2023 7th International Conference on Computer Applications in Electrical Engineering-Recent Advances (CERA) | 979-8-3503-0500-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/CERA59325.2023.10455227

Automatically Classify each sentence from Indian Legal Judgements to its Semantic Role

Niranjan Adsul
Department of Computer Engineering
College of Engineering Pune Technological University
Pune, India
niranjana21.comp@coep.ac.in

Dr. Pratiksha Deshmukh
Department of Computer Engineering
College of Engineering Pune Technological University
Pune, India
dpr.comp@coep.ac.in

Abstract— **In the past decade, automation has gained traction in Indian law. To perform many downstream tasks like judgment classification, judgment prediction, Named Entity Recognition (NER), etc., digitally understanding Indian legal case documents/judgments is the most important task. In legal judgement documents, sentences perform different semantic functions. Many models have sought in recent years to use contextual awareness between sentence sequences using Conditional Random Field (CRF) over manually handcrafted features, Hierarchical BiLSTM CRFs, etc., to assimilate the interrelationship between consecutive sentences. We propose a model for Semantic Role Labelling (SRL) of each sentence in Indian legal judgements. Our model out-performs previous models by using Named Entity replacement followed by Graph Attention (GAT) mechanism to generate context-aware sentence embeddings for each judgement sentence. An efficient SRL model makes downstream task of summarizing each semantic segment of judgement, more efficient and speed-up judgement analysis/study by law practitioners to deliver justice to all. A multi-class classifier neural network receives these context-aware embeddings and classifies them into one of thirteen semantic categories. The task is challenging since legal judgement documents are not properly structured. Also, the semantics roles may be subjective.**

Keywords—Rhetorical Role Labeling (RRL), Sequential Sentence Classification (SSC), Semantic Role Labeling (SRL), Legal Judgment Segmentation, Graph Attention (GAT), Name Entity Recognition (NER), Multi-Class Classification

I. Introduction

The Indian Judicial System (IJS) is well on its way to being fully digitalized, with judgement data being readily accessible online on each court's website. As a result, there is a growing demand for efficient automated information retrieval systems to help organize, process, and retrieve this information while also presenting it in a way that is suitable and user-friendly.

According to Supreme Court of India's National Judicial Data Grid (https://njdg.ecourts.gov.in/scnjdg/), the number of registered cases that are pending as of 09/16/2023 just in the Top Court are 64,975. This leads to delays in the administration of justice and negative impact on justice quality.

Semantic classification of sentences is one of the tasks for information retrieval from legal case documents. Effective semantic classification of sentences in court case documents is facilitated by natural language processing tools. Using so, legal experts can more efficiently obtain information while spending less time studying these previous judgements. Sentences in legal case judgements are not well organized. They have no fixed pattern, and their structure varies with domain. If each sentence in the court case ruling is analyzed separately, contextual information will be lost, and the model will not be strong enough to identify sentences correctly. The context and meaning of a statement in a judicial document is often determined by context from nearby sentences, leading to the term Sequential Sentence Classification (SSC) to be used to describe this difficulty.

We propose a model to perform SRL that is based on an input representation that may capture the neighborhood context to solve the neighborhood problem and is inspired by the Graph Attention (GAT) [1] mechanism used in Graph Neural Networks (GNN). The initial sentence embeddings are obtained from pretrained Bidirectional Encoder Representations from Transformers (BERT) model law-ai/INLegalBERT [2]. These non-contextual initial embeddings of sequence of sentences from legal judgement are transformed into context-aware embeddings using GAT [1] mechanism. In the beginning, we used a dataset of manually annotated judgements [3] that contained sentences for thirteen semantic role labels.

In earlier attempts [4], handmade characteristics, such as Language-based cues that suggest semantic functions, the sequential ordering of labels, etc., were used to automate the semantic role identification of sentences in Legal Judgements. Furthermore, these capabilities are limited to a few domains, need legal expertise, which is expensive to get, and might not be relevant across domains. The Deep Neural Network learning approach does away with the necessity for manually created features and instead automatically learns them given enough data.

