Explainable Statute Prediction from Legal Documents: Observations and Challenges

Anonymous Author(s)

ABSTRACT

Given a case situation, explainable statute prediction is an important and challenging problem. While there has been work on statute prediction, there has been limited work focusing on the explainability aspect of the same. To our knowledge, all of the state of the art approaches try to highlight explanations as an artefact of the underlying model. Recent studies show that such explanations do not align well with the legal experts' notion of explanation and hence such system-generated explanations are not likely to satisfy the legal experts – the end users of the prediction system. In this paper, we perform a preliminary study to understand the complexity of the statute prediction task and the difficulty in incorporating human explanations into the framework. The task turns out to be more complicated for the legal domain. In this paper, we report initial results and motivate to solve the task by considering explanations provided by the legal experts as an integral part of the model.

KEYWORDS

Statute retrieval, Explanation, Retrieval

ACM Reference Format:

Anonymous Author(s). 2018. Explainable Statute Prediction from Legal Documents: Observations and Challenges. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Countries on the higher side of population spectra such as India face the issue of a huge pendency in legal cases across all level of the judiciary. The problem, especially in India, is due to several factors, primary being inadequate strength of judges. Apart from that, limited working hours of courts during the COVID pandemic has increased the backlog. Side by side, client to expert ratio is significantly high. In a developing country like India, appointing more manpower to clear the backlog seems a bit far fetched. Thus the role of technology in assisting the legal machinery to speed up the process seems to be one of the most promising ways to deal with this issue. In general, lawyers and judges have to manually go through the document and identify which statutes fit best to the document and this is a time-consuming and painstaking task. In this paper, our primary objective is to automate this task to some extent. However, another fold of this problem is trustworthiness. In a sensitive domain like law, it is very unlikely that a legal expert will use the prediction directly that does not include any clues or reasoning process along with the prediction. It becomes a prime requirement to provide some sort of explanation/supportive documents that can justify the prediction and it will also help the lawyers in understanding the case and arguing in the court.

In this paper, we propose a setting to jointly address both the tasks i.e., predicting the statutes and the sentences in the document that might be a possible reason for such predictions. For this task we consider 50 fact descriptions of legal documents collected from FIRE legal track [1]The most frequent 200 sections were selected from the Indian Supreme Court case documents between 1952 to 2018. Three of these laws are repealed, thus leaving us with a pool of 197 statutes. Then the documents citing these statutes were collected and 50 documents chosen at random, from which the facts are extracted. Legal experts manually went through these descriptions and annotated the corresponding sentences that are relevant for each of those statutes. These annotated sentences served as the expert explanations for these statutes. The task that we address in the paper is to identify these statutes and the sentences relevant these statutes from the input documents.

First, we check the distribution of the statutes over the documents. We observed that appearance of some statutes are much frequent than others and overall distribution follows a power law. Interestingly, explanations are also not disjoint i.e., same sentence sometimes appear as explanation for two closely related statutes (e.g., *s*2 and *s*20) and make the task even more complicated. Hence, simple word based detection of statutes may lead to many false positives.

Further, we do a posthoc analysis to identify the sentences that might be a reason for the statutes. For explanability, we use sentences as explanations, since just words or small phrases might not capture the context satisfactorily. For the purpose of quantifying explanability, we use a method that evaluated the contribution of each sentence toward the similarity score between fact and statute description. This method allows to assess that the presence of which sentence contributes to the relevance of a statute and to what magnitude. Thus for every fact-statute pair we can have the sentences from the fact description that justify the relevance of the said statute.

