A (short) Survey of State-of-the-Art in Legal Document Processing

Debasis Ganguly, University of Glasgow

Paheli Bhattacharya, IIT Kharagpur

Overview of Challenges in Legal-NLP tasks

- Longer documents as compared to other domains (e.g., news).
  - Over 4000 words on an average
- Difficult to create large datasets required for supervised NLP/IR methods
  - Requires annotators having expertise in Law -- expensive, time-consuming
- Writing style and document structure varies widely from one country to another

Popular Research Problems in Legal-NLP

- Retrieval / recommendation of precedents
- Legal judgement prediction
- Legal Statute Identification: Identifying relevant statutes from given facts
- Rhetorical labelling / semantic segmentation of case judgements
- Summarization of case judgements and other legal documents
- Finding similarity between case documents
- Legal citation network analysis
- Explainability of AI-Law models
- Training language models specific to the Law domain
- ... and many more

Legal Judgement Prediction

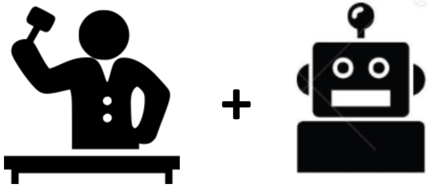
Dr. Debasis Ganguly

University of Glasgow

Why Judgment Prediction?

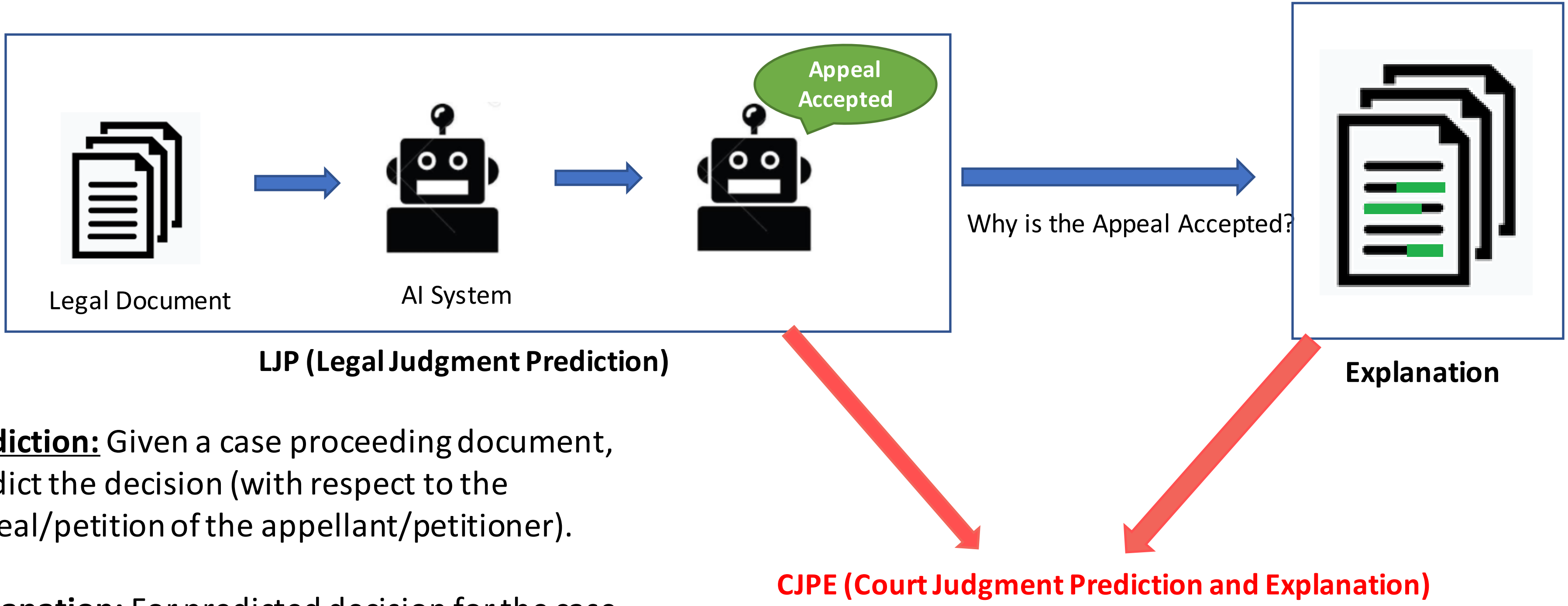
- In many highly populated countries (e.g., India), there exist a vast number of **pending backlog of legal cases** (Katju, 2019) that impede the judicial process.

- A system capable of **assisting/augmenting a judge** by suggesting the outcome of the court case will be useful.



- For the system to be of **practical utility**, in addition to prediction, it should **explain that outcome** in terms of how a legal practitioner understands the legal process.

CJPE Task



<u>Prediction:</u> Given a case proceeding document, predict the decision (with respect to the appeal/petition of the appellant/petitioner).

<u>Explanation:</u> For predicted decision for the case, explain the decision by predicting important sentences that lead to the decision.

Legal Text vs Non-Legal Text

Noisy and unstructured text

20. Section 10-A of the Act deals with permission for establishment of new medical college, new course of study, etc. Sub-section (7) of Section 10-A reads as follows:-

"(7) The Council, while making its recommendations under clause (b) of sub-

section (3) and the Central Government, while passing an order, either approving or disapproving the scheme under sub-section (4), shall have due regard to the following factors, namely—

(a) whether the proposed medical college or the existing medical college seeking to open a new or higher course of study or training, would be in a position to offer the minimum standards of medical education as prescribed by the Council under Section 19A or, as the case may be, under Section 20 in the case of postgraduate medical education.

Document contains some relevant information in the main text.

