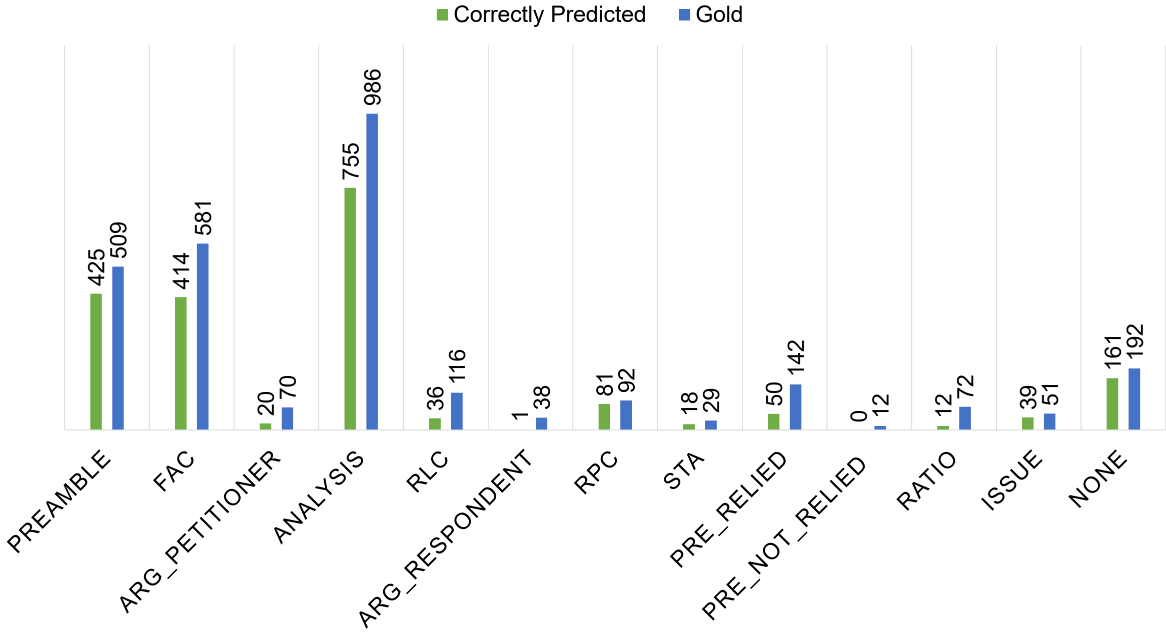

Figure 3: Statistics of correctly predicted labels on the development set.

cally occurs in the preamble. In the next example, the text is misclassified as a ruling by the present court (RPC) but the ground truth is a ratio (RATIO). It can happen because the sentence reflects a similar pattern as a final decision which is normally defined in an RPC.

Further, we investigate the performance of our system in Figure 3 by presenting a statistic of correctly predicted labels against the ground truth of the development set. We can observe that in most cases, our system correctly predicted nearly the same number of rhetorical roles as the gold labels of the development set. But in some cases, such as ARG_PETITIONER, RLC, ARG_RESPONDENT, STA, PRE_NOT_RELIED, and RATIO, our system lacks a considerable number of differences. The difference can occur since the dataset is quite imbalanced and very few training samples are provided for some of the rhetorical roles.

5 Conclusion

In this paper, we have described our proposed framework Legal-BERT. In the future, we intend to extend the corpus to include other legal domains with different structure. We also aspire to address the other two subtasks of SemEval-2023 task 6.

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Ben Hachey and Claire Grover. 2004. A rhetorical status classifier for legal text summarisation. In *Text Summarization Branches Out*, pages 35–42.

Prathamesh Kalamkar, Aman Tiwari, Astha Agarwal, Saurabh Karn, Smita Gupta, Vivek Raghavan, and Ashutosh Modi. 2022. Corpus for automatic structuring of legal documents. *arXiv preprint arXiv:2201.13125*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Vijit Malik, Rishabh Sanjay, Shouvik Kumar Guha, Shubham Kumar Nigam, Angshuman Hazarika, Arnab Bhattacharya, and Ashutosh Modi. 2021. Semantic segmentation of legal documents via rhetorical roles. *arXiv preprint arXiv:2112.01836*.

Ashutosh Modi, Prathamesh Kalamkar, Saurabh Karn, Aman Tiwari, Abhinav Joshi, Sai Kiran Tanikella, Shouvik Guha, Sachin Malhan, and Vivek Raghavan. 2023. SemEval-2023 Task 6: LegalEval: Understanding Legal Texts. In *Proceedings of the*

17th International Workshop on Semantic Evaluation (SemEval-2023), Toronto, Canada. Association for Computational Linguistics (ACL).

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992.

M Saravanan, Balaraman Ravindran, and S Raman. 2008. Automatic identification of rhetorical roles using conditional random fields for legal document summarization. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*.

Jaromir Savelka, Hannes Westermann, and Karim Benyekhlef. 2021. Cross-domain generalization and knowledge transfer in transformers trained on legal data. *arXiv preprint arXiv:2112.07870*.

Lucia Zheng, Neel Guha, Brandon R Anderson, Peter Henderson, and Daniel E Ho. 2021. When does pretraining help? assessing self-supervised learning for law and the casehold dataset of 53,000+ legal holdings. In *Proceedings of the eighteenth international conference on artificial intelligence and law*, pages 159–168.