2 RELATED WORK

The statute prediction task aims at finding relevant statute(s) for a given fact, and the reasoning behind the selection is justified as part of the explainability task. In several works the task of statute prediction is generalised as charge prediction, where rather than predicting the exact section (e.g. Indian Penal Code 320 , Constitution 14) of the law that deals with the case, the broad domain (e.g murder, intentional injury) to which the case belongs is retrieved.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

Previous works have mainly considered the former task, with little exploration done to explain the predictions. Paul et al. (2020) [10] use an attention-based model to predict relevant charges given the facts of a situation/case. They annotate a small fraction of their dataset with sentence-level charge annotations, and combine the primary task of document-level charge prediction with an auxillary task of sentence-level charge prediction. The document-level charge prediction is done by assigning weights to sentences as per sentence-level predictions and then obtaining document representation by aggregating sentence representations. Chao et al. (2019) [2] proposes a neural framework to extract important text snippets from the fact description and utilise them to improve charge prediction performance. They have integrated interpretability together with the prediction task. Unlike most other works which focus on prediction and then getting interpretation for the same, they extract rationales from input fact description and then predict the charge distribution. For rationale extraction they utilise reinforcement learning. This work closely resembles the direction of our work but differs on the implementation level. Rather than using rationales for prediction and improving prediction accuracy, we provide explanations for every prediction, justifying their suitability. We also quantify explanability rather than just using it as a means to the end, thus maintaining the possibility of comparing the performance with another explainable statute prediction system.

In contrast to the above two works, our work handles the more intricate statute prediction task. Two statutes can fall under the same charge thus making it more harder to track and capture the relevant statute as compared to the relevant charge.

Luo et al. (2017) [6] considers the charge prediction and statute prediction simultaneously. They propose an attention-based neural network method to jointly model both prediction tasks.

3 DATASET

In this section, first we briefly introduce our dataset and then elaborate the annotation process.

3.1 Dataset Details

The dataset is obtained from the Supreme Court case documents between 1952 to Feb, 2018. We take 50 documents from the FIRE AILA 2019 task [1]. From the legal cases, we take out fact sections and legal experts identify statutes and explanations i.e., sentences relevant for those statutes. We only consider the fact section because for future/new cases only fact part will be available to the system. We briefly describe our dataset and annotation strategy in the next part.

Statutes: Each Act (e.g. Indian Penal Code 1860) in law contains articles (e.g. Article 309 of the Constitution of India 1950) or sections (e.g. Section 120 of the Indian Penal Code 1860), which are referred to as statutes. A collection of 197 statutes relevant to the facts is utilized. Each statute is accompanied by a title and description, and denoted using identifiers S1 to S197. We provide an example statute from our dataset below.

Title: *Appellate jurisdiction of Supreme Court in appeals from High Courts in regard to civil matters* ***Desc:*** *1 (1) An appeal shall lie to the Supreme Court from any judgment, decree or final order in a civil proceeding of a High Court in the territory of India 2 if the High Court certifies under Article 134A- (a) that the case involves a substantial question of law of general importance; and (b) that in the opinion of the High Court the said question needs to be decided by the Supreme Court.·*

Fact Description: As mentioned before, we consider fact description part from the 50 chosen documents. Each 'fact' describes the chronology which resulted in filing of the case in the Supreme Court. Following preprocessing stages are performed to remove bias from the dataset and the model training phases:

1. The mention of statutes was removed.

2. Client/involved party names, dates, places were also anonymized to ensure that model should not focus on case specific individuals. For example we use 'Person' in place of names.

We showcase an example of a fact description of a legal case that deals with adulteration of food product.

This appeal arises from the judgment of the learned Single Judge of High Court dated 6th June, 1988 whereby the learned Single Judge declined to quash the prosecution of the petitioner. The petitioner therein has been prosecuted for selling adulterated supari on the basis of a certificate issued by the Director of Central Food Laboratory showing that the article of Food purchased from the accused contained 2000 mgs/kg.·

Ground Truth Labels: For the statute prediction task, we have the ground truth relevance labels for each query. For every fact description, the relevant mentioned statutes are extracted from the Supreme court document from which the facts of the cases were extracted. Thus for every fact we have the relevant ground truth labels from the pool of 197 statutes.

3.2 Data Annotation Process

The facts descriptions are annotated manually for explanations pertaining to each relevant statute. The annotations are done by a team of 5 legal experts. There is no standard measure for evaluating annotator agreements over textual data, thus metrics like ROGUE-1, ROGUE-2, ROGUE-L [5] are utilized.

