Project#7: Generating Explanations for Legal Judgment Prediction

Prateesh Awasthi[1], **Ashwani Kumar Srivastava**[2], **Sidhartha Watsa**[3],**Ayush Ranjan**[4],**Ayush Dhingra**[5]
[1]19807636, [2]190196, [3]190843, [4]19807216, [5]200240
[1]ME, [2]CHE, [3]ME, [4]ME, [5]ME
{prateesh, ashwanis, sidwat, ayranjan, ayushd20}@iitk.ac.in

Abstract

In this report, we have proposed and experimented with transformer-based neural network models to predict the judgment given the full legal document related to the case. This problem statement was inspired by the ILDC paper (Malik et al., 2021a). We have solved the basic problem in any legal judgment prediction task i.e., to capture long-range dependencies due to huge document size, along with attention mechanism, and have achieved better results using graph neural networks along with transformer models.

1 Introduction

With more than 59 lakhs pending cases in the country and adding up everyday , the legal judgement prediction task has the potential to have a substantial impact on the legal profession and society as a whole by enhancing the effectiveness, consistency, and fairness of legal decision-making by generating judgement much faster.
Given a complete legal document containing facts, evidence, arguments from both parties, a ruling by the lower court, issues, and precedents, the research problem is predicting the case's final verdict.
Creating an algorithm or model that reliably predicts the outcome of legal cases based on the facts of the case and relevant legal precedents is a crucial aspect of the legal judgement prediction problem.
A sub-part of problem is the creation of a model or algorithm capable of effectively balancing the numerous aspects and considerations that might affect the result of a legal case (such as employing an attention mechanism).

2 Problem Definition

The legal judgement prediction research problem is to develop an algorithm or model that reliably predicts the outcome of legal cases, given a complete legal document containing facts, evidence, arguments from both parties, a ruling from the lower court, issues, and precedents. (0 for verdict against the appellant and 1 for vice versa).

Legal data is often vast, unstructured, and constantly evolving, making it challenging to extract relevant information and identify patterns. Also legal decisions can be influenced by subjective factors, such as judicial discretion, legal interpretation, and societal context, which are difficult to quantify and model accurately.

Solving this issue necessitates resolving the difficulties posed by the complexity of legal data, unstructured data, and subjective factors, as well as developing robust and interpretable prediction models that can provide reliable insights into the likely outcomes of legal cases.

3 Related Work

1. *Indian Legal Documents Corpus for Court Judgment Prediction and Explanation: (Malik et al., 2021b)*
   This paper introduces ILDC (Indian Legal Documents Corpus), the first Law Corpus in an Indian setting containing $35K$ Indian Supreme Court cases annotated with original court decisions. The paper proposes the task of legal judgment prediction with explanations in the Indian setting based on ILDC dataset. It also depicts the results of experimentations such as $Classical\ Methods$(Logistic Regression, SVM and Random Forest with Doc2vec embeddings), $Sequence\ Models$(BiGRU with attention model), $Transformers\ Model$ (BERT, DistilBERT, RoBERTa, and XL-Net), $Hierarchical\ Transformers$ (XLNet with BiGRU) to perform judgment prediction. The best prediction model has an accuracy of 78% versus 94% for human legal experts, where the analysis of explanations by the proposed algorithm (hierarchical occlusion-based model for explainability) reveals a significant difference

in the point of view of the algorithm and legal experts for explaining the judgments, pointing towards scope for future research.

2. *Legal Judgment Prediction via Topological Learning: (Zhong et al., 2018)*
As legal judgment usually consists of multiple subtasks, such as the decisions of applicable law articles, charges, fines, and the term of penalty which depicts topological dependency amongst various subtasks. The paper formalizes the dependencies among the subtasks as a Directed Acyclic Graph (DAG) and propose a topological multi-task learning framework. It models the multiple subtasks in judgment prediction jointly under a multi-task learning framework using CNN based Neural Encoder for Fact Descriptions and a specific LSTM cell is employed for each task and the output of each task in the topological order where these subtask outputs are connected with a DAG to generate the final prediction . The experimental results show that this model achieves consistent and significant improvements over baselines on all judgment prediction tasks depicting need of exploiting these dependencies.