Legal Text vs Non-Legal Text

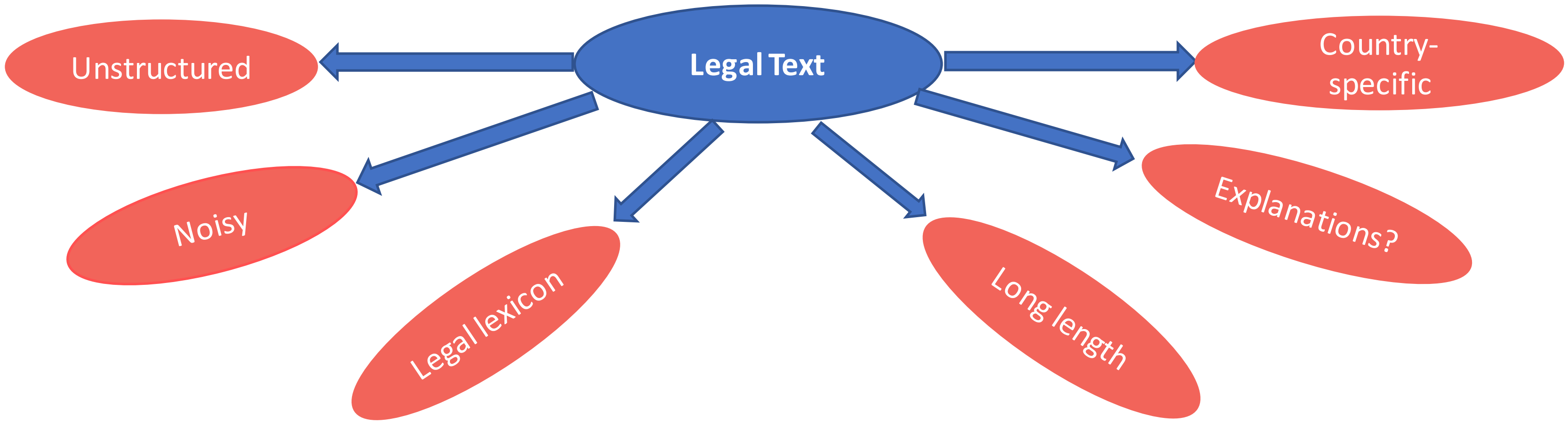
Domain Specific Lexicon

4. As the factual narration would evince, on 10th February, 2016, a team of assessors of the respondent No. 2 conducted verification assessment for grant of LOP for the academic year 2016-17. In the mean time, the Constitution Bench in Modern Dental College and Research Center and others v. State of Madhya Pradesh and others 1 constituted the Oversight Committee headed by Justice R.M. Lodha former CJI to oversee the functioning of the MCI. We shall refer the relevant paragraphs of the said judgment at a later stage. On 13th May, 2016, the report of the assessors team was considered by the Executive Committee of the respondent No.2 in its meeting dated 13.05.2016 and on 14.5.2016 the MCI recommended the disapproval of the scheme of the petitioner under Section 10-A of the Act for the academic year 2016-17. However, after Oversight Committee was constituted, the Central Government issued a public notice informing all the Medical 1 (2016) 7 SCC 353 Colleges to submit a compliance report concerning their respective colleges who had applied for LOP for 2016-17. As the facts would unfold, the 1st respondent sent the compliance report along with the reply of the MCI to the Oversight Committee for consideration which on 11.08.2016 approved the same for the year 2016-17 imposing certain conditions.

Legal and Formal text terminology

Legal Text vs Non-Legal Text

- Legal documents are quite long (over 3,000 tokens on average).
- Explaining prediction in legal documents is considerably more challenging.
- Legal text/proceedings are country-specific



Datasets for LJP

- English LJP datasets: European Court of Human Rights
  - 11.5k Cases from ECHR; human rights article violation, which article violated
  - https://archive.org/details/ECHR-ACL2019
- ILDC: Indian Legal Data Corpus
  - 35k Cases from the Indian Supreme Court; predict the decision with respect to the petitioner https://github.com/Exploration-Lab/CJPE
- HLDC: Hindi Legal Documents Corpus
  - 912k Cases from district courts from state of Uttar Pradesh (U.P.); Bail Prediction
  - https://github.com/Exploration-Lab/HLDC
- CAIL 2018: Chinese AI and Law challenge dataset
  - 2.6M Cases from Supreme People's Court of China; Charge Prediction, Relevant Articles Prediction
  - https://github.com/Exploration-Lab/HLDC
- English UK datasets: U.K. Supreme court
  - 5k Cases from U.K. Supreme court; predict the decision with respect to the petitioner
  - https://github.com/BStricks/legal_document_classifier_V2

English LJP datasets: European Court of Human Rights

Neural Legal Judgment Prediction in English, Ilias Chalkidis et al., ACL (2019)

Legal Judgment Prediction (LJP)- ECHR[1]

- **Input:**
    - Fact description of a case

- **Output:** **(Judgment Results)**
    1. Violation of a human rights article (binary classification)
    2. Type of violation, if any (multi-label classification, #classes = 66)
    3. Case importance detection (regression task)
        - scale from 1 (key case) to 4 (unimportant)

1. Neural Legal Judgment Prediction in English, Ilias Chalkidis et al., ACL (2019)

Statistics of the ECHR Dataset 2018

- Release a new English Cases from the **European** Court of Human Rights.

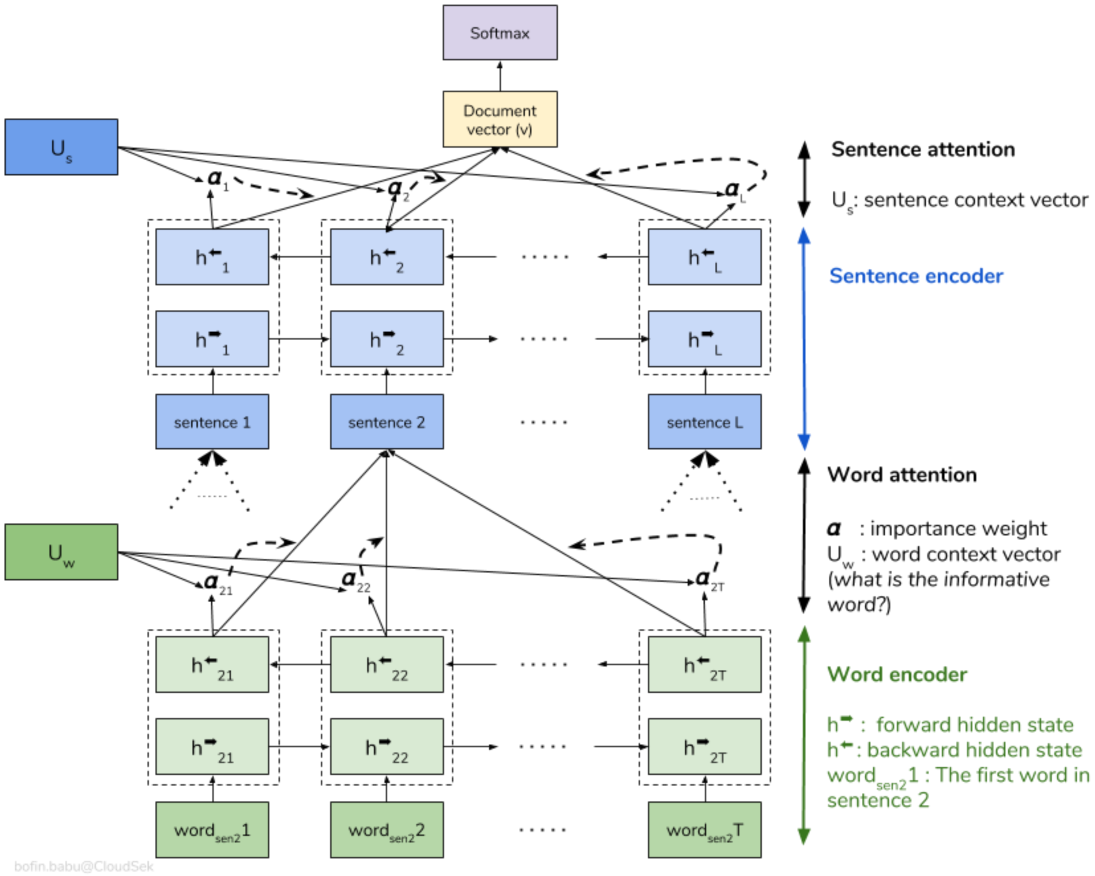
Subset	Cases	Year
Train	7100	1959 – 2013
Dev	1380	1959 – 2013
Test	2998	2014 – 2018
Total	**11478 = ~11.5K**	