A qualitative analysis of annotator agreement was also carried out which showcases the intricacies, subjectivity and comprehension challenges of the legal data. These characteristics of the legal data are not only computationally challenging but often even confuse the experts.

To consider an example, the expert annotations for the query AILAQ18 are highlighted in yellow.

These appeals involve a pure question of law as to whether an award by which residue assets of a partnership firm are distributed amongst the partners on dissolution of the partnership firm requires registration. Briefly the facts are that a partnership firm was constituted comprising of four persons belonging to the same family. Disputes and differences arose between the partners which were ultimately referred to arbitration. The arbitrators made an award on 2nd October, 1972. The award was challenged by way of objections filed under by some of the partners. The objection petition was contested by the other partners who prayed that the award be made a rule of the Court. The grounds of challenge to the award included misconduct on the part of the arbitrators as well as another ground that the award required registration. The trial Court accepted both the objections holding that there was misconduct on the part of the arbitrators as also that the award was required to be compulsorily registered and since it was not registered it was inadmissible in evidence. This decision of the trial court was challenged before the High Court by way of a Civil Revision filed. The High Court found that in the facts and circumstances of the case it could not be said that there was any legal misconduct on the part of the arbitrators. Thus the first ground of attack against the award was found to be unsustainable. However, the High Court accepted the finding of the trial Court on the second ground, that is, the award was required to be compulsorily registered. Since the award was unregistered, it could not be made a rule of the Court. Hence the present appeals.

Figure 1: Annotations by legal expert

For providing the explanations to the the statute prediction system (Task 2), we utilise the available expert annotations to evaluate the task.

4 METHOD

Considering statute prediction as a ranked retrieval task, we aim to calculate the similarity between fact-statute pair and rank them. We consider facts as query and compute the similarity between query and statute documents. We have 197 statutes in the corpus. Hence, we compute similarity score with each of the statutes and rank them in decreasing order. Finally, we retrieve $top - k$ statutes for a query q where k is the number of human-annotated relevant statutes for query q. Note that, the number of relevant statutes is different for different queries and our k is also adjusted based on that. To measure the similarity between query (fact) and the statutes, we consider both term-based, latent vector based, and contextual approaches. We consider the following approaches.

1. BM25: We use the general Okapi BM25 [13] model formulation, with the default parameter values (k_1 = 1.5 and b = 0.75). For every query, the score from BM25 model is computed with each of the 197 statutes, providing us with a ranked list after arranging the scores in descending order.
2. Word2vec: We use word2vec [9] to get word-embedding for every word in the query and document. Thereafter to obtain document-level representations we aggregate the word-embedding for the words in the document. We then use the cosine similarity score as a measure of similarity amongst query and statutes. We trained word2vec model on Indian legal data before getting the embeddings so that it can capture the context better. We use the embedding dimension 100.
3. Doc2vec: The doc2vec [4] is utilized as another approach to obtain document-embedding in the first step itself. After getting the document-level vector representations we follow the same steps and similarity metric as before. For implementation of word2vec and doc2vec, the Gensim Library [11] is used and we train them on our legal corpus.
4. BERT: Finally, we use contextual model BERT to generate the document embedding. We use a fine-tuned version of plain BERT model [3] that is trained on question-answering corpus [14]. We explored several trained models (RoBerta, DistillBert, mpnet) but fine-tuned BERT model [14] based document embeddings perform better than other approaches on our dataset. It maps the documents (both queries and statutes) to a 384 dimensional dense vector space. The model imposes a 512 token limit to calculate embedding; hence, tokens beyond 512 limit get truncated. Finally, we rank them using the same approach proposed before.

5 EXPERIMENT RESULTS

In this section, we first check the distribution of statutes over the corpus and report the results of our similarity based ranking strategies. Next, we report the results of posthoc explanation methods. Finally, we provide the limitations, domain specific challenges and open problems.