3. *When Does Pretraining Help? Assessing Self-Supervised Learning for Law and the CaseHOLD Dataset of 53,000+ Legal Holdings: (?)*
This paper provides insight about when one should engage in resource-intensive domain-specific pretraining (domain pretraining). More specifically, it was not clearly evident why there are only a few instances of substantial gains from domain-specific pertaining, although the legal language is widely seen to be unique. It hypothesizes that existing law-based tasks are easier, and hence domain pretraining does not help much. To depict this, it presented $CaseHOLD$ (Case Holdings On Legal Decisions), a new dataset comprised of over $53K+$ multiple choice questions to identify the relevant holding of a cited case. The paper also depicts the relevant performance gains on the $CaseHOLD$ dataset, and domain pretraining may be helpful if the task exhibits sufficient similarity to the pretraining corpus. The paper suggests Domain Specificity Score be considered before enabling domain-specific pretraining on law corpus.

4. *An Empirical Study on Cross-X Transfer for Legal Judgment Prediction(?)*
Cross-lingual transfer learning is understudied in the context of legal NLP, and not at all in Legal Judgment Prediction (LJP). This paper uses trilingual Swiss-Judgment-Prediction dataset to explore the transfer learning techniques in legal domain. It depicts cross lingual transfer improves the overall results across languages, especially when we use adapter-based fine-tuning. Further, the performance was also improved by augmenting the dataset with the machine-translated versions of the same document. . The paper also depicts that in legal areas and origin regions, models trained across all groups perform overall better, while they also have improved results in the worst-case scenarios.

5. *Causal Inference in Natural Language Processing: Estimation, Prediction, Interpretation, and Beyond:(Feder et al., 2022)* An important part of any legal decision is causality but the operating principle of the deep learning models is correlation and not causality. And during the learning process, the models learn many spurious correlations which have no direct impact on the result. To check for these spurious correlations, the authors suggested two tests namely Invariance Test and Sensitivity Tests which will check for the bias present in the data, and the model, in turn, increase our confidence in using these models in the high-stakes situation, which is the Legal judgment prediction.

6. *Causal Inference in Natural Language Processing: Estimation, Prediction, Interpretation, and Beyond:(?)* When given a graph as input, a GCN performs a convolution operation on each node. The GCN can extract features from each node's immediate neighbourhood by using the convolution operation, which takes into account both the node's own features and the features of its nearby nodes. The way a conventional convolutional neural network pulls features from an image's local neighbourhoods is comparable to this.

   A new collection of node features is produced by each convolutional layer, and these node characteristics can be fed into additional convolutional layers to extract ever-more complex

features. A set of node embeddings, which form a low-dimensional representation of the original graph, is the GCN's ultimate output. GCN's outperform recurrent networks in terms of capturing long range dependencies and thus can be useful for Legal Judgment Prediction.

4 Corpus/Data Description

Corpus (Avg. tokens)	**Number of docs (Accepted Class %)**		
	Train	**Validation**	**Test**
ILDC$_{multi}$ (3231)	**32305 (41.43%)**	**994 (50%)**	**1517 (50.23%)**
ILDC$_{single}$ (3884)	**5082 (38.08%)**		
ILDC$_{expert}$ (2894)	**56 (51.78%)**		

Figure 1: ILDC : Type of Datasets

ILDC 1 consists of $35k$ Indian Supreme Court case documents, with each document labels as "0" for verdict against the petition and "1" for vice versa. Out of these, 56 documents are annotated by 5 legal experts explaining the decision being made.

The other datasets with relatively large size are Chinese AI and Law Challenge Dataset (CAIL 2019) ($2.6million$ cases) and CaseHOLD Dataset ($53k$ multiple choice questions with prompts from a judicial decision). The detailed description of these datasets can be found in the cited papers.

5 Proposed Approach

1. **Model Input :** Since the input document will be very large, passing embeddings for individual words and creating attention vector of each of them would unduly increase the number of trainable parameters and make the model complex. Instead, we divide the document into chunks of fixed token size and create a representation vector of each of these using a suitable transformer model.
   The attention mechanism is used which enables the model to give different weights or importance to various words or phrases in the input text based on their relevance to the prediction task. This allows the model to selectively focus on relevant details while ignoring irrelevant or redundant data.

2. **Transformer Layer :** A transformer is a state-of-the-art method for creating text representation vectors. transformer is used to capture long-range dependencies between different parts of the input text, such as the facts, arguments, and legal precedents cited in a case. The transformer's self-attention mechanism can effectively capture these dependencies by allowing the model to attend to different parts of the input text and their relationships with each other

3. **Dense graph and GCN :** A fully connected dense graph is made out of embeddings from the transformer and a graph convolutional network is used to predict the judgement and gives the output '0' for verdict against the appellant and '1' for vice versa.

Evaluation : For the prediction task, a subset of ILDC dataset would act as evaluation dataset on which predictions will be tested. For explanation generation task, the extracted text will be compared with 56 documents in ILDC annotated with gold standard dataset using some suitable metric.

