Artificial Intelligence as Legal Research Assistant

Jhanvi Arora[a], Tanay Patankar[a] , Alay Shah[a] and Shubham Joshi[a]

[a] Lawnics Technologies, F-4 Raghushree Building, Jaisingh Highway, Bani Park, Jaipur, India

Abstract

Application of text retrieval and semantic segmentation has a lot of potential in changing the landscape of the legal research industry by making relevant information more accessible and affordable to anyone. In this working paper, we present a description of a few novel methods as a part of Artificial Intelligence for Legal Assistance (2020), an integral event of Forum for Information Retrieval Evaluation-2020. In the first part of the paper, we have identified the relevant prior cases and statutes for the provided query using approaches based on BM 25, Topic embeddings and Law2Vec embeddings. For the second part, we used BERT to semantically segment a legal case document into Seven pre-defined labels or "rhetorical roles". In the first task, our performance in P@10 and BPREF metrics positioned us in the top 2 ranking spots. On the other hand, our BERT implementation for the second task got us macro precision of .479, which is just .027 lower than the best performing approach.

Keywords [1]

nlp, word embeddings, topic embeddings, bm25, precedent retrieval, information retrieval, statute retrieval, bert, rhetorical role, classification, law

1. Introduction

Statutes, which are written law, and precedents, which are previous judgements delivered by a court are the essential sources of information used by lawyers for preparing their case for their trial by understanding the method by which the court dealt with comparative situations in the past. This is a largely a manual and tedious process, involving hours of scanning through a large number of cases to find the relevant cases. With a fast increment in the trend of digitizing legal documents, the development of a system which could help to shorten this search by suggesting the relevant statutes and judgements or by classifying them into relevant rhetorical roles would be of great significance. We present our approach to this problem by using Natural Language Processing and Information Retrieval techniques.

2. Problem Definition

Artificial Intelligence for Legal Assistance (AILA) was one of the tracks available for the Forum for Information Retrieval (FIRE) 2020[1]. The track had two tasks in which the first task was further bifurcated in two subtasks.

Similar to AILA 2019 [2], task 1 was segregated into two sub parts namely- Relevant Precedent Retrieval and the Relevant Statute Retrieval for the given query judgement. The Precedent documents pool to be considered for the relevance consisted of ~3260 case documents, while there were 197 statutes to be mapped to the given queries. The training data consisted of 50 queries mapped to the relevant case documents and the statutes, while the test set consisted of unseen 10 queries that ought to be mapped to the same. The submissions were to be evaluated by trec_eval toolkit format on the basis of four standard metrics- Mean Average Precision, BPREF, recip_rank and P@10 score.

[1] Forum for Information Retrieval Evaluation 2020, 16-20 December 2020, Hyderabad, India.
EMAIL: jhanvi@lawnics.com (A. 1); tanay@lawnics.com (A. 2); alay@lawnics.com (A. 3); sshubham@lawnics.com (A.4)

In task 2, 50 legal case documents were provided, with each sentence being classified into one of the 7 rhetorical roles, namely Facts, Ruling by Lower Court, Argument, Statute, Precedent, Ratio of the decision and Ruling by Present Court, bringing it to a grand total of 9380 sentences. A test set of 10 legal case documents following the same order were provided. The submissions were evaluated on the basis of the Macro Precision score, Recall score, F-Score and the Accuracy.

3. Related Work

In context of the work presented in [3] of Zhao et al. wherein the ensemble model of the BM25 establishes good proven efficiency in terms of the metrics stated in AILA 2019 [2], we derive the base idea of our approach from the same intuition. BM25 as an approach is based on the word-based characteristics of the document rather than the contextual mapping of the documents. Therefore, alongside preserving the word-based features of the documents, our method aims to inculcate the inner theme and content meaning-based features in the task 1. Also [4], proposing a similar approach, wherein BM25 was used to compute the similarity between the case documents and queries and BERT [5] being the component to map the relevant contextual information promises good results.