Statute	Freq	Frequent words
S2	13	deceased, accused, appellant(s),...
S197	6	persons, entered, armed, pistols,...
S20	5	sell, agreement, wrongful, transaction,...
S43	5	deceased, accused, rushed, towards,...
S6	5	officers, naval, blow, against,...
S115	4	accused, statement, recovery,...
S3	4	writ, petition, respondent, treated,...
S9	4	petition, allowed, sanctioned, judgment,...
S10	3	respondent, right, service, specific,...
S153	3	statement, accused, chain, allegedly,...

Table 1: 10 most frequent statutes

Model	MAP value
BM25	0.075
word2vec	0.0479
doc2vec	0.0413
BERT	0.1243

Table 2: MAP(Mean Average Precision) score over some baseline models

5.1 Statute Distribution

In this part, we want to measure the distribution of statutes across the fact descriptions. Some of the statutes are quite common and frequent whereas others occur only once. Side by side, we also check the distribution of words present in the marked explanation sentences to verify whether some words represent some specific statutes with high probability. Table 1 presents the distribution pattern of the most frequent statutes and their associated explanation/rationale words.

We also compute the similarity between two statutes based on the overlap of words between their explanation sentences. We have used Simpson similarity [7] to measure the closeness between two statutes.

$$Simpson - similarity(X, Y) = \frac{|X \cap Y|}{min(|X|, |Y|)}$$

In which X, Y are the set of explanatory words/sentences of the two corresponding statutes. We build a graph where each of the statutes represent a vertex and edge weight between them is decided based on the similarity coefficient of the two explanation sets. Further, we use Spectral clustering algorithm with PCA (Principal Component Analysis) to yield four clusters. The Figure 2 depicts four clusters corresponding to four corresponding colors. Red and blue clusters are concentrated to some extent.

We calculate the MAP (Mean Average Precision) score as an evaluation metric for our models. We observe that BERT gives the highest MAP score as compared to other three models.

5.2 Model Performance

In this part, we evaluate the performance of different ranking models proposed in Section 4. As it is a ranking task, we use the mean average precision metric (MAP) to compare their performances. MAP is an evaluation standard that provides across recall levels. For a single query, average precision is the average of precision value over the top-k ranked documents, after each relevant document is retrieved. The mean of this average precision value over all the

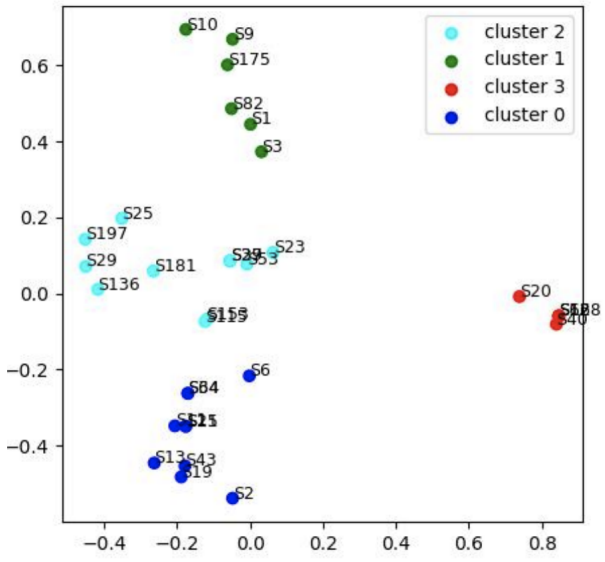

Figure 2: Four clusters by Spectral clustering algorithm

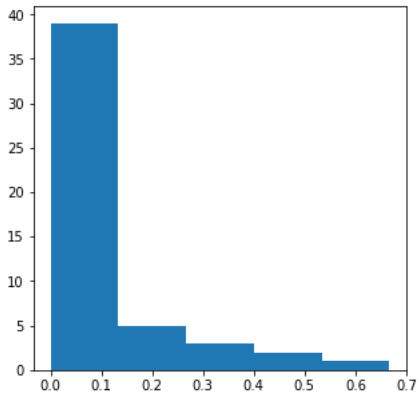

Figure 3: Histogram for Precision scores using BERT

queries/information needs gives the Mean Average Precision score. MAP is a standard evaluation measure and offers good stability and discrimination.