Models

- BiGRU-Attention[1] (**B**idirectional **G**ated **R**ecurrent **U**nit)
- HAN[2] (**H**ierarchical **A**ttention **N**etwork)
- BERT[3] (**B**idirectional **E**ncoder **R**epresentations from **T**ransformers)
- HIER-BERT (bypasses BERT's length limitation)

1. XU et al., ICML, 2015, Show, attend and tell: Neural image caption generation with visual attention
2. Zichao Yang et al., NAACL-HLT, 2016, Hierarchical Attention Networks for Document Classification
3. Jacob Devlin et al., NAACL-HLT, 2019, BERT: pre-training of deep bidirectional transformers for language understanding

Hierarchical Attention Networks[1]



1. Zichao Yang et al., NAACL-HLT, 2016, Hierarchical Attention Networks for Document Classification

Results for Task-1 (Binary Violation Results)

Model	Precision	Recall	F1
Non - Anonymized			
BiGRU-Att	87.1	77.2	79.5
HAN	88.2	78.0	80.5
BERT	24.0	50.0	17.0
Hier-BERT	**90.4**	**79.3**	**82.0**
Anonymized			
BiGRU-Att	**87.0**	70.5	70.9
HAN	85.2	**78.3**	**80.2**
BERT	17.0	50.0	25.4
Hier-BERT	85.2	78.1	80.1

Results for Task-1 (Binary Violation Results)

Case ID: 001-148227 **Violated Articles:** Article 3 **Predicted Violation:** YES (0.97%)

1. The applicant was born in 1955 and lives in Kharkiv .

2 . On 5 May 2004 the applicant was arrested by four police officers on suspicion of bribe - taking .
The police officers took him to the Kharkiv Dzerzhynskyy District Police Station , where he was held overnight .
According to the applicant , the police officers beat him for several hours , forcing him to confess .

3. On 6 May 2004 the applicant was taken to the Kharkiv City Prosecutor 's Office . He complained of ill-treatment
to a senior prosecutor from the above office . The prosecutor referred the applicant for a forensic medical examination .

4. On 7 May 2004 the applicant was diagnosed with concussion and admitted to hospital .

5. On 8 May 2004 the applicant underwent a forensic medical examination , which established that he had numerous
bruises on his face , chest , legs and arms , as well as a damaged tooth .

6. On 11 May 2004 criminal proceedings were instituted against the applicant on charges of bribe-taking . They were
eventually terminated on 27 April 2007 for lack of corpus delicti .

7. On 2 June 2004 the applicant lodged another complaint of ill - treatment with the Kharkiv City Prosecutor 's Office .

Attention over words (coloured words) and facts (vertical heat bars) as produced by HAN

Limitation

- No justification for neural models' predictions (Not Explainable)
- Attention scores are still far from being justifications (Not Trustworthy)
- Not enough data to train neural models (Small Dataset)

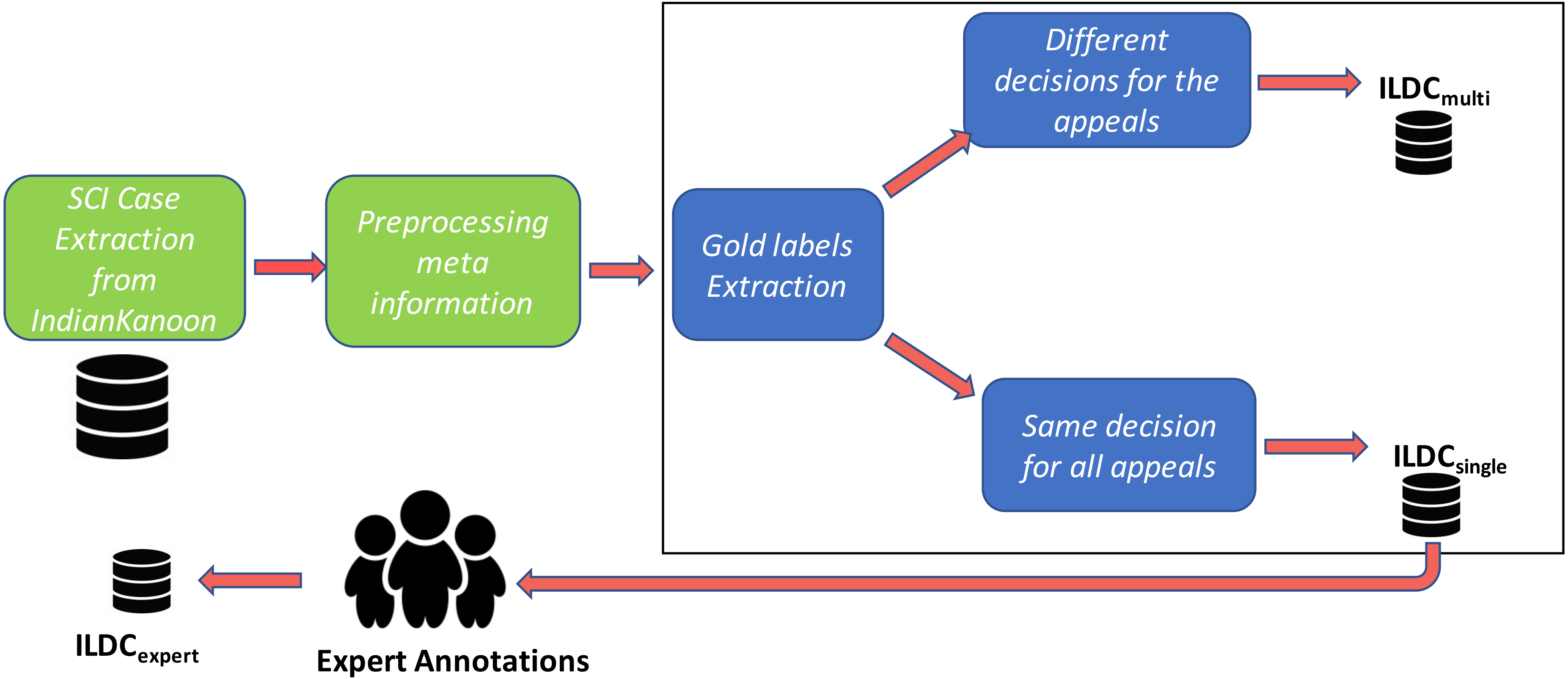
ILDC for CJPE: Indian dataset (English)

ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation, Malik et al., ACL (2021)

How CJPE is different?

- Predictive algorithms to explain decisions in CJPE task and for evaluation provide a separate test set with gold annotations.
- In **LJP** facts are explicitly provided but in **CJPE** we consider the only unannotated **unstructured** case (except the judgement)
- The **ILDC corpus** is the **largest legal** corpus for the Indian setting with **34,816** documents

ILDC Creation



ILDC Statistics

Corpus	Average tokens	Number of documents (%Accepted cases)		
		Train	Validation	Test
$\text{ILDC}_{\text{multi}}$	3231	32305 (41.43%)	994 (50%)	1517 (50.23%)
$\text{ILDC}_{\text{single}}$	3884	5082 (38.08%)		
$\text{ILDC}_{\text{expert}}$	2894	56 (51.78%)		

Temporal Aspects and Bias handling in ILDC

- Train-Test-Validation split was *NOT* based on any temporal consideration: (Why?)