With respect to the second task, the work presented in [6] used Bi-LSTM network and a CRF network on top of the Bi-LSTM network to classify legal case documents into rhetorical roles. In the context of the work done on DocBERT [7], wherein pre-trained BERT models were fine-tuned for document classification on four datasets, the base idea for our implementation was derived.

4. Relevant Precedent and Statute Retrieval

The work for each sub-task of Task 1 was submitted into two runs wherein the run 1 of the relevant document extraction was the same for each of the sub-tasks. Run 2 for sub-task differs and are applied on the basis of length and content of the relevant documents pool. The pre-processing strategy applied to the query, precedents (case documents) and statutes is constant throughout the Task 1.

4.1. Pre-Processing of Documents

The following steps entail the pre-processing of all the case documents and statutes to form a viable corpus on which the methodologies discussed further have been implemented. Also, the final queries posed for searching relevant documents are pre-processed in a similar manner. The language model used here for pre-processing is derived from spaCy [8].

1. Cleaning of Text: Each case document and query were parsed for the removal of certain noise inducing alphanumeric sequences, numbers and abbreviations present in the text that would otherwise not represent meaningful information.

2. Removing Entities: Inferring from the results of [9], all the named entities present in the text namely of the type- person, law, date, place, state, country and reference to any nationality were identified and removed from all the case documents, statutes and queries present in the dataset.

3. Omitting the Stop words and Lemmatization: To emphasize and give significance to contextual words, stop words were filtered out, while the words retained in the text were reduced to their root form using Pattern Lemmatize. To reason lemmatization over stemming, the objective was to even out text by assigning words with similar context a common word.

4.2. Precedent Retrieval:

In run 1, after we processed the documents, using the Named Entity Recognition we preserved important information containing content in the form of Noun and Adjectives. The underlying assumption of this cleaning is that most of the context is retained in the form of the above tags mentioned. Later to map the similarity between query and documents a word-based feature method Okapi BM25[2] is used. BM25 is a bag-of-words retrieval function that ranks a set of documents based on the query terms appearing in each document, regardless of the inter-relationship between the query terms within a document. It returns the documents scores in descending order and each score is normalised by dividing each score with the maximum score recorded.

In run 2, for LAWNICS_2, using the intuition highlighted by Angelov et al. [9], we aim to classify the context of the query and case documents and map them on the basis of topic classification. The underlying assumption of the method is that one case document/query contains several topics (can be referred to as context) in itself with each topic being of variable intensity. The aim was to extract the most highlighted topics by the case document and query.

The pre-processed case documents were treated as the corpus for learning the various varied topics present in the dataset. The Topic Embeddings learned by Top2Vec method [9], represent each topic by a set of 50 words that clustered closest to form that topic. The setup learned 38 topics present in the dataset.