Our method departs significantly from earlier work in the following ways:

A. *Initial sentence embeddings generation:*

We use a pretrained BERT [2] model, which was trained over enormous corpus of Indian legal judgements. Unlike previous work, we do not consider separate embeddings generated for each word of sentence but use the output embedding of BERT special token [CLS] that was appended before each sentence. As our task is to classify each sentence into its role label, embedding for [CLS] token carries more structural and semantic information for this task [5].

B. Use of GAT:

To obtain context-aware sentence embeddings that carry contextual information. This is because LSTM and derivatives can learn a lot of longer-term information from sequence of word tokens, but they can remember sequences of 100s, not 1000s or 10,000s or more. Whereas using multiple GAT [1] layers can easily help to convolve the neighborhood information in the embeddings of current sentence.

This model will make the downstream activities like Extractive/Abstractive Summarization of Court Judgments using Rhetorical Roles, Court Judgment Prediction using Rhetorical Roles, etc. This will ultimately reduce the case pendency by speeding up the process of justice administration.

II. Prior Work

Recent years have seen a lot of activity in the realm of lawful digital automation. There were many different datasets and tasks presented. Argument mining [6], Information extraction and retrieval [7], Event Extraction [8], prior case retrieval [9], summarization [10], and case prediction are a few examples.

Some recent works have focused on the development of annotated corpora and the challenge of automatically labelling rhetorical roles. Reference [11] created a corpus called TEMIS with 504 sentences that are syntactically and semantically analyzed. Corpus for Automatic Structuring [3].

Automatic labeling of rhetorical roles was first conducted on Indian legal judgements in [4], where CRFs were used to label seven rhetorical roles. Reference [12] created a method for determining if sentences are facts or not using fastText. Reference [13] splits United States legal judgements into semantic and topic-specific segments using Conditional Random Fields (CRF) with customized characteristics. A separate field of study, [14] contrasted machine learning algorithms with rule-based scripts for the purpose of identifying rhetorical roles. Rule-based scripts require substantially less training data. As a result of most of the earlier work using handcrafted features, which made models less effective, [15] suggested a model that employed the BiLSTM-CRF model with sent2vec features to categorize semantic roles in judgements from the Indian Court. The model was improved further by applying Hierarchical BiLSTM-CRF followed by Attention and embedding produced by pretrained BERT model over Indian Legal Judgement Documents in [16]. Also in the year 2023, [17] published a pretrained legal BERT model that is now accessible to the public.

Additionally, we observe the application of Multitask Learning (MTL) in [18] study, whereby label shift prediction (LSP-BiLSTM-CRF Model) serves as an auxiliary task and rhetorical role prediction as the primary task. For predicting and explaining court judgements, [19] have published a corpus of Indian legal papers.

Using the pretrained weights from BERT, [5] built a combined sentence representation that allows Transformer layers in BERT to directly utilize contextualized representations of all words in all sentences. They used BERT-specific tokens such [CLS] and [SEP], which are meant to capture the contextual/more significant information in a sentence and its surroundings [20].

A generalized graph convolution technique based on attention, known as the GAT mechanism, was proposed in [1]. To address shortcomings of earlier approaches that were using graph convolutions or their variants, an advanced neural network architecture that operates on graph data and uses self-attentional masked layers has been developed. In terms of contextual/neighborhood information capture in graph-structured data, this technique has produced state-of-the-art outcomes.

The contextualized sentence embeddings were obtained almost exclusively through the application of LSTM and its variants, such as Hier-BiLSTM and BiLSTM. However, as far as we are aware, GAT has not yet been used to produce contextual sentence embeddings in a task of automatically detecting semantic roles of sequential sentences in legal texts.

III. Dataset

The Semantic Role (SR) annotated legal case judgement dataset and corpus, which was published by [3], is used in this study. Rational behind selecting this dataset is that it contains legal judgements from Supreme Court of India, High Courts in different Indian states, and some district-level courts. The authors [3] have followed a transparent process of developing the corpus, that includes selection of multiple legal experts, data annotation, adjudication and then annotation quality assessment against Fleiss Kappa score [21] of 0.59.

