DAUIN
Department of Control and Computer Engineering

Master's Degree Thesis
in *Data Science and Engineering*

Court Judgment Prediction and Explanation based on Transformers

Supervisor: Prof. Luca Cagliero

Co-Supervisor: Dr. Irene Benedetto

Politecnico di Torino

Candidate: Salvatore Junior Curello

December 2023

Politecnico
di Torino

Introduction

Thesis Objectives

State-of-the-art

Methodologies

Results

Conclusions

Introduction

The vast majority of data in the legal domain are represented in text forms, such as:

- Judgment documents
- Legal agreements
- Contracts
- Legal opinions

Challenges

Legal Language Complexity

Navigating the legal domain is intricate due to technical terms, impeding proper understanding.

Document Length

Legal document are are typically quite long, verbose, and noisy.

Interpretability

It is required to understand the facts, follow the arguments, and apply legal rules to arrive at the final decision.

Limitations of common Language Models:

- they are typically trained on general language patterns
- they struggle with large volumes of data

Thesis Objectives

To implement an automated system that addresses two tasks.

Binary Classification Problem

Court Judgment Prediction (CJP)

Thesis Objectives

To implement an automated system that addresses two tasks.

Binary Classification Problem

Court Judgment Prediction (CJP)

Explainability Problem

Court Judgment Prediction and Explanation (CJPE)

State-of-the-art

(Zhong et al., 2018): incorporates multiple legal subtasks and DAG dependencies into judgment prediction [1].

- *Limitation*: Lower performances with unbalanced datasets

(Ye et al., 2018): court view generation from the fact description in a criminal case [2].

- *Limitation*: Explanations are incorporated into the training data

(Xin et al., 2018): provide explanations of charge predictions [3].

- *Limitation*: Explanations are incorporated into the training data

(Malik et al., 2021): generate predictions and produce explanations using the occlusions method [4].

- *Limitation*: Only Generic Transformer Models are analyzed

Dataset overview

Indian **L**egal **D**ocument **C**orpus (**ILDC**)

- **ILDCsingle**: single petition followed by a single decision or multiple petitions with an identical decision
- **ILDCmulti**: multiple petitions results in different decisions

Dataset overview

Indian **L**egal **D**ocument **C**orpus (**ILDC**)

- **ILDCsingle**: single petition followed by a single decision or multiple petitions with an identical decision
- **ILDCmulti**: multiple petitions results in different decisions

	Number of documents		
Corpus	**Train**	**Validation**	**Test**
Single	5082	994	1517
Multi	32305		
Expert	56		