We showcase a histogram (Figure:3) that displays the precision scores using the BERT model, over all 50 queries.We observe that running the system on several queries does not retrieve any relevant statute in the top-k documents. This demonstrates the difficulty of the task.

From Table 2, it is evident that simple similarity based unsupervised methods are not able to predict the statutes quite well. Contextual pretrained models performs little bit better than standard approaches. However, this highlights the challenges associated with the task. In general, we don't have enough annotated data for statutes to train neural models. Hence, it is an interesting direction to check how document ranking models trained on standard query-document pairs perform over the legal query-statute pairs.

5.3 Explainability using Occlusion Model

Along with the prediction, we also try to find out the correctness and relevance of the prediction. We do a posthoc analysis of the predictions. For a query, we take a statute that is relevant to it (as per gold standards not system prediction). Thereafter we want to analyze which sentences in the query (fact) description explain the relevance of the statute. Thus, we require a way to assign importance to the sentences in the query, in order to utilize them for explanation.

We deploy the occlusion model [8] to move towards a quantitative approach to explainability. We calculate scores (BM25 score, cosine similarity for word embedding based models) between statute description and query to complete the task of ranked retrieval of statutes, and refer to it as the original similarity score. Further, we calculate these scores by deleting one sentence from the fact description (query), maintaining other sentences intact. Thus we have the similarity scores between a statute and 'n' number of modified queries. The modified query stands for the original query minus one of the sentences and 'n' is the number of total sentences in the query. Thus we hypothesize that the removal of a sentence that results in the biggest dip from the original similarity score, must be the most important sentence. This sentence must be the top contender that explains the prediction. Using this logic we obtain a ranked list (up to the number of sentences annotated by the legal expert) of sentences that explain the prediction.

To understand this logic in mathematical terms we consider the fact description as a sequence of sentences, $Q = [S_1, S_2, ..., S_n]$. The similarity score between a query, Q and a statute St is represented as , $sim(Q, St)$. When a sentence, say the i^{th} sentence is deleted from the original query Q, the new query is denoted as $Q^{-i} = [S_1, S_2, ..., S_{i-1}, S_{i+1}, ..., S_n]$. Thus for every $i \in [1, n]$ we have n reformulated queries and correspondingly n similarity scores, $sim(Q^{-i}, St)$. Thus now we order these scores in descending order of the dip in similarity score value, $dip = sim(Q, St) - sim(Q^{-i}, St)$, where $i \in [1, n]$. Thus we get our ranked list of sentences as per the dip in similarity score value i.e. the argument i from the dip arranged in descending order. As stated above the most sentence that contributes most to the explanation is the first sentence in the ranked list.

We compare the explanations provided by the system with that provided by the legal expert. We a priori know that for the query AILAQ18, we have S137 one of the relevant statutes. The expert provided annotations are shown in yellow in Figure 1. The explanations provided by the system for same query-statute pair are highlighted below in green.

Thus we can observer the common sentences that both the legal expert and our system deem to be the explanations for the choice of S137 for query AILAQ18.

To evaluate the performance of our occlusion model we use document embedding provided by BERT, and cosine similarity as our choice of similarity measure. We calculate 'Precision at k' for each query to evaluate the occlusion model. The value of k for every query-statute pair is taken to be the number of annotated sentences by the legal expert. The average score over all queries is 0.206.

