How to Cite:
Lalitha, Y. S., Raju, N. V. G., Teja, V. R., Sravani, P., & Reddy, E. S. (2022). AI enabled legal assistance system: A case study. *International Journal of Health Sciences*, *6*(S3), 6835–6844. https://doi.org/10.53730/ijhs.v6nS3.7553

AI enabled legal assistance system: A case study

Y. Sri Lalitha
Professor, Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, JNTUH, Telangana, India
*Corresponding author email: srilalitham.y@gmail.com

Dr. N. V. Ganapathi Raju
Professor, Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, JNTUH, Telangana, India

V. Ram Teja
Students, Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, JNTUH, Telangana, India

P. Sravani
Students, Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, JNTUH, Telangana, India

E. Sridhar Reddy
Students, Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, JNTUH, Telangana, India

***Abstract*---**Given a Case, finding the related prior cases and their judgements is a time consuming job of a Lawyer. The lawyer has to go through huge volumes of law books and prepare his case. An automated tool that retrieves the relevant past cases and their judgments is a very useful application for lawyers. It is a complex task especially in Indian context, the cases and their judgments are un-structured, and there is no standard format of case and judgement presentation. It is understandable for a lawyer but, a most difficult task for a machine. The work here presents a case study to retrieve judgments given in the past for a given factual description. The dataset considered for this work is selected from FIRE-2019, AILA track. The previously developed models showed best average precision of 0.149 using BM25, which itself demonstrates the challenging aspect of the given task. In this work LDA, a probabilistic algorithm for Topic Modelling is explored and studied. The proposed method has shown improved precision.

***Keywords*---**prior legal cases retrieval, LDA for legal facts, probabilistic data retrieval.

Introduction

Common Law system, which is followed by most countries (e.g., UK, USA, Canada, Australia, India etc.) has two primary sources – Precedents and Statutes. Precedents are the prior cases decided in the Courts of law. Statutes are bodies of written law, such as the Constitution of a country. Legal practitioners use past cases, facts and legal concepts to bolster their arguments before the judge for the desired outcome in a case. In terms of the principles, situations with similar facts should obtain similar decisions. Identifying the facts in past cases is a necessary but time-consuming undertaking for a legal practitioner. In this paper, we offer an algorithm for retrieving documents containing referenced facts and preparing legal reasonings based on them. Because these records are many, having an automated programme that can return previous cases will be advantageous to a lawyerThere is approximately 15 to 20 percent difference in the results when we consider only English and Hinglish sentences.

Related Work

In the areas of common law, attorneys consider present court decisions (precedents) as the source of the law. Case quotations, forensic refer¬¬ences, are valuable means for litigation, which allows attorneys to construct and present their argument convincingly. It goes hand in hand with 'Stand by the cases', in which the case under review with the same facts sufficiently should have the same conclusions as the precursors based on the same circumstances. Excerpts from existing law are utilized to show the principles and facts that illustrate how the principles are in the present case. Quotation analysis can assist legal practitioners to identify which principles have been applied in a particular case and which facts have influenced the outcome and which are important for finding similarities between the two cases. Generally, when considering that the two cases are less likely to be the same in all respects, legal experts should consider a variety of preconceived notions, each of which highlights certain facts and legal principles that underpin their argument (or may be used against arguments). Therefore, it is important that each side of the legal dispute identifies the basis for appropriate litigation that supports the legal claims made during the legal dispute. As the subject of general law grows ever stronger, analysing people's quotes is more complex and informative and timely.

This study aims to use machine learning techniques to automatically identify legal cases and judgments. Much work has been done in the area of quotation analysis in scientific research, much of the work can be used to identify previous legal cases and judgments. More recently there has been an increase in interest in legal citation studies, in which there appear to be two major guidelines: using quoted network analysis [4] and segregation programs that allow one to estimate the '' treatment '' status of a quoted case [5, 6]. Attorneys are frequently required to identify a set of similar instances that can be used to support the argument since the common law standards of law are defined and coded in current

decisions The study of legal authorities is frequently carried out with a variety of methods, which according to [1] can be grouped into three categories — primary, secondary, and tertiary materials to make decisions. Generally, Legal experts use a combination of information from primary, secondary, and tertiary sources.
The work in [2] alludes to the fact that the increasing number of available citations is causing overload of information, with the court emphasising the necessity to only cite instances that established the principle of law, rather than sources that were just used to explain it. [2] He criticises modern verdicts for being overly long, difficult to understand, and containing proportions "hidden in the sea of extraneous information." In the field of law there has been a surge in interest in citation studies, with two main avenues emerging: applying network analysis to citations and using citation data to create chevaliers. The goal of the study in [3] was to automatically extract "argumentation-relevant information from a corpora of legal judgement documents" and "construct new arguments utilising that information.". The BM25 score between each question and each applicant was counted in [7] after pre-screening. The top two documents returned as a result of this search are considered important. The vectors of documents are read using Doc2Vec in the instance of doc2Vec. The standard points between each question and each contender were then calculated, and the top two documents were returned as a consequence.

