Legal Case Document Summarization: Extractive and Abstractive Methods and their Evaluation

Abhay Shukla[1] Paheli Bhattacharya[1] **Soham Poddar[1]** Rajdeep Mukherjee[1]
Kripabandhu Ghosh[2] Pawan Goyal[1] Saptarshi Ghosh[1]

[1]Indian Institute of Technology, Kharagpur [2]Indian Institute of Science Education and Research, Kolkata

Motivation and Contributions

- Law practitioners have to read through hundreds of case judgements/rulings, but case documents are generally very long and complex.
- Developed three legal case judgement summarization datasets from case documents from the Indian and UK Supreme Courts
- Reproduce/apply representative methods from several families of summarization models on these datasets, including some state-of-the-art models.
- First paper that analyses the relative performances of a wide spectrum of extractive vs abstractive summarizers on legal documents

Main Insights

- In many cases, we observe general (domain-agnostic) methods to perform better than domain-specific methods.
- Using models pretrained on legal corpora, like Legal-Pegasus, consistently improves performance.
- Chunking-based approach performs better for legal documents, especially with fine-tuning
- *Law Experts* advise to not only evaluate the full-document summaries, but also representation of different rhetorical segments in a legal case document (such as Facts, Final Judgement)
- Even though ROUGE scores achieved by the best extractive models are at par with those achieved by the best abstractive models, the practitioners often prefer the extractive summaries over the abstractive ones.

Datasets

- **IN-Abs:** Indian case documents with abstractive summaries from Legal Information Institute of India website.
- **IN-Ext:** Test dataset annotated by 2 LLB graduates of Indian case documents with extractive summaries.
- **UK-Abs:** UK case documents with abstractive summaries from the UK Supreme court website.

Our Dataset Statistics

Dataset	Compression Ratio	Avg # Tokens		#Docs	
		Doc	Summ	Test	Train
IN-Ext	0.31	5,389	1,670	50	7030
IN-Abs	0.24	4,378	1,051	100	
UK-Abs	0.11	14,296	1,573	100	693

Methods

- **Extractive**
  - Unsupervised and Supervised models
  - Domain-Agnostic and Domain-Specific models
  - **Label Selection for Supervised:** Greedily pick sentences according to Avg. ROUGE-1,2&L scores.
- **Abstractive**
  - *Pretrained*
    - Split into chunks of N words and summarize each of them. (N = Max Input Sequence length of a model)
    - Models meant for long documents – Longformer
    - Hybrid Extractive and Abstractive models.
  - *Similarity methods used to generate fine-tuning data*
    - MCS - Mean of token-level embeddings obtained using SBERT

Results on the IN-Ext dataset

Algorithm	ROUGE Scores			BERTScore
	R-1	R-2	R-L	
Extractive Methods (U: Unsupervised, S: Supervised)				
Pacsum_bert (U)	0.590	**0.410**	0.335	**0.879**
LetSum (U)	**0.591**	0.401	**0.391**	0.875
SummaRunner (S)	0.532	0.334	0.269	0.829
BERT-Ext (S)	0.589	0.398	0.292	0.85
Finetuned Abstractive Methods				
BART_MCS	0.557	0.322	0.404	**0.868**
Legal-Pegasus_MCS	**0.575**	**0.351**	**0.419**	0.864
Legal-LED	0.471	0.26	0.341	0.863

Experimental Setup and Evaluation

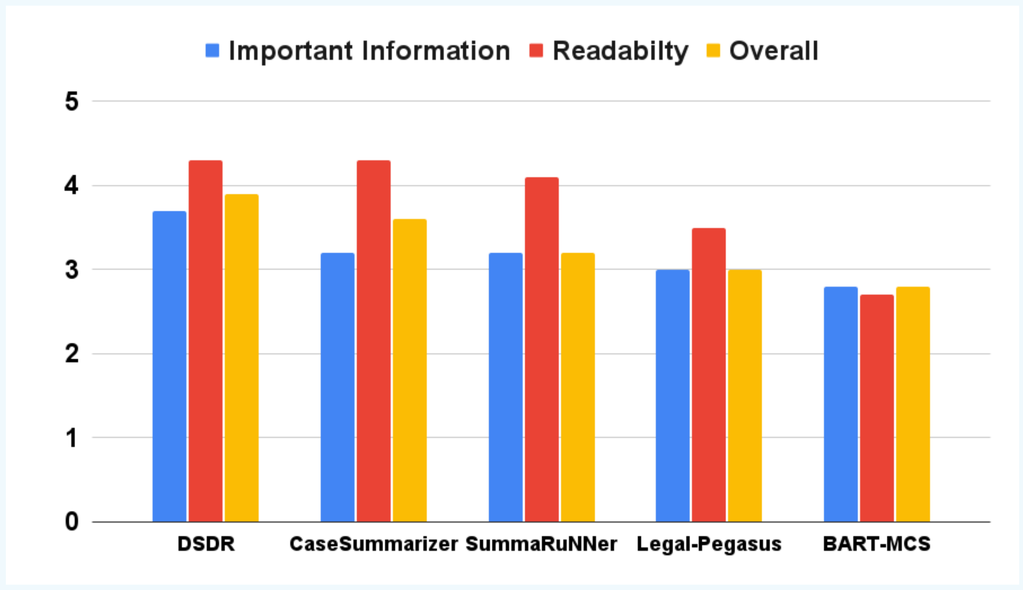
- **Target Summary Length:** #words in the reference summary.
- **Automatic Evaluation**
  - ROUGE-1,2,L and BERTScore
  - Document-wise and segment wise evaluation
- **Human Expert Evaluation**
  - Quality of Important Information, Readability and Overall score on 1-5 Likert Scale given by 3 Law Experts.
  - Document-wise and segment-wise evaluation

* *Segment wise evaluation in paper*

Results on the UK-Abs dataset

Algorithm	ROUGE Scores			BERTScore
	R-1	R-2	R-L	
Extractive Methods (U: Unsupervised, S: Supervised)				
DSDR (U)	0.484	0.174	0.221	0.832
CaseSummarizer (U)	0.445	0.166	0.227	0.835
SummaRunner (S)	**0.502**	**0.205**	**0.237**	**0.846**
Finetuned Abstractive Methods				
BART_MCS	**0.496**	**0.188**	**0.271**	**0.848**
Legal-Pegasus_MCS	0.476	0.171	0.261	0.838
Legal-LED	0.482	0.186	0.264	0.851

Results of Evaluation by Human Experts



More Details about our Work

Paper and Dataset available at:
arxiv.org/abs/2210.07544
github.com/Law-AI/summarization