

Finding Important Arguments from a Legal Case

Daniel Konstantynowicz¹, Franciszek Grzegorz Wojciechowski¹ and Procheta Sen¹

¹*Department of Computer Science, University of Liverpool, United Kingdom*

Abstract

Within the legal field, the strength of an argument can be the deciding factor in the outcome of a case. Lawyers find themselves spending hours going over and analysing legal precedents and relevant statutes to build a compelling argument, and with the exponential growth in case data availability it can be useful for a lawyer if an artificial intelligence tool can automatically show important arguments present in a similar case from the past. In this work we propose an approach to estimate the importance of arguments using Natural Language Processing techniques. As a first step arguments are extracted from a legal case, and then the importance of each argument is estimated in an automated way. We explored both supervised and unsupervised approach to estimate the importance of arguments in a legal case.

Keywords

Argument Importance, Natural Language Processing, Information Extraction

1. Introduction

Within the legal field, the strength of an argument can be the deciding factor in the outcome of a case. Lawyers find themselves spending hours going over and analysing legal precedents and relevant statutes to build a compelling argument. With the exponential growth in data availability and the need to be able to keep up with competitors, there has been an increasing call for tools which could be used to assist in research and analysis of legal cases.


Natural Language Processing (NLP) is a branch of artificial intelligence concerned with giving computers the ability to understand text and spoken words in much the same way human beings can. NLP combines rule-based modelling of the human language, with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to „understand,“ its full meaning, complete with the speaker or writer,“s intent and sentiment ¹.

In this work, we use NLP techniques to analyse legal documents and to identify the strengths of the arguments present within each case. The findings of this work will provide insights into the potential of NLP in assisting lawyers with preparing for cases, as well as help assist judges by showing what the collection of past precedents have dictated. We explored both supervised (i.e. where training data is required) and unsupervised approaches (i.e. where no training data is required) to estimate the importance of arguments. We used data from Indian legal proceedings to estimate the importance of arguments. To the best of our knowledge, this is the first work on estimating the importance of legal arguments using NLP techniques.

CMNA'23: Workshop on Computational Models of Natural Argument, December 1st, 2023, online



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

¹<https://www.ibm.com/topics/natural-language-processing>

2. Proposed Methodology

The first step to estimate the argument importance is to automatically extract arguments from a legal case. To this end, we apply a method proposed in [1] to automatically extract arguments from the legal case. In the following subsection we have provided a brief overview of the proposed approach for automatically extracting arguments from the legal case.

2.1. Argument Extraction [1]

The study in [1] showed that a legal case can be broken down into several rhetorical roles. The different labels are a) Facts, b) Ruling by Lower Court, c) Argument, d) Statute, e) Precedent, f) Ratio of decision, g) Ruling by Present Court. Facts refer to the chronology of events that led to the filing of the case, and how the case evolved over time in the legal system. Ruling by Lower Court refer to the judgments given by lower courts (Trial Court, High Court). Argument refers to the discussion on the law that is applicable to the set of proven facts. Statute refers to the established laws, which can come from a mixture of sources. Precedent refers to the prior case documents. Ratio of decision refers to the application of the law along with reasoning/rationale on the points argued in the case. Ruling by Present Court refers to the ultimate decision and conclusion of the Court.

[1] used a Hierarchical BiLSTM CRF model [2] to automatically assign one of the seven rhetorical labels mentioned above to each sentence of a legal case document. They used a manually labelled data by legal experts to train the BiLSTM CRF model. In the context of this research, we are only interested in sentences belonging to Argument and Ratio of Decision category in a legal document.

2.2. Argument Importance

Once arguments are extracted, we explored both unsupervised and supervised approaches to estimate the importance of a particular argument. Each one of them is described as follows.

2.2.1. Unsupervised Approaches

For unsupervised approach, clustering methods were used to identify unique arguments from the set of arguments obtained from the method described in Section 2.1. The motivation for using clustering is to investigate whether unique arguments can contribute to the most important arguments in a legal case.

Clustering Based Approaches We specifically used DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering algorithm in this research scope. DBSCAN groups similar data points based on their density. It has the advantage of finding arbitrarily shaped clusters and identifying noise (outliers) in the data when the popular k-means clustering algorithm has no ability to automatically determine the number of clusters, lacks the ability to identify clusters with arbitrary shapes, and does not perform well with clusters of different densities. To cluster the set of arguments, we have used two different representation methodologies to present each argument sentence. Each one of them are described as follows.

TF-IDF Vector Based Clustering TF-IDF Vectorizer converts each sentence in a vector of floating point numbers. The main idea behind TF-IDF is to give more importance to words that appear less frequently in the entire corpus but more frequently in an input instance, thus helping to identify key terms that distinguish an input instance.

BERT Based Clustering The second representation approach used for clustering is BERT[3]. BERT captures context of a sequence of words from both ways (i.e. both left and right context). BERT was pre-trained on a large corpus of text that can learn in general language. As a result of this BERT is able to represent the semantic meaning present in a sentence better than the TF-IDF approach.

2.2.2. Supervised Approaches

For supervised approach, we trained a neural network to estimate the importance of an argument. The neural network takes an argument as an input, this is a vector that states which arguments are and are not present. Depending on the pre-processing done the vector changes in size and therefore the number of neurons in each layer of the model will change accordingly. The models designed for this project was a sequential neural network, which were chosen due to their adaptive learning, self-organization, and fault-tolerance capabilities([4]). The model consists of an input layer, 6 hidden layers. Each hidden layer utilises a sigmoid function and then an output layer.