There are 247 legal papers in the corpus from criminal and tax law categories. Total number of sentences in corpus is 28986. The division of documents into thematically coherent chunks known as SRs facilitates the computerised analysis of documents. The usage of 12 SRs and a NONE label is suggested in this paper. These are the following specifics and definitions for each SR:

A. Preamble (PREAMBLE)

This refers to the metadata for the court order document. A typical judgement would begin with the name of the court, the parties' information, the names of the solicitors and judges, and a headnote. Usually, a keyword like (JUDGEMENT or ORDER) would be used to round out this section. In the beginning of some papers, there are also HEADNOTES and ACTS sections. These are included in the Preamble as well.

B. Facts (FAC)

It describes the sequence of actions that resulted in the case being filed and how it progressed (for example, submitting a complaint at police station or an application to the court, etc.).

Submissions to the court, its processes, and a synopsis of earlier court decisions.

C. Ruling by Lower Court (RLC)

Instead of being filed directly in the higher courts, appeals are made from subordinate courts. The documents therefore include decisions made by the subordinate courts (District/High Court) considering the current appeal (to the Supreme Court or high court).

D. Issues (ISSUE)

In certain judgements, the crucial details on which the conclusion must be rendered are mentioned. The Court's framed legal questions are issues.

E. Argument by Petitioner (ARG PETITIONER)

Arguments made by the petitioners' solicitors. This category includes precedent cases that petitioner attorneys argue, but when the court analyses them later, they fall into either the relied-upon or not-relied-upon category.

F. Argument by Respondent (ARG RESPONDENT)

Respondent's solicitor's arguments. This category includes precedent cases that Respondent's attorneys argue, but when the court analyses them later, they fall into either the relied-upon or not-relied-upon category.

G. Analysis (ANALYSIS)

These are the court's perspectives. This includes the way the evidence, facts, preceding cases, and laws have been discussed in court. discussions on whether the law is relevant to the current instance. Court observations that are not legally binding. It is the parent tag for the following three tags: STATUTE, PRE-NOT RELIED, and PRE RLEIED. i.e., every statement that fits into one of these three categories should be labelled "ANALYSIS".

H. Statute (STA)

This comprises materials that the court uses to explain established laws; these texts may be drawn through variety of data sources, including Act and their sections, various article, formulated rule, issued order, summoned notice, Notifications, and direct quotes from acts. The law will be marked with the tags Analysis and Statute.

I. Precedent Relied (PRE RELIED)

Texts in which the court analyses earlier case materials, deliberations, and rulings that were cited in making final decisions. Both the tags Analysis + Precedent will be present on Precedent.

J. Precedent Not Relied (PRE NOT RELIED)

Texts where the court examines earlier case materials, arguments, and rulings that weren't used as the basis for its final judgements.

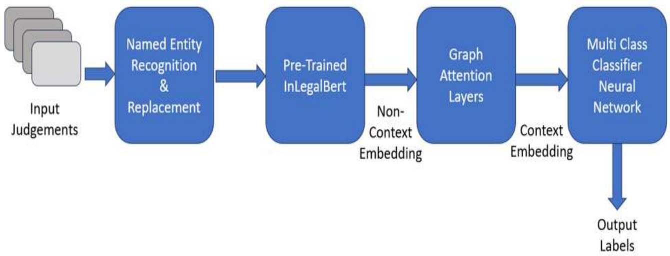


Fig. 1. *Semantic Role Labelling Model*

K. Ratio of the decision (Ratio)

Included in this is the primary justification offered for using any legal rule in relation to the legal question. The outcome of the court's analysis is this. It often emerges just before the choice is made. It is distinct from the "Ratio Decidendi" that is taught in the legal academic curriculum.

L. Ruling by Present Court (RPC)

Final judgement + conclusions + order of the court resulting from the logical/natural conclusion of the argument.

M. NONE

A sentence is marked as NONE if it does not fit into any of the categories.