Methodologies

CJP

- Non-Hierarchical Transformer Models
- Hierarchical Transformer Models

Methodologies

CJP

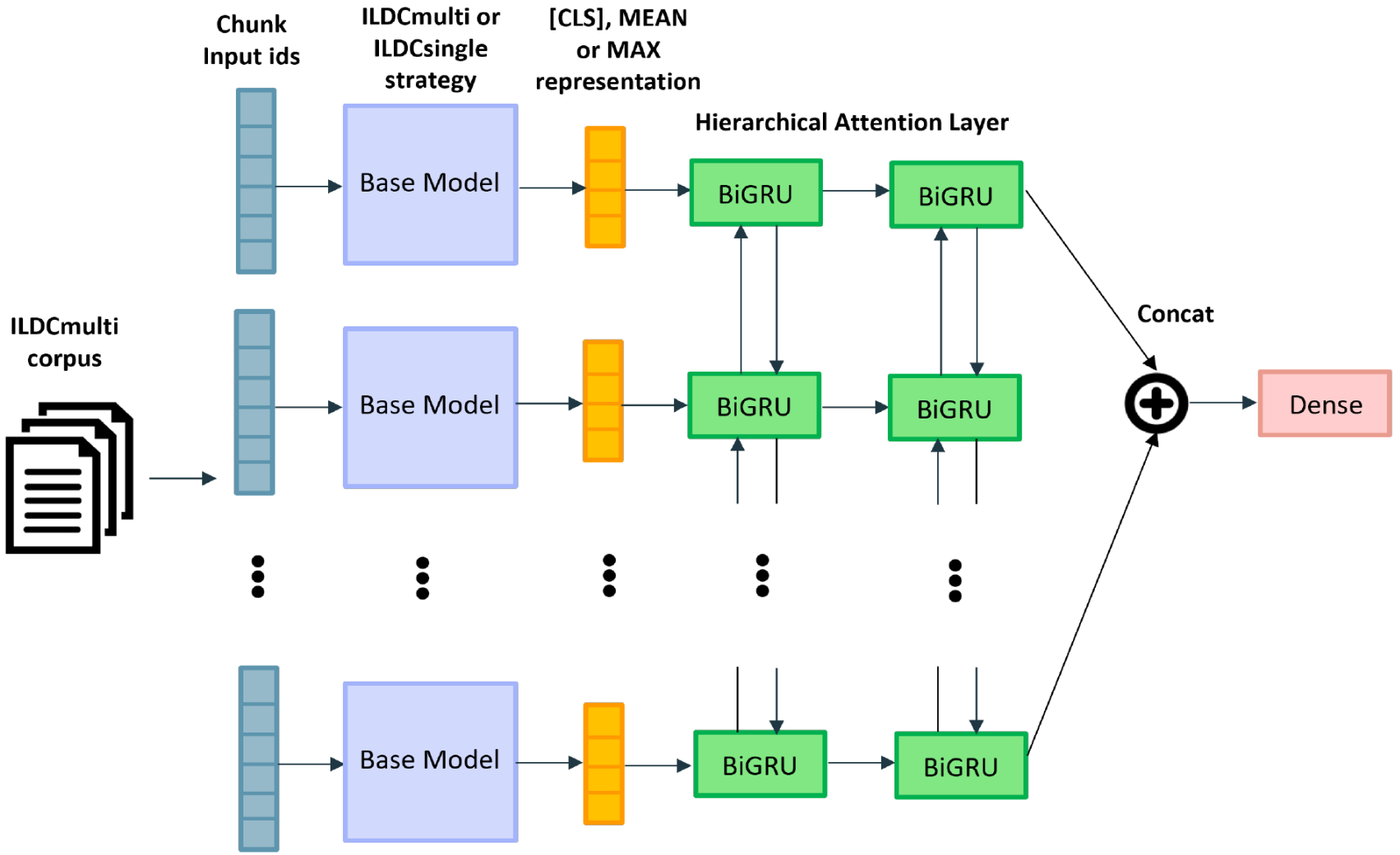
- Non-Hierarchical Transformer Models

Trained on the last N tokens of each document

where N is the max number of tokens supported by each model.

Methodologies

Hierarchical Transformer Models



Methodologies

CJPE

1. Split each doc in **ILDCexpert** in chunks of 512 tokens with an overlap of 100.

Methodologies

CJPE

1. Split each doc in **ILDCexpert** in chunks of 512 tokens with an overlap of 100.

2. Explanation generation:
   - **Occlusions method**: extract the best sentences in each chunk
   - **Attention Mechanism**: extract chunk with higher attention weight

Politecnico di Torino

Court Judgment Prediction

Results

CJP Results

- LegalLSGBERT is the top performing model

		Non-Hierarchical Transformers
Model	**Tokens**	**F1-score**
Baseline	512	71.77
CaseLawBERT	512	75.51
LegalBERT	512	75.32
RoBERTa	512	69.62
LegalLED	1024	68.97
LED	1024	67.90
LegalLSGBERT	2560	**82.01**

CJP Results

- LegalLSGBERT is the top performing model
- Domain-specific Models outperform Generic Models