  - **System's aim** is to **identify standard features** of judgments
  - **If not**, such a system is likely to **fail** in application since, in the future **laws might get amended or replaced.**

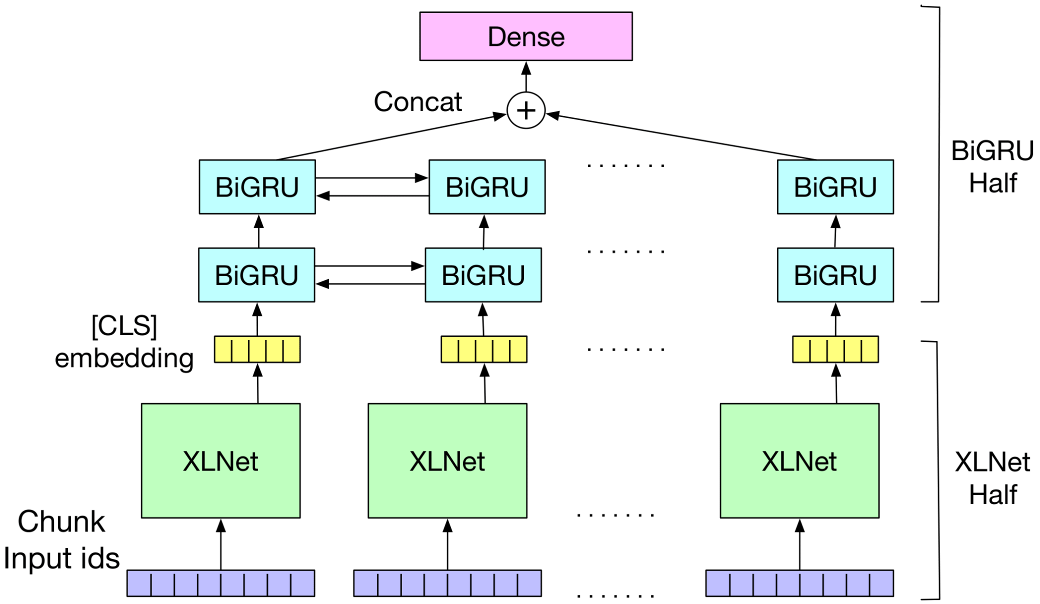
- We have not made any specific choice about any **specific law** or any **category of cases** (random sampling).

- Names of the **judge(s), appellants, petitioners**, etc., were **anonymized**. (**Why?**):
  - A **strong indicator** of the case outcome
  - The system should focus on the **facts and applicable law.**

Judgement Prediction

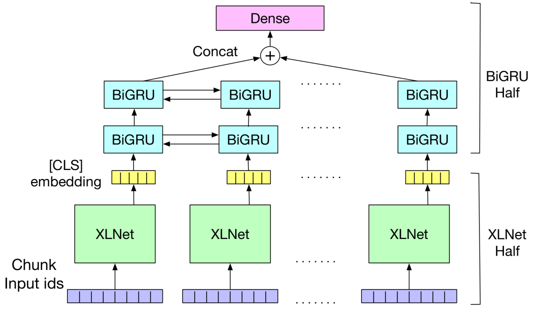
Model	Macro F1(%)	Accuracy(%)
Sequential Models		
Sent2vec + BiGRU + attention	59.66	58.31
GLoVe + BiGRU + attention	64.35	60.75
HAN	59.77	59.53
Transformers Models		
BERT Base (first 512 tokens)	59.06	57.65
BERT Base (last 512 tokens)	68.31	67.24
RoBERTa	71.77	71.26
XLNet	71.07	70.01
Hierarchical Models		
XLNet + BiGRU	**77.79**	**77.78**
Hierarchical Models with Attention		
XLNet + BiGRU + attention	77.07	77.01