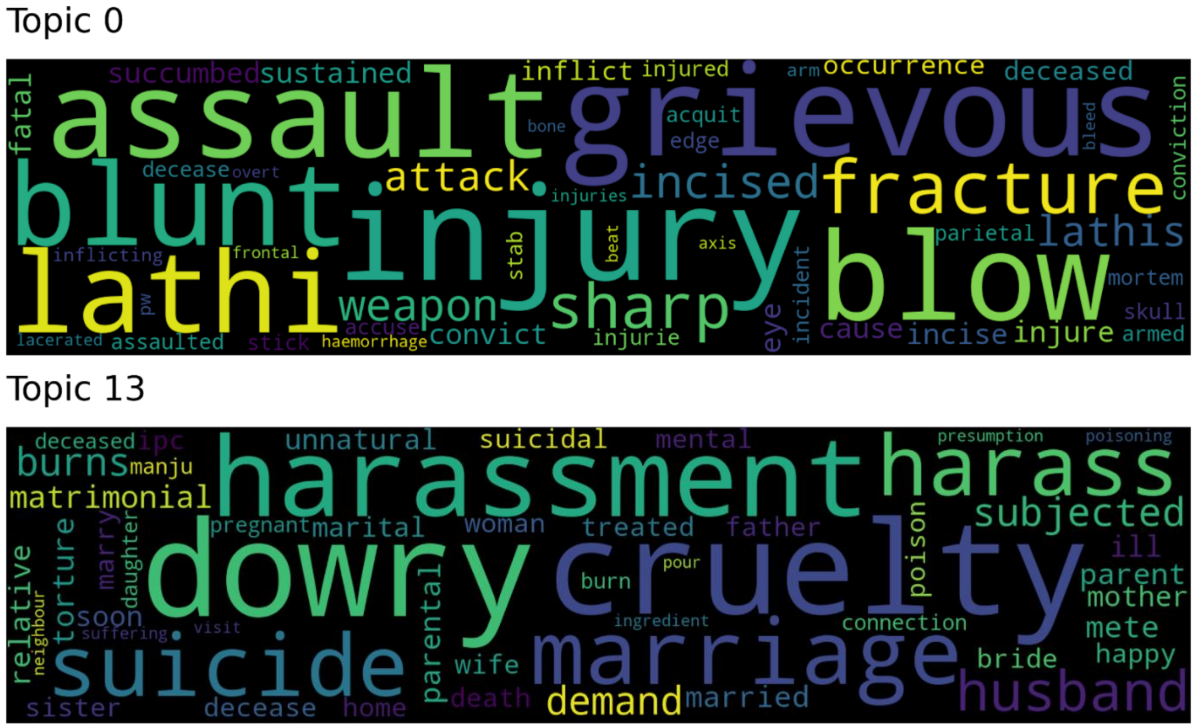

Figure 1: Instance of 2 topic-word clouds amongst the 38 topics learnt

The topic word-cloud of topic 0 and topic 13 from Figure 1 pretty much give a gist of context of each topic. Topic 13 hints at a dowry related context, while topic 0 relates to an assault related context. Both the topics can be jointly present in varying intensity in one case document.

To represent each query, we extract all the words that represent any topic amongst the 38 topics previously learnt on the corpus. We derive the semantic similarity of each query with case-docs by cosine-similarity.

Preserving the obtained scores, we also derive the Okapi bm25 relevance scores for each query with the given case documents. Normalizing and weighing both the scores equally, the final similarity score for each query with the case documents is generated. The score is scaled from 0 to 1, with 0 indicating lowest extent of relevance and 1 highest.

4.3. Statute Retrieval:

For run 1 of task 1b, the same setup as the Precedent Retrieval LAWNICS_1 is followed. We retain only the Noun and Adjectives in query and statutes and rank it in the order of similarity on the basis of Okapi bm25 method.

In the run 2 of task 1b, LAWNICS_2, we apply the pre-trained embedding model, Law2Vec [10] to obtain considerable mapping between the given queries and statutes. However, instead of traditional distance measure of cosine similarity, the text from both comparison ends was converted into a bag of words representation and the metric of soft cosine similarity deriving implementation from [11] was applied. The soft cosine similarity takes into consideration the individual word distances in contrast to cosine similarity which maps the distance between two documents as whole.

4.4. Results:

Table 1
Precedent Retrieval Results

Run	Method	MAP	BPREF	recip_rank	P@10
LAWNICS_1	BM25(Noun & Adjectives)	0.1085	0.0756	0.1607	0.08
LAWNICS_2	Topic Embeddings+ BM25	0.1288	0.0913	0.1586	0.1

The experimental method of weighing BM25 scores and Topic embedding submitted in second run achieves a P@10 score of 0.1 which stands the highest with all other parameters of MAP, BPREF and recip_rank ranking amongst the top 10 submissions for Task.

Table 2
Statute Retrieval Results

Run	Method	MAP	BPREF	recip_rank	P@10
LAWNICS_1	BM25(Noun & Adjectives)	0.2962	0.2812	0.4607	0.13
LAWNICS_2	Law2Vec Embeddings+ Soft Cosine Similarity	0.1996	0.151	0.4843	0.09

The proposed method of retaining only noun and adjectives in the text with BM25 submitted as the first run of the task achieves better results, while amongst all the methods submitted it achieves a BPREF score of 0.2812 which stands second with all the other parameters of MAP,P@10 and recip_rank amongst the top 10 of all the submissions.