		Non-Hierarchical Transformers
Model	**Tokens**	**F1-score**
Baseline	512	71.77
CaseLawBERT	512	75.51
LegalBERT	512	75.32
RoBERTa	512	69.62
LegalLED	1024	68.97
LED	1024	67.90
LegalLSGBERT	2560	**82.01**

Domain-specific Models Generic Models

CJP Results

- LegalLSGBERT is the top performing model
- Domain-specific Models outperform Generic Models
- Hierarchical models further improve the performances

		Non-Hierarchical Transformers	Hierarchical Transformers
Model	**Tokens**	**F1-score**	**F1-score**
Baseline	512	71.77	77.79
CaseLawBERT	512	75.51	**81.12**
LegalBERT	512	75.32	**78.90**
RoBERTa	512	69.62	**72.55**
LegalLED	1024	68.97	**69.45**
LED	1024	67.90	**69.05**
LegalLSGBERT	2560	82.01	**82.13**

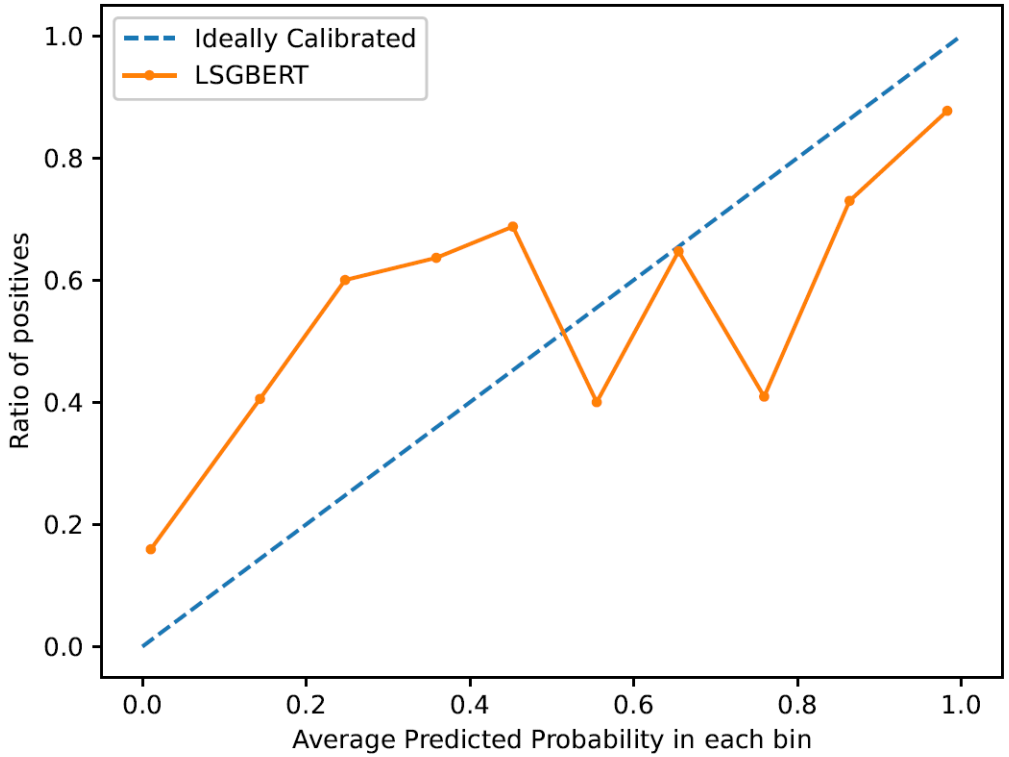
CJP Results

- LegalLSGBERT is the top performing model
- Domain-specific Models outperform Generic Models
- Hierarchical models further improve the performances