The experiment performed in this work is carried out over these 247 judgements. The input is available in Json format. The Json contains unique id for each judgement and sentences from each judgement are mentioned under this unique id along with their respective semantic labels. The entire text of judgement is also mentioned in this Json.

This complete judgement text is useful for inputting to Named Entity recognition model.

TABLE I. LIST OF NAMED ENTITIES IN A LEGAL JUDGEMENT

Named Entity	*Description*
COURT	If taken directly from the preamble, the name of the court that issued the current verdict. If taken from judgement sentences, the name of any court mentioned
PETITIONER	Names of the present case's appellants
RESPONDENT	Names of respondents, defendants, and/or opponents from the present case
JUDGE	If taken from a preamble the names of the judges from the current case. Names of the judges from both the most recent and earlier instances, if known from judgement sentences
LAWYER	Names of the legal representatives for each party
DATE	Any date in the judgment
ORG	Name organizations apart from the court
GPE	Geopolitical Locations like states, cities, etc.
STATUTE	Names of ACT/Laws in Judgement

Named Entity	*Description*
PROVISION	Articles, articles, orders, and rules under a Statute
PRECEDENT	As precedent, the judgement cited all previous court decisions
CASE_NUMBER	Other case numbers mentioned in the judgment (apart from precedent)
WITNESS	List of witnesses in the most recent decision
OTHER_PERSON	All person except petitioner, respondent, judge, and witness

IV. PROPOSED WORK

In the following paper, we suggest a multi class classifier neural network to classify each sentence from a legal case judgement into one of the 13 labels. These 13 labels are semantic functions/roles of the sentences in the judgement. The semantic role of each sentence in legal judgement depends not only on its own structure but also on the structure of its neighboring sentences. As a result, the problem of SSC can be used to simulate semantic role labelling. *Fig. 1* serves as a reference for the model block diagram.

A. Input Judgements

Parsing the input dataset of semantic role labelled judgement sentences. This input is then given to the Named Entity Recognition (NER).

B. NER and Named Entity replacement

The development of additional legal artificial intelligence applications requires the identification of Named entities from legal documents. Compared to frequently used named entities like Person, Organization, and Location, named entities in legal texts are a little different and more precise. List of Named Entities can be referred from *Table 1*.

The model we utilized is a transformer-based legal NER model [22] that is pre-trained over an enormous volume of legal judgement. After passing judgement sentences to the NER model, we obtain the named entities from each judgement. We replace each entity from all sentences with its respective Named Entity label.

Named Entity Recognition and replacement makes the legal judgement sentences more generalized.

C. Generate non-contextual/Initial sentence embeddings

Output of the previous block are generalized named entities replaced in a sequence of sentences. But these sentences need to be converted to embeddings that are understood by classifier neural network. We use a pre trained Bert model that is trained over large corpus of legal case judgments from Indian law courts [2].

The output embeddings generated for each sentence from this pre trained model are independent of other sentences from the judgment. Hence these embeddings are called non-contextual embeddings. The detailed mechanism to obtain these non-contextual embeddings is shown in *Fig. 2*.

Every pre trained BERT model takes a sentence as its input and performs tokenization. During this tokenization step the sentence tokens are prepended by a special token [CLS] and appended with a special token [SEP]. The pre trained Bert generates embeddings for each token of this sentence. In the case of classification task, we need to use the embedding of this [CLS] token. As the embedding for [CLS] token carries more structural and semantic information for this task [5].