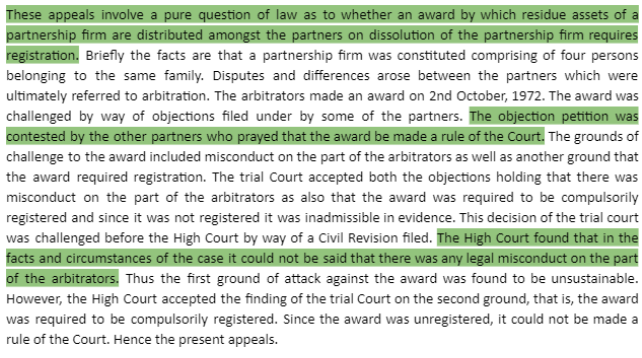
These appeals involve a pure question of law as to whether an award by which residue assets of a partnership firm are distributed amongst the partners on dissolution of the partnership firm requires registration. Briefly the facts are that a partnership firm was constituted comprising of four persons belonging to the same family. Disputes and differences arose between the partners which were ultimately referred to arbitration. The arbitrators made an award on 2nd October, 1972. The award was challenged by way of objections filed under by some of the partners. The objection petition was contested by the other partners who prayed that the award be made a rule of the Court. The grounds of challenge to the award included misconduct on the part of the arbitrators as well as another ground that the award required registration. The trial Court accepted both the objections holding that there was misconduct on the part of the arbitrators as also that the award was required to be compulsorily registered and since it was not registered it was inadmissible in evidence. This decision of the trial court was challenged before the High Court by way of a Civil Revision filed. The High Court found that in the facts and circumstances of the case it could not be said that there was any legal misconduct on the part of the arbitrators. Thus the first ground of attack against the award was found to be unsustainable. However, the High Court accepted the finding of the trial Court on the second ground, that is, the award was required to be compulsorily registered. Since the award was unregistered, it could not be made a rule of the Court. Hence the present appeals.

Figure 4: Annotations by the system

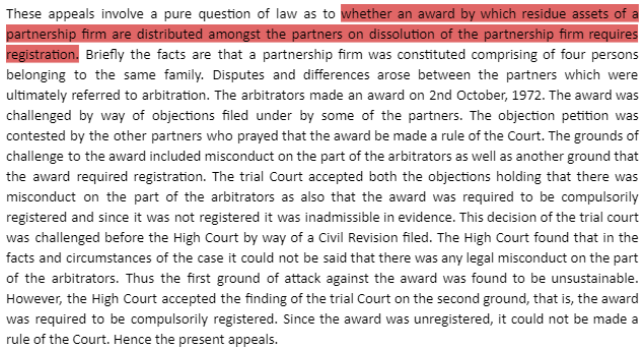
These appeals involve a pure question of law as to whether an award by which residue assets of a partnership firm are distributed amongst the partners on dissolution of the partnership firm requires registration. Briefly the facts are that a partnership firm was constituted comprising of four persons belonging to the same family. Disputes and differences arose between the partners which were ultimately referred to arbitration. The arbitrators made an award on 2nd October, 1972. The award was challenged by way of objections filed under by some of the partners. The objection petition was contested by the other partners who prayed that the award be made a rule of the Court. The grounds of challenge to the award included misconduct on the part of the arbitrators as well as another ground that the award required registration. The trial Court accepted both the objections holding that there was misconduct on the part of the arbitrators as also that the award was required to be compulsorily registered and since it was not registered it was inadmissible in evidence. This decision of the trial court was challenged before the High Court by way of a Civil Revision filed. The High Court found that in the facts and circumstances of the case it could not be said that there was any legal misconduct on the part of the arbitrators. Thus the first ground of attack against the award was found to be unsustainable. However, the High Court accepted the finding of the trial Court on the second ground, that is, the award was required to be compulsorily registered. Since the award was unregistered, it could not be made a rule of the Court. Hence the present appeals.

Figure 5: Annotations common amongst expert and system

5.4 Observation and Challenges

This initial study on explainable statute detection opens a couple of challenges. We list them as follows:

(1) We observe that simple similarity based matching does not work well. Even traditional term based matching algorithm (BM25) does not work. This suggests that distribution of legal queries and vocabularies used are quite different. Hence, incorporating domain knowledge is important for model performance. On the other hand, it is very difficult to gather large amount of training data for such tasks. Apart from that, legal system of different countries follow different patterns which complicates the task many fold. Hence, one of the challenges is to achieve comparable performance with minimal amount of training data.