3. Experiment Setup

Dataset We used a series of 3, 564 cases crawled by from Thomson Reuters Westlaw India website⁽²⁾. It contains cases ranging from 1951 to 2016 and fits into the following 5 categories: Land & Property, Constitutional, Criminal, Intellectual Property, Labour and industrial law. For unsupervised approach we have used cases from all the categories. However, for supervised approach we focused only on criminal cases. The reason for choosing only criminal cases for supervised approach is that training a neural network should be better on similar types of cases.

To train the model it is given an importance score for each argument in the legal case, this is found by measuring the similarity between each argument and the ratio of decision. The resulting importance scores are then normalised by dividing each score by the sum of all the argument similarity scores. The output score vector has the same structure as the input argument vector, with each index in the vector referencing the index of the argument in the dictionary. However this time, instead of a binary input, if an argument is present its normalised similarity score is provided otherwise the value is set to 0.

Pre-Processing As a preprocessing step, we removed all the redundant information from the legal case document. For example, information about citation and involving parties in each legal case is a redundant information for our proposed approaches. Hence these information was removed from the beginning and end of each legal case document. URLs and non-English

²<https://www.westlawasia.com/>

characters were also removed from each case document using regex pattern. We used sentence tokenizer available in NLTK³ to convert each case document into an array of sentences.

Evaluation There is no publicly available dataset where the importance of arguments are manually labelled by legal experts. As a result of this, we opted for an implicit judgement of our proposed approach for finding important arguments. We first took the final decision segment (sentences belonging to Ratio of Decision category) corresponding to each legal case. It has been observed that the most important arguments presented during the court proceeding is again described in the final judgement section. Hence we compute the similarity of the predicted important arguments and the final judgement section. If that similarity is greater than a particular threshold then we consider that predicted argument as one of the important arguments. If out of K predicted arguments only k_1 are important then the accuracy for that particular instance is computed as $k_1/k * 100$. We set the similarity threshold for our case as 0.4.

4. Results

Results for Unsupervised Approaches Figure 1 shows the visualization of the clusters using both TF-IDF technique. It can be easily observed that with the increase in the number of clusters the data points corresponding to different clusters are observed nearby. As a result of this, it can be concluded that a decrease in the number of clusters will help to identify more unique arguments from a case. We also had similar observation for BERT based clustering approach. We manually observed the results obtained from the unsupervised approach and our conclusion was that it was not being able to identify any important arguments. Hence we finally opted for supervised approach.

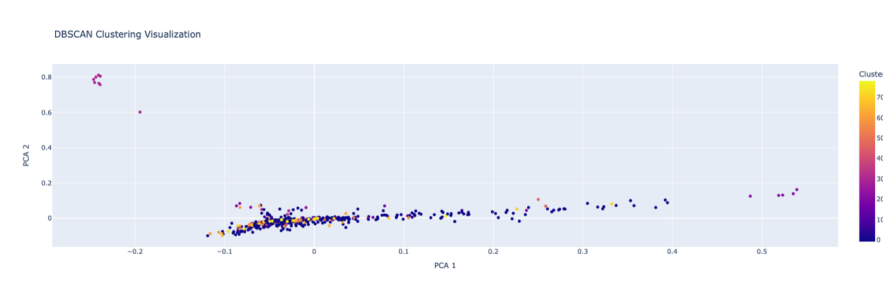


Figure 1: Visualization of Data Points Corresponding to Different Clusters Using TF-IDF Vectors.

Results for Supervised Approaches Figure 2 shows that the best performing version of the proposed model gives 15% accuracy with Adagrad optimizer.

³<https://www.nltk.org/>

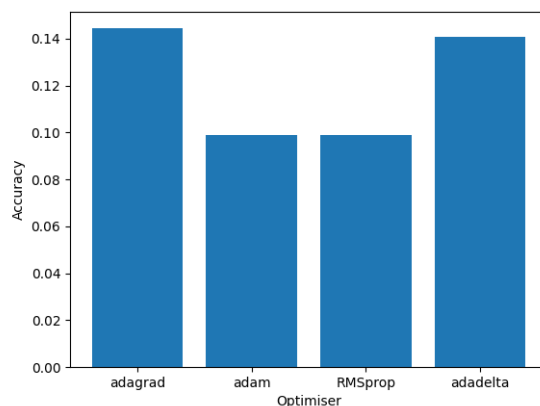


Figure 2: Graph Showing Accuracy of the Neural Model Against Different Types of Optimizers

5. Conclusion

In this work we investigated how we can use existing supervised and unsupervised NLP techniques to estimate the importance of arguments in a legal case. The best performance obtained from the neural approach solution was 15%. However, we would like to mention that this is a work in progress. The results described in this paper is the output of initial investigation. The major challenge in finding important arguments is the lack of labelled data and the noise present in legal case data. We hope to achieve better solution with more sophisticated techniques in future.

References

- [1] P. Bhattacharya, S. Paul, K. Ghosh, S. Ghosh, A. Wyner, Identification of rhetorical roles of sentences in indian legal judgments, CoRR abs/1911.05405 (2019). URL: <http://arxiv.org/abs/1911.05405>. arXiv:1911.05405.
- [2] Z. Huang, W. Xu, K. Yu, Bidirectional lstm-crf models for sequence tagging, 2015. arXiv:1508.01991.
- [3] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186.
- [4] S. Bhattacharya, An overview of neural approach on pattern recognition, 2020. URL: <https://www.analyticsvidhya.com/blog/2020/12/an-overview-of-neural-approach-on-pattern-recognition/>.