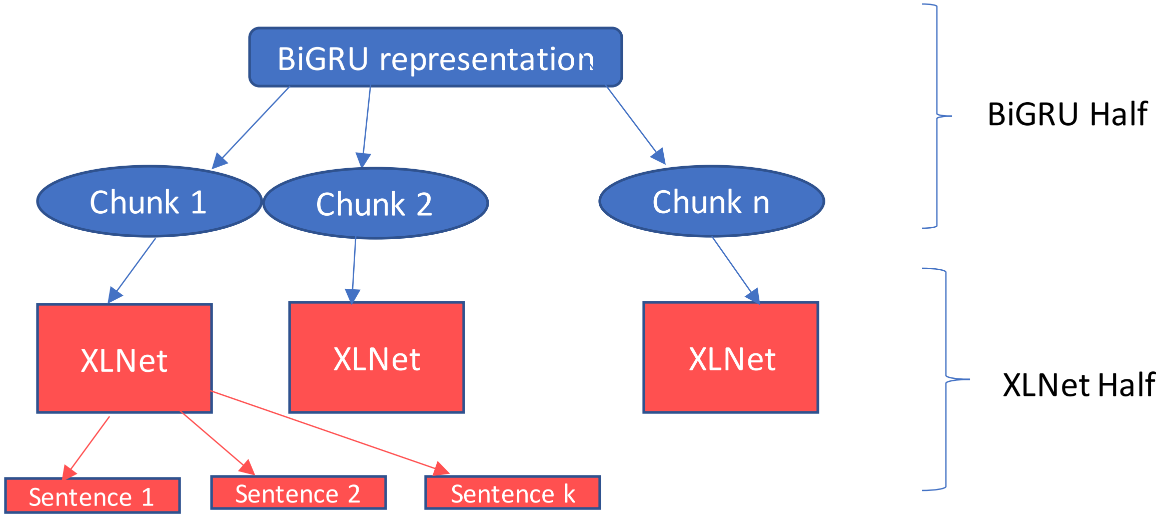
Explanation Methodology



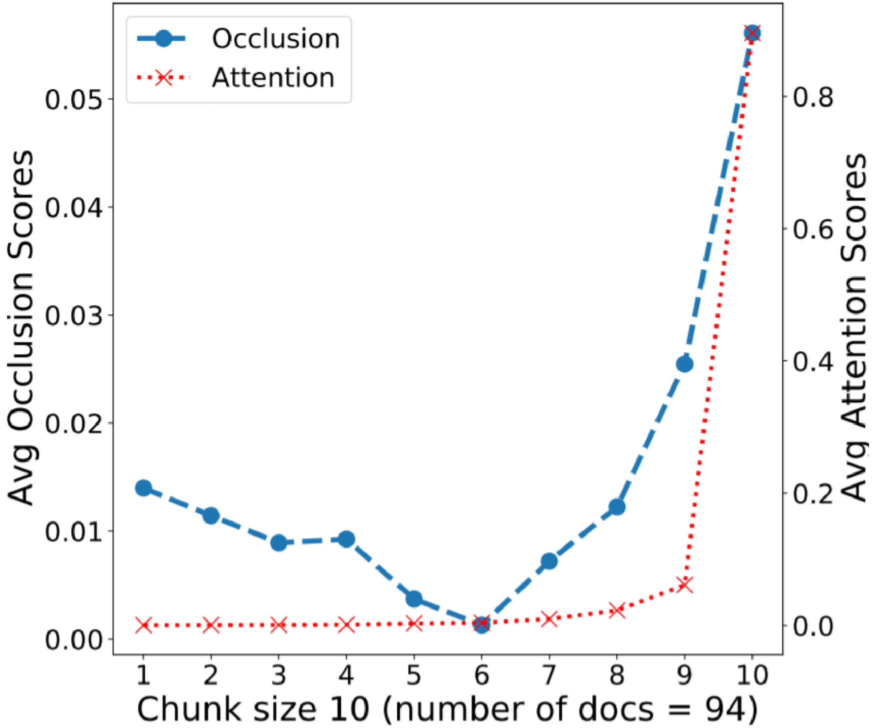
Explanation Methodology

- **Mask** a chunk embedding
- Calculate **masked probability of the label**
- **Compare** with **unmasked probability** & get the chunk importance score.





Insights from chunk scores



Judgment explanation

Metric	Explainability Model vs Experts				
	Expert				
	1	2	3	4	5
Jaccard Similarity	0.333	0.317	0.328	0.324	0.318
Overlap-Min	0.744	0.589	0.81	0.834	0.617
ROUGE-L	0.439	0.407	0.423	0.444	0.407
BLEU	0.16	0.28	0.099	0.093	0.248
Meteor	0.22	0.3	0.18	0.177	0.279

Summarization of Legal Case Documents

Paheli Bhattacharya

Indian Institute of Technology Kharagpur, India

Motivation : Automatic Case Document Summarization

- Lengthy, free-flowing, unstructured, dense legal text

- Reading and comprehending the full documents is difficult even for lawyers

- Automatic summarization systems

Case Document (https://indiankanoon.org/doc/749642/)

Kelvinator Of India Ltd vs The State Of Haryana on 23 August, 1973

Equivalent citations: 1973 AIR 2526, 1974 SCR (1) 463

Author: H R Khanna

Bench: Khanna, Hans Raj

CIVIL APPELLATE JURISDICTION : Civil Appeal No. 2005 (NT) of 1972:.

Appeal by special leave from the order dated the 11th April 1972 of the Punjab and Haryana High Court at Chandigarh, in General Sales Tax Reference No. 8 of 1970.

N.A. Palkhivala, H. L. Sibal, J. B. Dadachanji, A. K. Verma, Kapil Sibal and S. C. Agnihotri, for the appellant. Y.M. Tarkunde, Narendra Goswami and M. N. Shroff, for the respondent.

S. T. Desai, and 1. N. Shroff, for the intervener.

The Judgment of the Court was delivered by KHANNA, J. This appeal by special leave by M/s. Kelvinator of India Ltd. is directed against the judgment of Punjab & Haryana High Court whereby that court answered the following question referred to it by the Sales Tax Tribunal Haryana in favour of the department and against the appellant "Whether on the facts and circumstances of the case, the agreement between M/s. Kelvinator of India (Assessee) M/s. Spencer & Co. Ltd., Messrs Blue Star Engineering Co,, and M/s. General Equipment Ltd., in pursuance of which the refrigerators manufactured by M/s, Kelvinator of India at Faridabad moved to Delhi were merely for distribution of goods between the principal and his agents or were agreements of sale between two parties?"

The matter relates to the assessment year 1965-66, i.e. the period from April 1, 1965 to March 31, 1966. The appellant company has a, factory at Faridabad in Haryana. It manufactures refrigerators, deep freezers, compressors and other similar articles. The factory went into production in 1964.

....

We are concerned in the present case with clause (a) ....The word "occasions" is used as a verb and means 'to cause or to be the immediate cause of'. In the case of Tata Iron and Steel Co. Ltd. v. S. R. Sarkar and Ors; (1) Shah J. (as he then was) speaking for the majority observed that a transaction of sale is subject to tax under the Act on the completion of the sale. A mere contract of sale is not a sale within the definition of "sale" in section 2(g). A sale being, by the definition, transfer of property becomes taxable under section 3(a) "if the movement of goods from one State to another is under a covenant or incident of the contract of sale". In Ben Gorm Nilgiri Plantations Co. Cooncor & Ors. v. Sales Tax Officer, Special Circle, Ernakulam & Ors(2) this Court dealt with the provisions of

.....

We accordingly accept the appeal and set aside the judgment of the High Court ..... The appellant shall be entitled to the costs from the respondent of this Court as well as in the High Court. V.P.S.
Appeal allowed.

7580 words

Summary

Section 6 of the Central Sales Tax Act, 1956, makes every dealer liable for payment of tax under the Act on all sales-effected by him during interstate trade or commerce. A sale of' goods can be held to have taken place in the course of interstate trade under s. 3 (a) if it can be shown that the sale has occasioned the movement of goods from one State to another, that is, if, (i) there is a sale, (ii) there is actual movement of goods from one State to another, and (iii) the sale and movement of the goods formed integral parts of the same transaction. A sale being. by the definition in the Act, transfer of property, to be exigible to tax under the Act it must be shown that tie movement was the result of a covenant or incident of the contract of sale' The movement of goods which takes place independently of a contract of sale would not fall within the ambit of s. 3(a). There must be acontract of sole preceding the

....

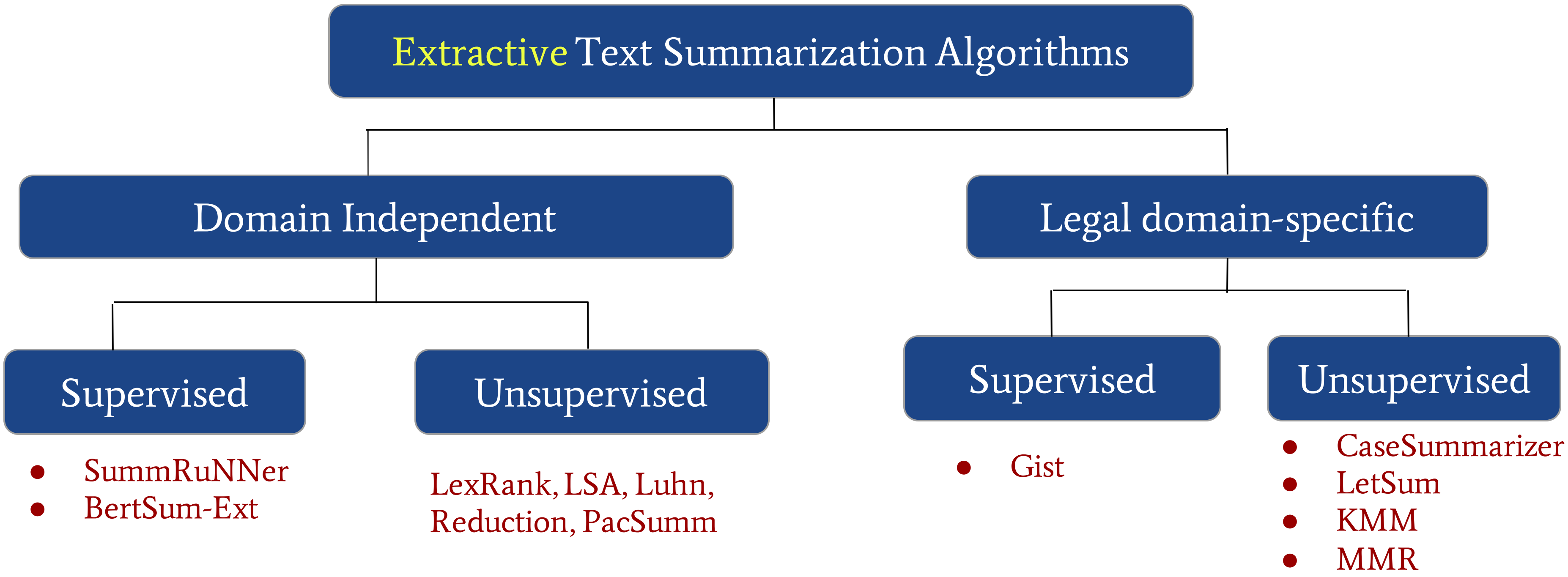
....