5. Rhetorical Labelling:

In this task, we had to classify a given legal case document into one of the seven rhetorical roles mentioned before.

5.1. Pre-Processing of Documents:

Each legal case document was parsed for the removal of alphanumeric sequences, abbreviations and ASCII characters present in the text that would not help in providing any meaningful inference.

5.2. Methodology:

As defined in [5], "BERT is designed to pretrain deep bidirectional representations from unlabelled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be finetuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as language inference or sentence classification."

While going through the training dataset, we realized a high imbalance in the count of each class. The count ranged from the highest being 3624 to the lowest being 262. To ensure that the minority class was not left out while training, we split each class in an 80:20 and aggregated them into the training and testing sets respectively.

For the first run we finetuned BERT$_{\text{base}}$ model with a fully connected hidden layer of size 768x64 and an output layer to generate the predictions. For the second run, we followed a similar approach to the first run but we removed the hidden layer, keeping the output layer. The model was trained for 4 epochs with a learning rate of 3e-6 and a token size of 512. A small batch size of 8 was used due to hardware limitations. The result of the second run was higher with results in Table 3. Overfitting could be a possible reason for the low accuracy of the first run. With the second run, we secured a position in the top 10 of the submissions.

Table 3
Submission Results

Run	Accuracy	Macro Precision	Macro Recall	Macro F-Score
lawnics_1	0.152	0.208	0.164	0.119
lawnics_2	0.584	0.479	0.479	0.435

6. Conclusion and Future Work:

While extracting the precedents, the method of weighing different thematic contexts using Top2Vec [9] with the word-based features from Okapi BM25 achieves higher result in terms of BPREF, MAP and P@10 score. In the proposed work the achieved efficiency is based on the topic-based context extraction, better context disambiguation could promise a higher efficiency. Alongside, better content filtering to overcome the challenges posed by the long text- such as good abstraction techniques can indeed lead to better results in the legal domain.

With respect to the rhetorical labelling of case documents, further hyperparameter tuning could promise a greater accuracy. Furthermore, creating an ensemble of the work of P. Bhattacharya [6] could help in increasing the precision of accurately labelled rhetorical roles in documents falling within the legal domain.

7. References

[1] P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, P. Mehta, A. Bhattacharya, P. Majumder, Overview of the Fire 2020 AILA track: Artificial Intelligence for Legal Assistance. In Proc. of FIRE 2020 - Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020.

[2] P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, P. Mehta, A. Bhattacharya, P. Majumder, Overview of the Fire 2019 AILA track: Artificial Intelligence for Legal Assistance. In Proc. of FIRE 2019 - Forum for Information Retrieval Evaluation, Kolkata, India, December 12-15, 2019.

[3] Zhao, Zicheng, et al. "FIRE2019@ AILA: Legal Information Retrieval Using Improved BM25." FIRE (Working Notes). 2019.

[4] Gain, Baban, et al. "IITP in COLIEE@ ICAIL 2019: Legal Information Retrieval using BM25 and BERT." (2019).

[5] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pretraining of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[6] Paheli Bhattacharya, Shounak Paul, Kripabandhu Ghosh, Saptarshi Ghosh, Adam Wyner, “Identification of Rhetorical Roles of Sentences in Indian Legal Judgments,” in Proc. Jurix, 2019.

[7] Adhikari, A. Ram, R. Tang, and J. Lin, “Docbert: Bert for document classification,” arXiv preprint arXiv:1904.08398, 2019.

[8] Honnibal, Matthew, and Ines Montani. "spaCy library." (2018).

[9] More, Ravina, et al. "Removing Named Entities to Find Precedent Legal Cases." FIRE (Working Notes). 2019.

[10] Chalkidis, Ilias, and Dimitrios Kampas. "Deep learning in law: early adaptation and legal word embeddings trained on large corpora." Artificial Intelligence and Law 27.2 (2019): 171-198.

[11] Řehůřek, Radim, and Petr Sojka. "Gensim—statistical semantics in python." Retrieved from genism. org (2011).