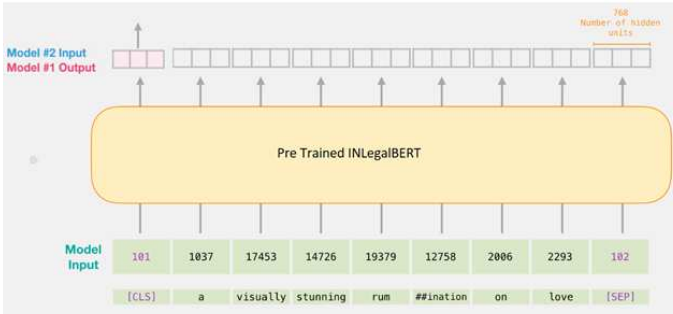


Fig. 2. *Model to generate initial non-contextual sentence embeddings.*

Fig. 3. *Sequence of Judgement sentences modelled as graph.*

Node Index	0	1	2	3
0	1	1	0	0
1	1	1	1	0
2	0	1	1	1
3	0	0	1	1

Fig. 4. *Sample Adjacency Matrix for judgement sentences.*

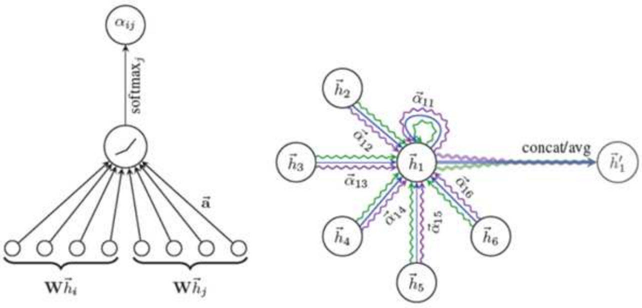
D. Generate context-aware sentence embeddings

Output of the previous step are non-contextual embeddings which do not carry any information of neighborhood sentences. To contextualize the embeddings, we propose to model the sequence of sentences in judgement as a graph. Refer to the graph shown in *Fig 3*.

If a judgment has n sentences, then the judgment graph will contain n nodes. Each node will be connected to its previous and next node, this will form a structure like a singly linked list. The edges will be undirected. The graph structure will be captured using an adjacency matrix of dimension $AdjMat_{NxN}$. Adjacency matrix for a judgement with 4 sentences will look like shown in *Fig. 4*.

As the embeddings from the previous step have dimensions as 1x768, each node will also have feature vector of same dimension and for n nodes the feature vector matrix is denoted by h_{Nx768}.

The input embeddings h_{Nx768} and $AdjMat_{NxN}$ are passed through 20 GAT [1] layers to get the context-aware embeddings h'_{Nx768} (h-prime) for a judgment. These embeddings are aware of their neighborhood sentences. The work of GAT to obtain h' can be understood from *Fig. 5*.



$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i \| \mathbf{W}\vec{h}_j]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\text{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i \| \mathbf{W}\vec{h}_k]\right)\right)}$$

$$\vec{h}'_i = \sigma\left(\sum_{j\in\mathcal{N}_i}\alpha_{ij}\mathbf{W}\vec{h}_j\right).$$

Fig. 5. *Working of a single GAT layer to generate context-aware sentence embeddings.*

Fig. 5 shows the approach of how the context-aware embeddings are calculated for a node using single GAT layer that can convolve information of immediate neighborhood for each node in the graph. Multiple such GAT layers can convolve information from many more neighborhood nodes into the current node of focus.

During h' calculation each node is given a different attention score. Each neighbor is not treated equally. This helps to convolve neighbors based on order of alpha score. Thus, helps model to remember important neighbor sentences.

Output of this step is context-aware embeddings h'_{Nx768} (h-prime) for all judgments.

E. Multi-class Classifier Neural Network

Previous step generates contextual embeddings for all the sentences in all judgements. These contextual embeddings are then used as inputs to a classifier neural network. It classifies the input embeddings into 13 output asemantic role labels.

Classifier has 4 layers and an output layer with dimension as given below.

(layer_1): Linear(in_features=768, out_features=1024, bias=True)

(layer_2): Linear(in_features=1024, out_features=512, bias=True)

(layer_3): Linear(in_features=512, out_features=128, bias=True)

(layer_4): Linear(in_features=128, out_features=64, bias=True)

(layer_out): Linear(in_features=64, out_features=13, bias=True)

(relu): ReLU()

(dropout): Dropout(p=0.2, inplace=False)

(batchnorm1): BatchNorm1d(1024, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)

(batchnorm2): BatchNorm1d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)

(batchnorm3): BatchNorm1d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)

(batchnorm4): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)

V. RESULTS

The model was tested against 5767 context-aware sentence embeddings. It overall performed better than the baseline model proposed in paper [3] to classify the input judgement sentences into respective semantic role labels.