Therefore, it was the orders which were placed in Delhi by the distributors and the acceptance thereof by the appellant that resulted in the mutual agreement of sale. The distribution agreement with each distributor only provided the framework within which the different contracts of sale were to be entered into by the distributor with the appellant, and the distribution agreement and contract of sale were distinct transactions. [474B-G] (b)It is not correct to say that the distributor with whom the first agreement was entered into was bound to purchase all the products of the appellant. The words 'the sale would be as mutually agreed upon from time to time' Would lose all significance if that

....

[477 H-478H] (2)There was no movement of refrigerators from Faridabad to Delhi under a contract of sale. [476G] 465 (a)If there is a choice before the parties of so arranging their matters that lift one case they would have to incur liability to, pay tax and in the other case the liability to pay tax would not be attracted, they would Prefer the latter course'There is nothing illegal or impermissible to a party so arranging its affairsthat the liability to pay tax would not be attracted or would be reduced. [476C-D] The appellant could have sold the refrigerators at either of the two case, it was not the distributor but the insurer who would have to bear the loss and the transit insurance expenses were borne by the appellant.

1770 words


Extractive Text Summarization Algorithms
Domain Independent
Legal domain-specific
Supervised
Unsupervised
Supervised
Unsupervised
SummRuNNer
BertSum-Ext
LexRank, LSA, Luhn,
Reduction, PacSumm
Gist
CaseSummarizer
LetSum
KMM
MMR

Abstractive Summarization Methods

- Pointer Generator
- BERTSum-Abs
- BART
- Pegasus
- BigBird
- Longformer
- ... and others

- Pre-trained versions – usually trained over news summarization datasets
- Fine-tuned versions – fine-tuning over legal (doc, summary) pairs likely to yield better results

Abstractive Summarization Methods

- Pointer Generator
- BERTSum-Abs
- BART
- Pegasus
- BigBird
- Longformer
- ... and others

- Pre-trained versions – usually trained over news summarization datasets

- Fine-tuned versions – fine-tuning over legal (doc, summary) pairs likely to yield better results

→ Legal-Pegasus – https://huggingface.co/nsi319/legal-pegasus

→ Legal-Longformer https://huggingface.co/nsi319/legal-led-base-16384
Already fine-tuned over legal summarization data (litigation releases and complaints from US courts)

Datasets for Legal Case Document Summarization

Dataset	# documents	Summary Type	Avg #words	Avg compression ratio	Jurisdiction
IN-Abs	7130	Abstractive	4,378	0.24	India
UK-Abs	793	Abstractive	14,296	0.11	UK
IN-Ext	50	Extractive	5,389	0.3	India

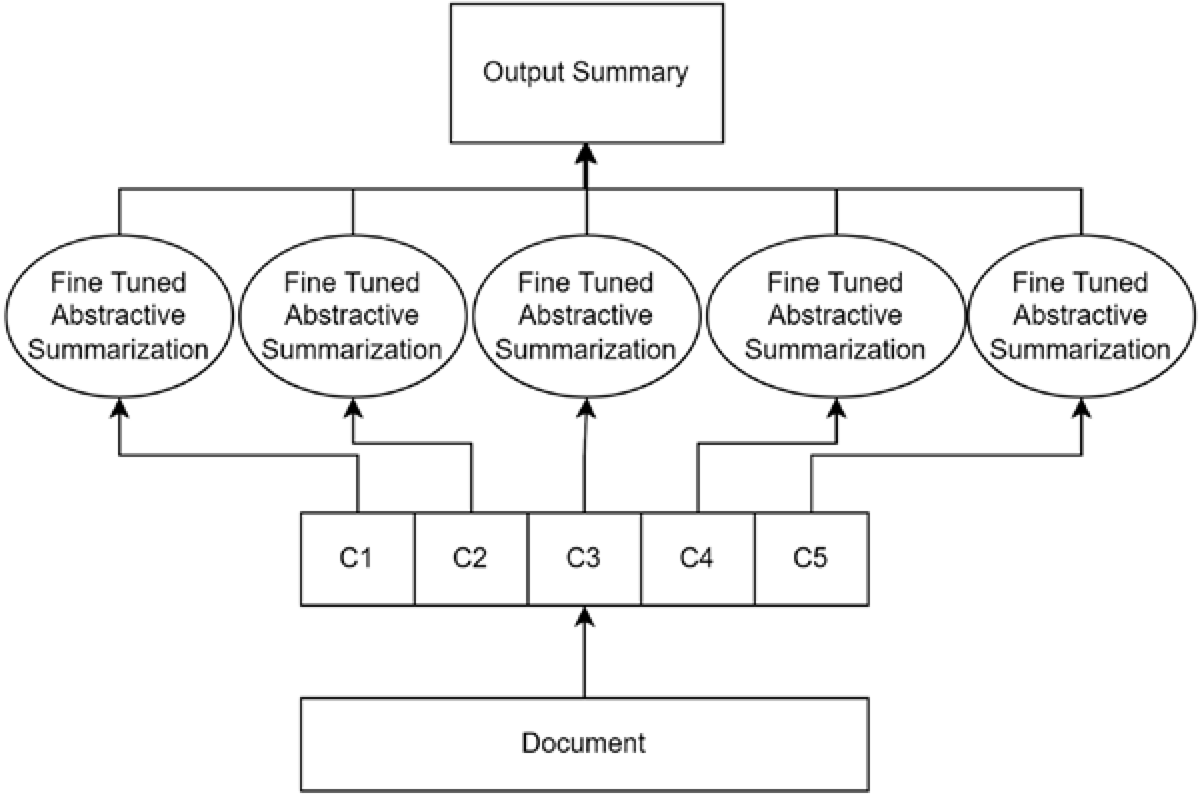
Challenge in Applying Abstractive Models on Legal Docs

