Last Meeting Update

Paper: Identification of Rhetorical Roles of Sentences in Indian Legal Judgments
Task: Semantic Segmentation (sentence level)

Facts (FAC): Chronology of events that led to filing the case, and how the case evolved over time
Ruling by Lower Court (RLC): Verdict of the lower Court and the ratio behind the judgement
Argument (ARG): Court's discussion on the law that is applicable to the set of proven facts
Statute (STA): Established laws from Acts , Sections, Articles, Rules, Order, Notices, Notifications
Precedent (PRE): Prior case documents. Instructions similar to statute citations
Ratio of the decision (Ratio): Reason given for the application of any legal principle
Ruling by Present Court (RPC): Ultimate decision / conclusion of the Court

Trained Hier-BiLSTM-CRF with Pretrained embeddings with their data
Model tested on our data [1 ADJ ; 1 SAT]

Table 3. Macro Precision, Recall and F-score of the baseline methods and neural network-based methods. Best performances highlighted in boldface.

Category	Method	Variations	Precision	Recall	F-score
Baselines (CRF with handcrafted features)	Features from [2]	-	0.4138	0.3308	0.4054
	Features from [4]	-	0.4580	0.4196	0.3250
	Features from [4] and [2]	-	0.5070	0.4358	0.4352
Neural models	Hier-BiLSTM	Pretrained emb	0.8168	0.7852	0.7968
		Random initialization	0.5358	0.5254	0.5236
	Hier-BiLSTM-CRF	Pretrained emb	**0.8396**	**0.8098**	**0.8208**
		Random initialization	0.6528	0.5524	0.5784

Directions for Legal Case Files

Semantic Segmentation : Segment out context/relevant info
Text Simplification: Simplify text and then extract context/relevant info

Semantic Segmentation - I	**Semantic Segmentation - II**	**Text Simplification**
1. Use the base papers' labels that perform well on our data i.e if they can classify relevant information	1. Define our own labels 2. Use the annotation approach & methodology from the base paper	1. Simplify the entire document to improve understanding and extraction 2. Then extract the information
Cons: 1. Irrelevant labels 2. Not trained using our data	Cons: 1. Annotation of about 9k sentences from 50 relevant case files 2. Crawl 50k documents for sentence embedding pretraining	Cons: 1. Challenging with limited data 2. No relevant pre existing dataset for this task
Next Step: Simplify extracted context/info and then map to regulation regulation	Next Step: Simplify extracted context/info and then map to regulation regulation	Next Step: Map to regulation regulation

Original Sentence	Simplified Sentence
Owls are the order Strigiformes, comprising 200 bird of prey species.	An owl is a bird. There are about 200 kinds of owls.

Next Meeting:

1. Analysis of the results using their best model on our data
2. Relevant papers for text simplification (legal domain) that can help
3. Any relevant datasets available