Table 2 shows the F1 scores for our model. Since this is a multi-class classification model with unbalanced input for each label, F1 score can show better understanding of the model's learning. Standard Evaluation metrics like:

Precision (ratio of correctly predicted positive class to all predicted positive classes), **Recall** (ratio of correctly predicted positive classes to all existing positive classes), **F1 Score** = 2*(Precision * Recall) / (Precision + Recall)

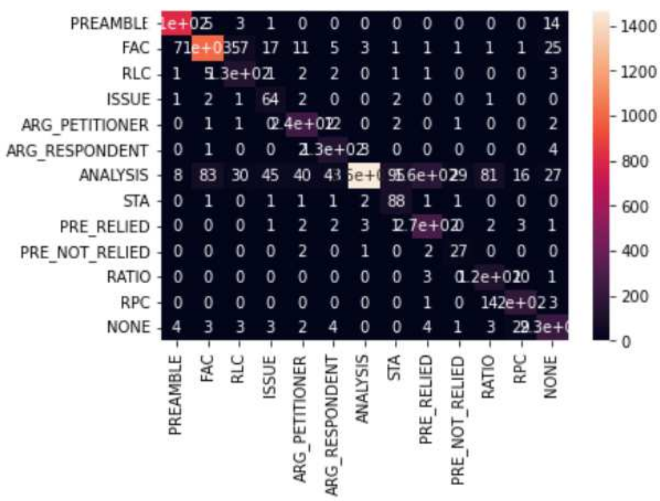


Fig. 6. *Confusion Matrix for the proposed model.*

TABLE II. F1 SCORES FOR EACH LABEL FROM PROPOSED MODEL

Label	*Precision*	*Recall*	*F1 Score*
PREAMBLE	00.97	00.97	00.97
FAC	00.91	00.89	**00.90**
RLC	00.58	00.89	**00.70**
ISSUE	00.48	00.88	00.62
ARG_PETITIONER	00.79	00.93	**00.85**
ARG_RESPONDENT	00.65	00.93	**00.77**
ANALYSIS	00.99	00.69	00.81
STA	00.46	00.92	00.62
PRE_RELIED	00.61	00.95	**00.74**
PRE_NOT_RELIED	00.45	00.84	**00.59**
RATIO	00.54	00.90	**00.67**
RPC	00.77	00.92	**00.84**
NONE	0.74	0.80	0.77
Overall weighted avg.	**0.88**	**0.83**	**0.84**

It is evident that our model performs better than the baseline model [3] in predicting 8 Semantic Role labels viz. FAC, RLC, ARG_PETITIONER, ARG_RESPONDENT, PRE_RELIED, PRE_NOT_RELIED, RATIO, and RPC. Our model performs like baseline model in predicting PREAMBLE. But the overall weighted F1 score **0.84** is way better than the baseline [3] model.

Confusion matrix shown in *Fig. 6* shows that our model is confused during predicting label ANALYSIS and misclassify into other labels as the judge tend to discuss some facts, issues, statutes, and rational as a part of analysis.

VI. SUMMARY AND NEXT STEPS

This work has proposed a GAT based Semantic Role labelling model that classifies sentences from legal judgements into 13 different Semantic Roles. We also used a pre-trained INLegalBERT model to generate initial sentence embeddings. The model shows similar trends in predicting the roles as human annotators for some of the roles. While it is still under performing for few roles.

This work can be extended further to finetune the model to perform better for the labels that are currently being misclassified. Also, the model can be trained over a greater number of judgements to increase its performance.

REFERENCES

[1] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, Yoshua Bengio. 2017. Graph Attention Networks. arXiv:1710.10903

[2] Paul, Shounak and Mandal, Arpan and Goyal, Pawan and Ghosh, Saptarshi. 2023. Pre-trained Language Models for the Legal Domain: A Case Study on Indian Law, Proceedings of 19th International Conference on Artificial Intelligence and Law - ICAIL 2023

[3] Prathamesh Kalamkar,Aman Tiwari, Astha Agarwal, Saurabh Karn, Smita Gupta, Vivek Raghavan, Ashutosh Modi. 2022. Corpus for Automatic Structuring of Legal Documents.