- Limit on the number of input tokens
- Many case docs do not fit in any of these models except Longformer
    - Several docs in UK-Abs do not fit fully even in Longformer

Model	Input token limit
Pointer-Generator	400
BertSum-Abs	512
Pegasus	512
BART	1024
BigBird	4*1024
Longformer	16*1024

Dataset	**Avg #words**
IN-Abs	4378
UK-Abs	14,296

Solution: Chunking with Fine-Tuning over Legal Data



Challenges:

- How to get chunk-summary pairs for fine-tuning such models?
- Loss of information across the independent chunks

Extractive vs. Abstractive

- Extractive models trained over legal data
- Pre-trained abstractive models fine-tuned over legal data

Results on the IN-Abs dataset

Method	Type	ROUGE-1 F	ROUGE-2 F	ROUGE-L F
Gist	Extractive	0.471	0.238	0.308
SummaRuNNer	Extractive	0.493	0.255	0.274
BART	Abstractive	**0.495**	**0.249**	**0.330**
Legal-Pegasus	Abstractive	**0.488**	**0.252**	**0.341**

Extractive vs. Abstractive

- Extractive models trained over legal data
- Pre-trained abstractive models fine-tuned over legal data

Results on the IN-Abs dataset

Method	Type	ROUGE-1 F	ROUGE-2 F	ROUGE-L F
Gist	Extractive	0.471	0.238	0.308
SummaRuN Ner	Extractive	0.493	0.255	0.274
BART	Abstractive	**0.495**	**0.249**	**0.330**
Legal-Pegasus	Abstractive	**0.488**	**0.252**	**0.341**

- We also conducted evaluation of summaries by Law experts
- Law experts often prefer summaries by simple extractive models
- Lot of scope for improvement of legal summarization models

Resources

1. Bhattacharya et.al., A comparative study of summarization algorithms applied to legal case judgments, ECIR 2019

2. Bhattacharya et.al., Incorporating Domain Knowledge for Extractive Summarization of Legal Case Documents, ICAIL 2021 (Best Student Paper Award)

3. Shukla et.al., Legal case document summarization: Extractive and abstractive methods and their evaluation, AACL-IJCNLP 2022

Data & Codes : https://github.com/Law-AI/summarization

Similarity between Legal Case Documents

Motivation for Legal Case Doc Similarity

- In a particular case document, not all relevant prior-cases are cited -- only a few are
- Recommendation and Search for similar *uncited* case documents
- Analogy : Search similar patents & scientific articles

Estimate the similarity of two legal case documents

Case Document (https://indiankanoon.org/doc/749642/)

Kelvinator Of India Ltd vs The State Of Haryana on 23 August, 1973

Equivalent citations: 1973 AIR 2526, 1974 SCR (1) 463

Author: H R Khanna

Bench: Khanna, Hans Raj

CIVIL APPELLATE JURISDICTION : Civil Appeal No. 2005 (NT) of 1972:.

Appeal by special leave from the order dated the 11th April 1972 of the Punjab and Haryana High Court at Chandigarh, in General Sales Tax Reference No. 8 of 1970.

N.A. Palkhivala, H. L. Sibal, J. B. Dadachanji, A. K. Verma, Kapil Sibal and S. C. Agnihotri, for the appellant. Y.M. Tarkunde, Narendra Goswami and M. N. Shroff, for the respondent.

S. T. Desai, and 1. N. Shroff, for the intervener.

The Judgment of the Court was delivered by KHANNA, J. This appeal by special leave by M/s. Kelvinator of India Ltd. is directed against the judgment of Punjab & Haryana High Court whereby that court answered the following question referred to it by the Sales Tax Tribunal Haryana in favour of the department and against the appellant "Whether on the facts and circumstances of the case, the agreement between M/s. Kelvinator of India (Assessee) M/s. Spencer & Co. Ltd., Messrs Blue Star Engineering Co,, and M/s. General Equipment Ltd., in pursuance of which the refrigerators manufactured by M/s, Kelvinator of India at Faridabad moved to Delhi were merely for distribution of goods between the principal and his agents or were agreements of sale between two parties?"

The matter relates to the assessment year 1965-66, i.e. the period from April 1, 1965 to March 31, 1966. The appellant company has a, factory at Faridabad in Haryana. It manufactures refrigerators, deep freezers, compressors and other similar articles. The factory went into production in 1964.

....

We are concerned in the present case with clause (a) ....The word "occasions" is used as a verb and means 'to cause or to be the immediate cause of'. In the case of Tata Iron and Steel Co. Ltd. v. S. R. Sarkar and Ors; (1) Shah J. (as he then was) speaking for the majority observed that a transaction of sale is subject to tax under the Act on the completion of the sale. A mere contract of sale is not a sale within the definition of "sale" in section 2(g). A sale being, by the definition, transfer of property becomes taxable under section 3(a) "if the movement of goods from one State to another is under a covenant or incident of the contract of sale". In Ben Gorm Nilgiri Plantations Co. Cooncor & Ors. v. Sales Tax Officer, Special Circle, Ernakulam & Ors(2) this Court dealt with the provisions of

.....

We accordingly accept the appeal and set aside the judgment of the High Court ..... The appellant shall be entitled to the costs from the respondent of this Court as well as in the High Court. V.P.S.
Appeal allowed.

Precedent Citations

Other similar cases :

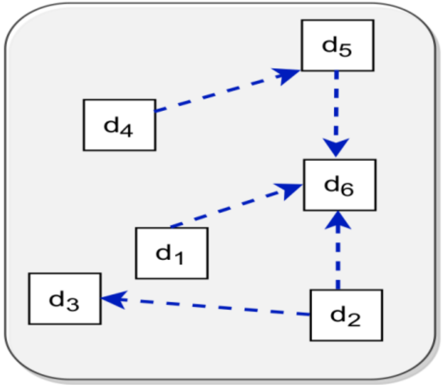
- M/s Sahney Steel and Pressworks Ltd v The Commercial Tax Officer And Others
- M/S Hyderabad Engineering Industries v State Of Andhra Pradesh
- Union of India and Another v K. G. Khosla and Company Ltd and Others