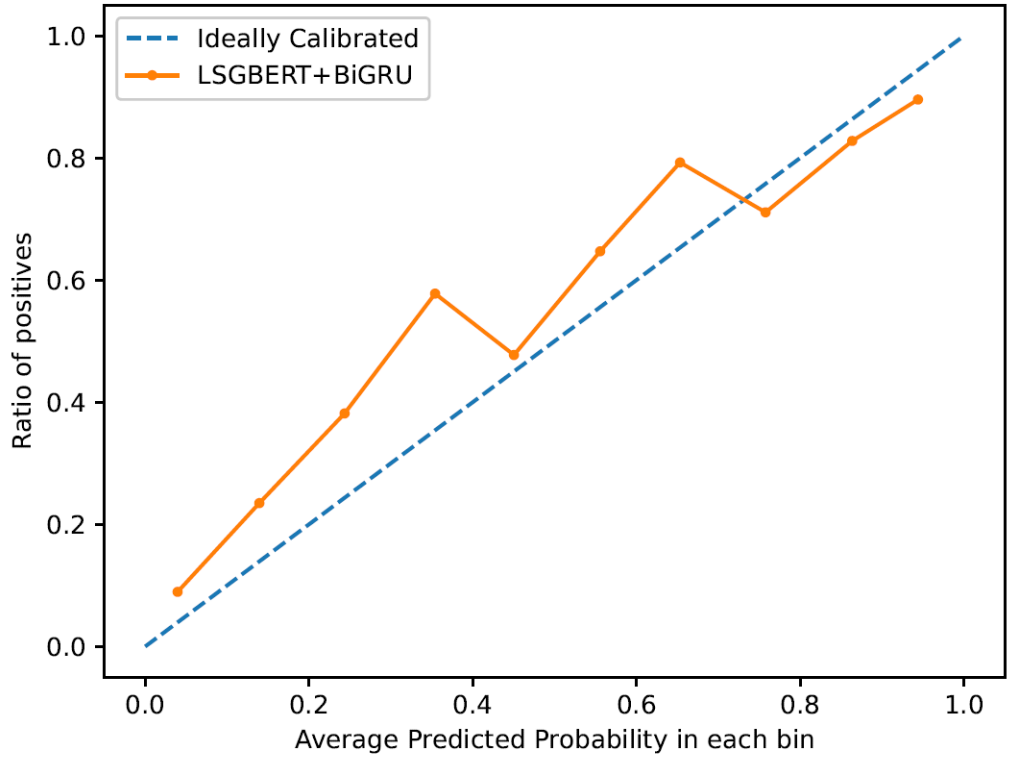
		Non-Hierarchical Transformers	Hierarchical Transformers
Model	**Tokens**	**F1-score**	**F1-score**
Baseline	512	71.77	77.79
CaseLawBERT	512	75.51	**81.12**
LegalBERT	512	75.32	**78.90**
RoBERTa	512	69.62	**72.55**
LegalLED	1024	68.97	**69.45**
LED	1024	67.90	**69.05**
LegalLSGBERT	2560	82.01	**82.13**