6 Experiments and Results

6.1 Distilbert+GCN

Model	Accuracy	macro F1
DistilBert +GCN+ Attention+Dense Layer	0.704	0.702
DistilBert+ GCN (2-Layers)	0.707	0.710
DistilBert + DenseGCN	0.69	0.695
DistilBert + BiGRU + GCN	0.687	0.688

6.2 Explanation Generation

We have extracted the attention scores from the encoded representations of the input data and used them to generate the explanations.

7 Error Analysis

The analysis of the inferences had shown that capturing the long-range dependencies to do the prediction is a tough task. The models have failed to connect the facts, evidence and arguments. These errors are better handled by the graph-based neural network.

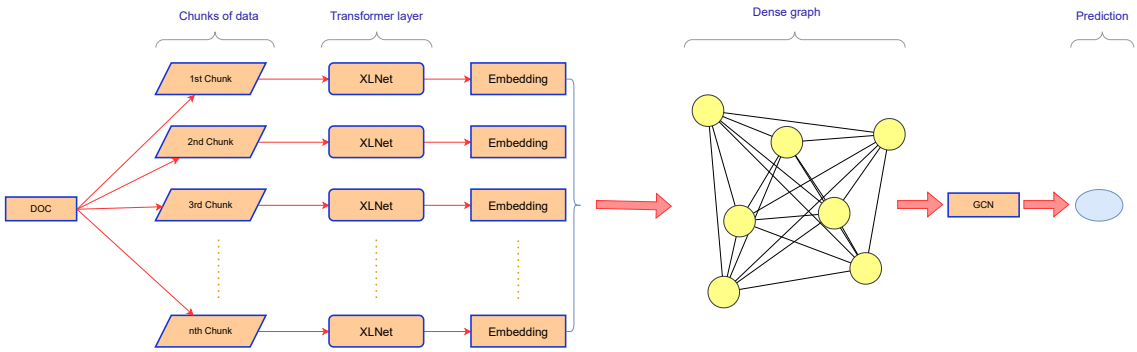

Figure 2: Basic Pipeline

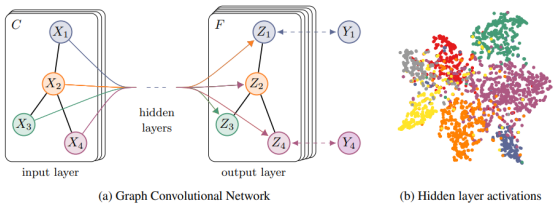

Figure 3: GCN

8 Future Directions

Identification of Rhetorical Roles of Sentences in Indian Legal Judgments, Paheli Bhattacharya et al.. Rhetorical Roles for each statement in the document can be evaluated using Heirarchial Bi-LSTM Attaching the corresponding rhetorical roles with each sentence in the document. Grouping the sentences according to rhetorical roles assigned to it. Then following the pipeline proposed above in the paper to train the paper.

9 Individual Contribution

Name	Work
Prateesh	**ILDC, Cross-X transfer Learning Transformer Interpretability Beyond Attention ,GCN**
Ashwani	**ILDC, LJP Topological learning, Domain-specific Pretraining**
Sidhartha	**ILDC, Causal Inference in NLP, Cross-X Transfer for LJP**
Ayush R	**ILDC, Topological learning, Causal Inference in NLP**
Ayush D	**ILDC, Presentation, Report,GCN**

10 Conclusion

We have achieved better results for legal judgment prediction using BERT+GCN model compared to Bert+BiGRU of ILDC paper

References

Amir Feder, Katherine A. Keith, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar, Zach Wood-Doughty, Jacob Eisenstein, Justin Grimmer, Roi Reichart, Margaret E. Roberts, Brandon M. Stewart, Victor Veitch, and Diyi Yang. 2022. Causal inference in natural language processing: Estimation, prediction, interpretation and beyond.

Vijit Malik, Rishabh Sanjay, Shubham Kumar Nigam, Kripa Ghosh, Shouvik Kumar Guha, Arnab Bhattacharya, and Ashutosh Modi. 2021a. Ildc for cjpe: Indian legal documents corpus for court judgment prediction and explanation.

Vijit Malik, Rishabh Sanjay, Shubham Kumar Nigam, Kripabandhu Ghosh, Shouvik Kumar Guha, Arnab Bhattacharya, and Ashutosh Modi. 2021b. ILDC for CJPE: Indian legal documents corpus for court judgment prediction and explanation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4046–4062, Online. Association for Computational Linguistics.

Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. 2018. Legal judgment prediction via topological learning. In *Proceed-*

ings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3540–3549, Brussels, Belgium. Association for Computational Linguistics.