Legal Doc Similarity : Network-based

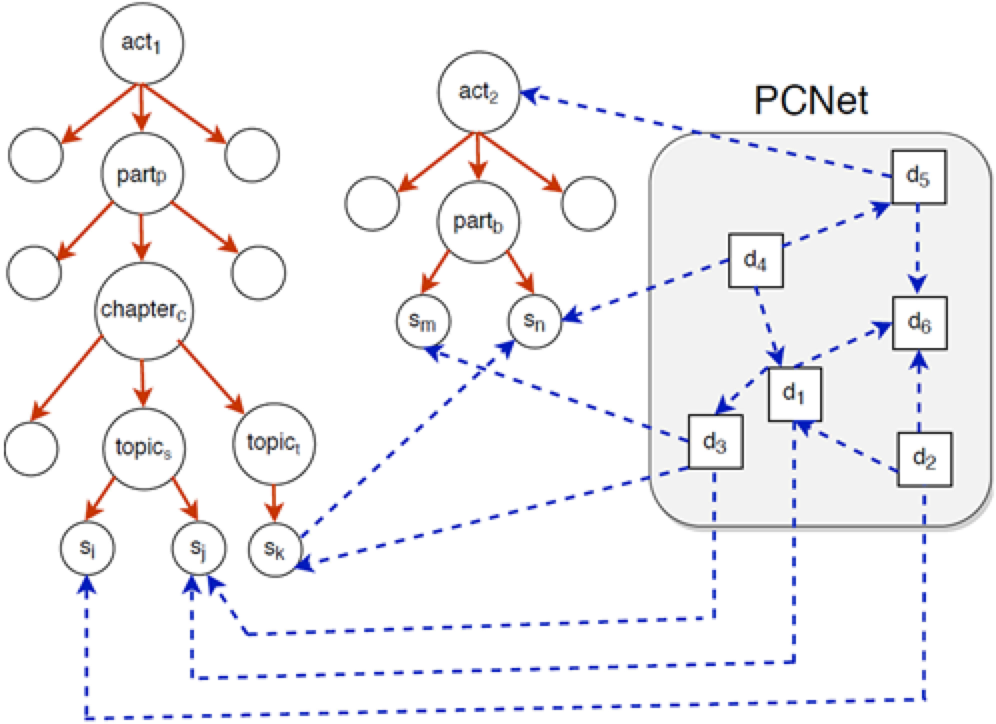
Network-based :

- Used **precedent citation networks (PCNet)**
  - Vertices : case documents
  - Edges : if a case document has cited another case document
- Misses important signals inherent in Statutes
  - Rich source of legal domain knowledge

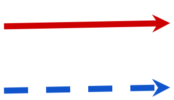
PCNet



Hier-SPCNet (Hierarchical Statute + Precedent Citation Network)



- Augment PCNet, with a network of Statutes (Hier-SPCNet) encompassing:
  - The structure of the Statutes
  - Citation information within the Statutes
- Heterogeneous Network



Hierarchy links

Citation links

Similarity measures – Bibliographic Coupling, Co-citation, Node2Vec Metapath2Vec (encodes domain knowledge)

Combining Network & Text for Legal Case Doc Similarity

- Legal case documents have a rich network which provide useful domain-knowledge
- The textual content of the case documents --> Unsupervised similarity measures : TF-IDF, Doc2Vec
- Combine information from both the sources

Combining Network & Text for Legal Case Doc Similarity

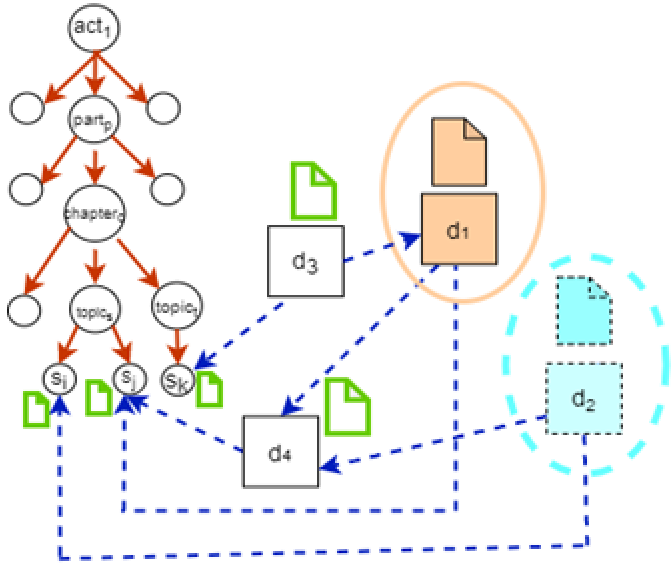
1. **Value Combination (n : network sim., t : text sim.)**
   - avg = average (n,t)
   - max = maximum(n,t)

2. **Embedding combination (n: network emb., t: text emb.)**
   - avg = element-wise-avg(n,t)
   - max = element-wise-max(n,t)
   - concat = concatenation(n,t)

Combining Network & Text for Legal Case Doc Similarity

1. Value combination (n : network sim., t : text sim.)
2. Embedding combination (n: network emb., t: text emb.)

3. **Graph-based Combination**



- Give the graph as input to node embedding algorithms : TADW, GCN, GraphSage

- These representations are expected to fuse :
  - Text : given as node feature
  - Network structure from this input graph itself

Combining Network & Text for Legal Case Doc Similarity

1. Value combination (n : network sim., t : text sim.)
2. Embedding combination (n: network emb., t: text emb.)
3. Graph-based Combination

4. **Neural Network based :** NN-Map+Conc[1], Autoencoder[2]
    - Learn a shared embedding space by learning to reconstruct the text emb. from the network emb. (or vice-versa)
    - Multi-modal techniques to combine text & image -- have not been explored to combine text & network

[1] Collell et.al., Imagined Visual Representations as Multimodal Embeddings, AAAI 2017
[2] Wang et.al. Associative Multichannel Autoencoder for Multimodal Word Representation, EMNLP 2018

Combining Network & Text : Results

Type	Function	Method	Correlation	MSE	Fscore
Text based	–	Doc2Vec	0.777	0.0381	0.774
Network based	–	Hier-SPCNet-ICF-m2v	0.725	0.0427	0.779
Text + Network Combination	Value combination	Value-Average	0.781	0.0390	0.828
		Value-Max	0.761	0.0375	0.788
	Unsupervised Embedding Combination	Emb-Average	0.749	0.0382	0.803
		Emb-Max	0.745	0.0612	0.705
		Emb-Conc	0.781	0.0386	**0.829**
	Graph-based combination	Paper2Vec	0.651	0.0598	0.745
		TADW	0.706	0.0619	0.705
		GCN	0.353	0.2483	0.385
		GraphSage	0.597	0.1611	0.678
	Neural Network based combination	NN-Map+Conc	**0.836**	**0.0332**	**0.829**
		NN-Map+Wtd.Conc	0.782	0.0424	0.715
		AutoEncoder	0.729	0.2558	0.328

- Best performance is noted in NN-Map+Conc
- Emb-Conc is competing
- Graph methods does not perform well
  - Domain-knowledge, e.g., through metapaths
  - Relational GCNs

Resources

1. Bhattacharya et.al., Hier-SPCNet A legal statute hierarchy-based heterogeneous network for computing legal case document similarity, SIGIR 2020

2. Bhattacharya et.al., Legal case document similarity: You need both network and text, IPM 2022

Data: https://github.com/Law-AI/document-similarity

Backup

Challenges in Legal Case Document Summarization

- Very few datasets available for training legal summarization models
- Expensive to get reference summaries written by Law experts
- Documents are lengthy (thousands of words) → cannot be easily handled by most abstractive summarization models

Challenges in Legal Doc Sim

- Case documents are long, complicated and unstructured
- The notion of Legal Case document similarity is not well-defined
  - 2 docs are similar if the experts judge them to be similar
- Lack of annotated data
  - Expensive to obtain expert annotations
  - Unsuitable for supervised methods (eg. DL models)