Calibration

Non-Hierarchical LegalLSGBERT



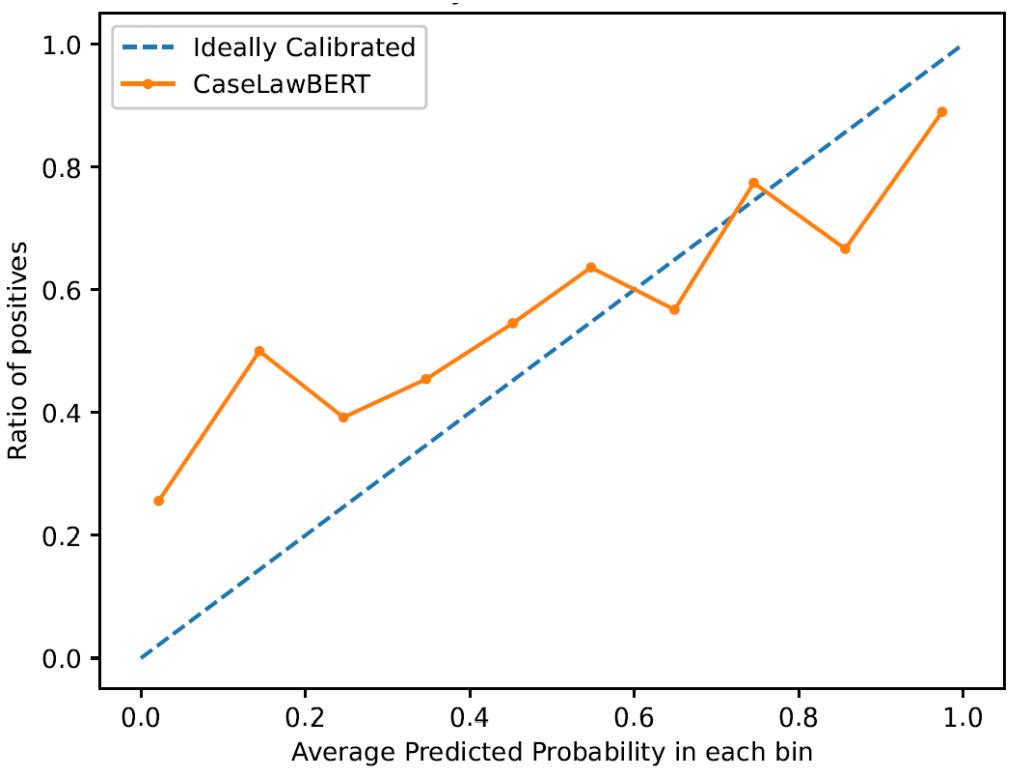
Hierarchical LegalLSGBERT



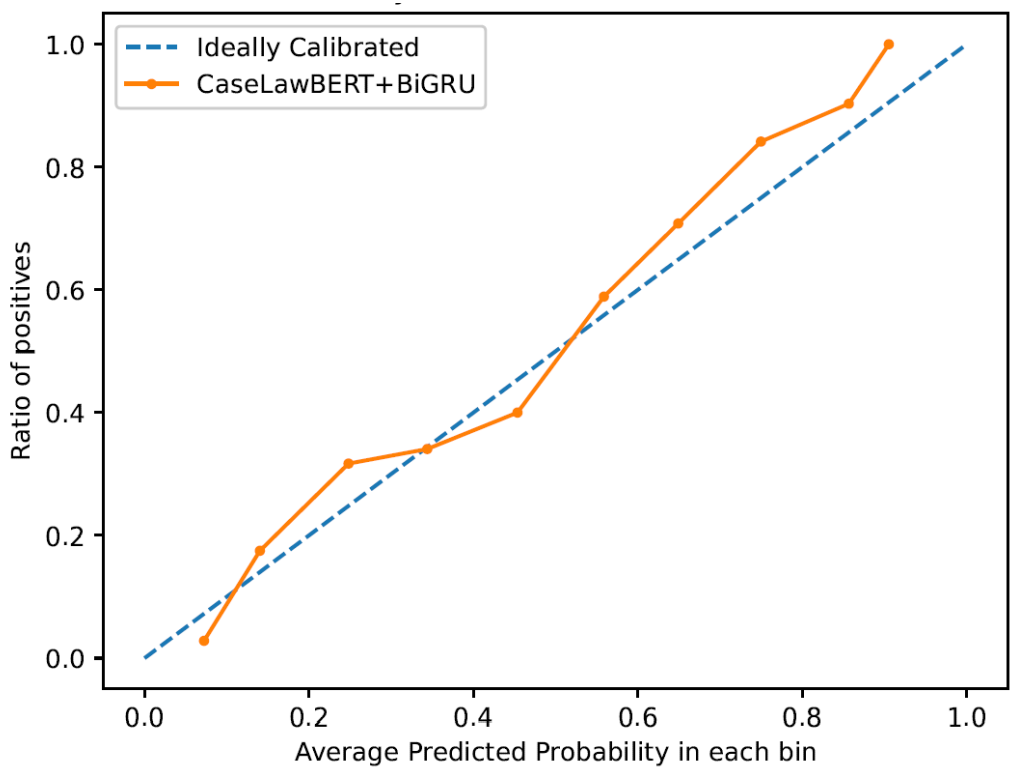
Hierarchical models, besides being more performant, are also more calibrated