[4] 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.

[5] Arman Cohan, Iz Beltagy, Daniel King, Bhavana Dalvi, Daniel S. Weld (2019). Pretrained Language Models for Sequential Sentence Classification. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (2019) 3693-3699

[6] AdamWyner, Raquel Mochales-Palau, Marie-Francine Moens, and David Milward. 2010. Approaches to text mining arguments from legal cases. In Semantic processing of legal texts, pages 60–79. Springer.

[7] Vu Tran, Minh Le Nguyen, and Ken Satoh. 2019. Building legal case retrieval systems with lexical matching and summarization using a pre-trained phrase scoring model. In Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, pages 275–282.

[8] Nikolaos Lagos, Frederique Segond, Stefania Castellani, and Jacki O'Neill. 2010. Event extraction for legal case building and reasoning. In International Conference on Intelligent Information Processing, pages 92–101. Springer.

[9] Peter Jackson, Khalid Al-Kofahi, Alex Tyrrell, and Arun Vachher. 2003. Information extraction from case law and retrieval of prior cases. Artificial Intelligence, 150(1-2):239–290.

[10] Marie-Francine Moens, Caroline Uyttendaele, and Jos Dumortier. 1999. Abstracting of legal cases: the potential of clustering based on the selection of representative objects. Journal of the American Society for Information Science, 50(2):151–161.

[11] Giulia Venturi. 2012. Design and development of temis: a syntactically and semantically annotated corpus of italian legislative texts. In Proceedings of the Workshop on Semantic Processing of Legal Texts (SPLeT 2012), pages 1–12.

[12] Isar Nejadgholi, Renaud Bougueng, and Samuel Witherspoon. 2017. A semi-supervised training method for semantic search of legal facts in canadian immigration cases. In JURIX, pages 125–134.

[13] Jaromir Savelka and Kevin D Ashley. 2018. Segmenting us court decisions into functional and issue specific parts. In JURIX, pages 111–120.

[14] Vern R Walker, Krishnan Pillaipakkamnatt, Alexandra M Davidson, Marysa Linares, and Domenick J Pesce. 2019. Automatic classification of rhetorical roles for sentences: Comparing rule-based scripts with machine learning. In ASAIL@ ICAIL.

[15] Paheli Bhattacharya, Shounak Paul, Kripabandhu Ghosh, Saptarshi Ghosh, and Adam Wyner. 2019. Identification of rhetorical roles of sentences in indian legal judgments. CoRR, abs/1911.05405.

[16] Paheli Bhattacharya. 2021.DeepRhole: deep learning for rhetorical role labeling of sentences in legal case documents

[17] Paul, Shounak and Mandal, Arpan and Goyal, Pawan and Ghosh, Saptarshi. 2023. Pre-trained Language Models for the Legal Domain: A Case Study on Indian Law, Proceedings of 19th International Conference on Artificial Intelligence and Law - ICAIL 2023

[18] Malik, V., Sanjay, R., Guha, S. K., Nigam, S. K., Hazarika, A., Bhattacharya, A., and Modi, A. (2021a). Semantic Segmentation of Legal Documents via Rhetorical Roles. CoRR, abs/2112.01836.

[19] Malik, V., Sanjay, R., Nigam, S. K., Ghosh, K., Guha, S. K., Bhattacharya, A., and Modi, A. (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, August. Association for Computational Linguistics.

[20] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. CoRR, abs/1810.04805.

[21] Fleiss, J. L., Levin, B., and Paik, M. C. (2013). Statistical methods for rates and proportions. john wiley & sons

[22] Prathamesh Kalamkar, Astha Agarwal, Aman Tiwari, Smita Gupta, Saurabh Karn, Vivek Raghavan. 2022. Named Entity Recognition in Indian court judgments.