(2) The occlusion results show low overlap between system suggested explanation and those annotated by legal experts. This is consistent with the observation of Malik et al. [8]. This suggests that we need to consider expert annotations as a part of the prediction task. Posthoc explanations [12] in general can't explain global nature of the model. To this end, we need a larger set of legal-domain specific training data (as opposed to 50 explanation-annotated case documents in the current paper). However, we have to minimize this effort by transferring some knowledge from the existing annotated data for other countries or related tasks.

6 CONCLUSION

In this paper, we analyzed the challenges in generating explanations for the statute prediction task. We realized that we need sizeable number of case facts annotated with explanations by legal experts. However, annotation process is also time-consuming and cumbersome. Hence, the objective is to accomplish the task with minimum amount of human supervision. We will explore other semi-supervised and transfer learning strategies. Side by side, system explanations are also not coherent with the the ground truth that indicates the difficulty in the learning process of the model. Hence, explainability should be part of the model development process. We also look forward to generating the same and designing a machine learning model that integrates expert-generated explanations as a part of the statute prediction task.

REFERENCES

[1] Paheli Bhattacharya, Kripabandhu Ghosh, Saptarshi Ghosh, Arindam Pal, Parth Mehta, Arnab Bhattacharya, and Prasenjit Majumder. 2019. Overview of the FIRE 2019 AILA Track: Artificial Intelligence for Legal Assistance. In *Working Notes of FIRE 2019 - Forum for Information Retrieval Evaluation, Kolkata, India, December 12-15, 2019 (CEUR Workshop Proceedings, Vol. 2517)*. CEUR-WS.org, 1–12. http://ceur-ws.org/Vol-2517/T1-1.pdf

[2] Wenhan Chao, Xin Jiang, Zhunchen Luo, Yakun Hu, and Wenjia Ma. 2019. Interpretable Charge Prediction for Criminal Cases with Dynamic Rationale Attention. *Journal of Artificial Intelligence Research* 66 (11 2019), 743–764. https://doi.org/10.1613/jair.1.11377

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[4] Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*. PMLR, 1188–1196.

[5] Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*. Association for Computational Linguistics, Barcelona, Spain, 74–81. https://aclanthology.org/W04-1013

[6] Bingfeng Luo, Yansong Feng, Jianbo Xu, Xiang Zhang, and Dongyan Zhao. 2017. Learning to Predict Charges for Criminal Cases with Legal Basis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Copenhagen, Denmark, 2727–2736. https://doi.org/10.18653/v1/D17-1289

[7] Vijaymeena M K and Kavitha K. 2016. A Survey on Similarity Measures in Text Mining. *Machine Learning and Applications: An International Journal* 3 (03 2016), 19–28. https://doi.org/10.5121/mlaij.2016.3103

[8] Vijit Malik, Rishabh Sanjay, Shubham Kumar Nigam, Kripabandhu Ghosh, Shouvik Kumar Guha, Arnab Bhattacharya, and Ashutosh Modi. 2021. 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)*. Association for Computational Linguistics, Online, 4046–4062. https://doi.org/10.18653/v1/2021.acl-long.313

[9] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[10] Shounak Paul, Pawan Goyal, and Saptarshi Ghosh. 2020. Automatic Charge Identification from Facts: A Few Sentence-Level Charge Annotations is All You Need. In *Proceedings of the 28th International Conference on Computational Linguistics*. International Committee on Computational Linguistics, Barcelona, Spain (Online), 1011–1022. https://doi.org/10.18653/v1/2020.coling-main.88

[11] Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, 45–50. http://is.muni.cz/publication/884893/en.

[12] Marco Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings NAACL*. 97–101.

[13] Karen Spärck Jones, S. Walker, and Stephen E. Robertson. 2000. A probabilistic model of information retrieval: Development and comparative experiments. *IP&M* 36, 6 (2000), 779–808, 809–840.

[14] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings EMNLP*. 38–45.