Calibration

Non-Hierarchical CaseLawBERT



Hierarchical CaseLawBERT



Hierarchical models, besides being more performant, are also more calibrated

Politecnico di Torino

Court Judgment Prediction and Explanation

Results

CJPE Results

Occlusions Method (40% of the best sentences)

Model	Jaccard Similarity	Overlap Min	Overlap Max	ROUGE-1	ROUGE-2	ROUGE-L
Baseline	0.324	**0.719**	0.383	0.451	0.297	0.424
LegalLSGBERT	**0.325**	0.673	0.398	0.486	0.326	0.466
CaseLawBERT	**0.325**	0.663	**0.404**	**0.490**	**0.336**	**0.474**

- Domain-specific Models outperform the baseline
- Despite using the same methodology, the models are more explainable.

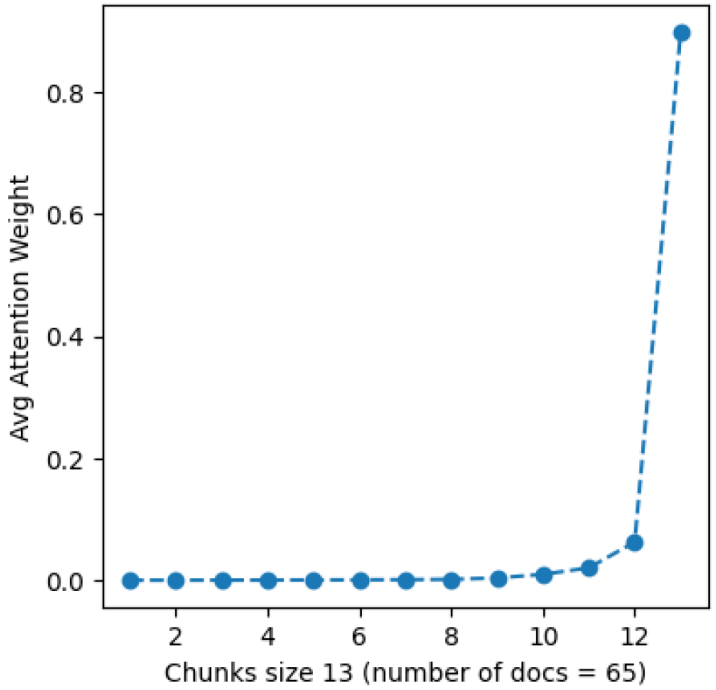
CJPE Results

Attention Mechanism

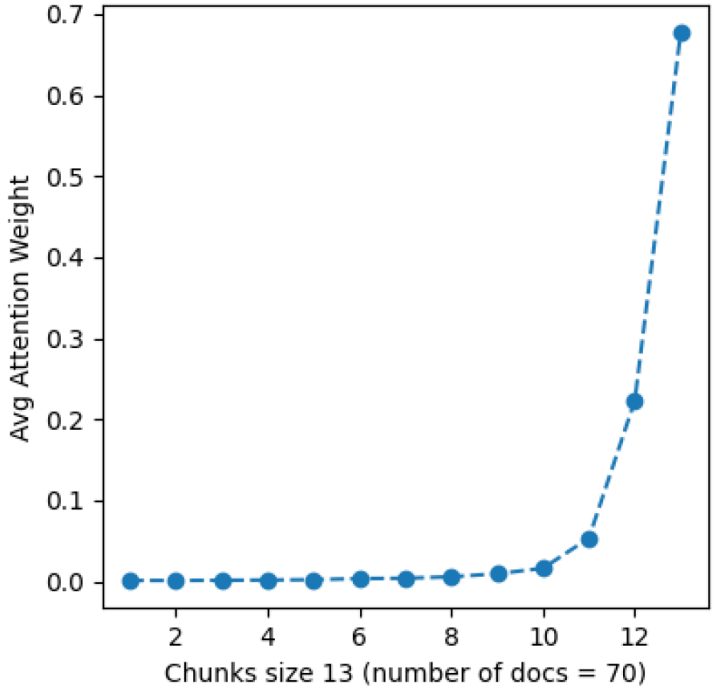
Model	Jaccard Similarity	Overlap Min	Overlap Max	ROUGE-1	ROUGE-2	ROUGE-L
Baseline	**0.324**	0.719	**0.383**	**0.451**	**0.297**	**0.424**
LegalLSGBERT	0.187	0.765	0.206	0.281	0.173	0.269
CaseLawBERT	0.226	**0.771**	0.250	0.334	0.217	0.322

- Drop in performance caused by explanations that are 75% shorter

CJPE Results



(a) LegalLSGBERT



(b) CaseLawBERT

- Typically, last chunks are the most representative ones.
- Last chunk contains many fewer tokens compared to the previous ones.

CJPE Results

Attention Mechanism with Revised chunk division

Model	Jaccard Similarity	Overlap Min	Overlap Max	ROUGE-1	ROUGE-2	ROUGE-L
Baseline	**0.324**	0.719	**0.383**	0.451	0.297	0.424
LegalLSGBERT	0.187	0.765	0.206	0.281	0.173	0.269
CaseLawBERT	0.226	**0.771**	0.250	0.334	0.217	0.322
LegalLSGBERT (end)	0.320	0.747	0.366	0.451	0.318	0.436
CaseLawBERT (end)	0.323	0.743	0.370	**0.458**	**0.323**	**0.444**

- Performances are similar to the baseline but the explanations are 40% shorter.

Conclusions

Findings:

- Hierarchical Transformer models not only offer better performance but also greater stability.
- Domain-specific Models outperform Generic Models in both CJP and CJPE.

Future research line:

- Pretraining on Indian legal documents
- Multi-modal Learning

Thank you for your attention

Bibliography

[1] Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. «Legal Judgment Prediction via Topological Learning». In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Ed. by Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii. Brussels, Belgium: Association for Computational Linguistics, Oct. 2018, pp. 3540–3549. doi: 10.18653/v1/D18-1390. url: https://aclanthology.org/D18-1390 (cit. on p. 31).

[2] Hai Ye, Xin Jiang, Zhunchen Luo, and Wenhan Chao. «Interpretable Charge Predictions for Criminal Cases: Learning to Generate Court Views from Fact Descriptions». In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). Ed. by Marilyn Walker, Heng Ji, and Amanda Stent. New Orleans, Louisiana: Association for Computational Linguistics, June 2018, pp. 1854–1864. doi: 10.18653/v1/N18-1168. url: https://aclanthology.org/N18-1168 (cit. on p. 33).

[3] Xin Jiang, Hai Ye, Zhunchen Luo, WenHan Chao, and Wenjia Ma. «Interpretable Rationale Augmented Charge Prediction System». In: Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations. Ed. by Dongyan Zhao. Santa Fe, New Mexico: Association for Computational Linguistics, Aug. 2018, pp. 146–151. url: https://aclanthology.org/C18-2032 (cit. on p. 33).

[4] Vijit Malik, Rishabh Sanjay, Shubham Kumar Nigam, Kripa Ghosh, Shouvik Kumar Guha, Arnab Bhattacharya, and Ashutosh Modi. ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation. 2021. doi: 10.48550/ARXIV.2105.13562. url: https://arxiv.org/abs/2105.13562 (cit. on pp. 2, 24, 31–34, 36–38, 40, 41, 47–49, 51, 54, 58–60, 69).