The main strategy in [8-10] was to index case files and use different BM25, TFIDF, and Word2Vector ways to represent the document vector in order to extract keywords as a query. Texts were measured using the Euclidean range. The test approach used in [11] was mostly dependent on document keywords. To begin, they extracted keywords from the searches using TF-IDF and text placement. They then recover keywords using the vector space model, language model, and BM25 model. A number of topic words were chosen for testing. The results were obtained using the vector space model in conjunction with the TF-IDF approach.In [12] After pre-processing the documents, they used Glove, Doc2Vec and TF-IDF based methods. In [13] the language model assorting algorithm of Indri, Lucene and with Dirichlet Similarity were used. Hypothesis: Can we apply Probabilistic approach to group the related cases and retrieve the most relevant judgements for a given query case. As clustering is to group the most relevant documents together in a cluster. Searching a cluster for relevant prior cases will optimize the process. We experimented Topic modelling technique for the task.

Methodology

Dataset Collection

The dataset was obtained through FIRE 2019, and it consists of 2194 case documents provided by the Supreme Court of India. After registering for the conference we obtained Query document and relevance judgements prior cases.

Pre-processing

Pre-processing is a crucial step in Machine Learning Process. Better results are achieved with pre-processed data. Different text pre-processing techniques are

studied to sift noisy and non-informative text from the dataset. Several Text cleaning methods are applied to extract crisp, error-free data from large collection of Legal Documents. We have applied Converting Words to Lower Case, Tokenization, Stop Word Removal, Lemmatization and Stemming.

Table 1
Sample Dataset of Legal Cases

	Article
0	This appeal, by special leave, by the appellan...
1	In this appeal, by special leave, the appellan...
2	This appeal is brought by special leave from t...
3	The question involved in this case is whether ...
4	Leave granted.Ruling by Present Court\nRespond...
...	...
56	Accused Pathan Hussain Basha,was married to Pa...
57	In these appeals the dispute relates to paymen...
58	Leave granted in SLP (CrlFacts\nNo.3737/2014Fa...
59	This appeal is preferred against the judgment ...
60	This appeal by special leave is directed again...

Feature Engineering

To determine the classification rules, the features must be expressed using numeric values. This stage involves extracting essential features from raw text and numerically expressing the extracted features. Filter out tokens that appear in 0.3% of the dataset. The formed Document Feature vector is the training set for Model Construction.

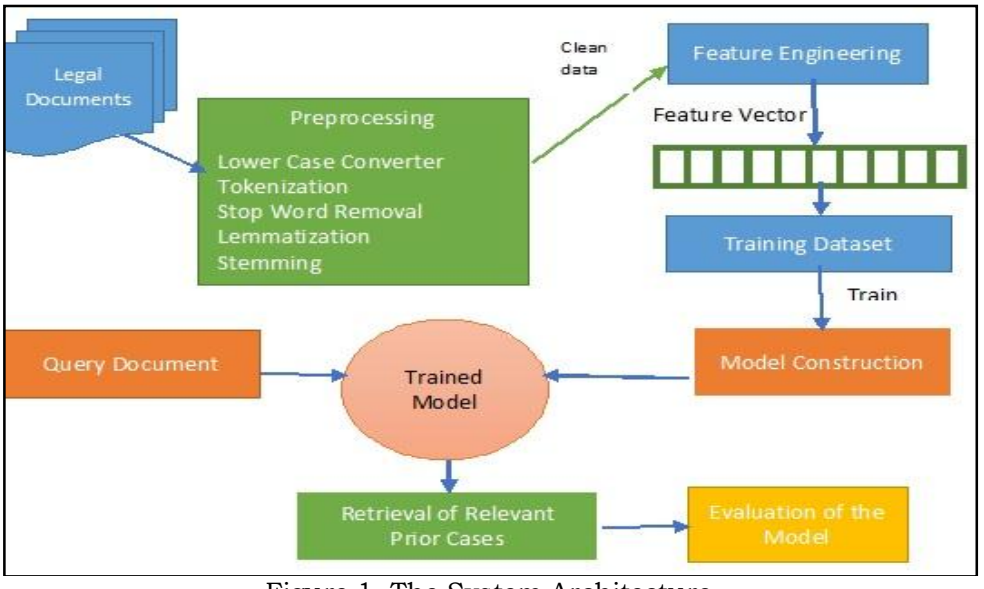

Figure 1. The System Architecture

Model Construction

Previous works demonstrated various methods. We choose to work with a model that is not tried in previous works and therefore selected Topic Modelling with Latent Dirichlet Allocation. It is scalable, computationally fast, and most importantly, it provides simple and understandable topics that are similar to those assigned by the human mind when reading a document. While LDA is most commonly employed for unsupervised tasks like topic modelling and document clustering, it may also be utilised as a feature extraction technique for supervised tasks like text classification. The latent subjects that make up documents are concealed in the data. Dirichlet refers to the distribution of distribution, as well as the distribution of topics in papers and the distribution of words within the topic. Topics are assigned to documents and document terms are assigned to topics based on distributions of distributions.

$\Theta(td) = P(t/d)$
Is Prob. distri. of Topics in Documents

$$\Theta(td) = P(t/d)$$

Is Prob. distri. of Words in Topics

P(w/d) is Prob. distri. of W in D
given by dot product of $\Theta(td)$ *and* $\Phi(wt)$ *for each topic t*

$$P(w|d) = \sum_{t=1}^{T} P(w/t)P(t/d)$$

Decompose the probability distribution matrix of words in a document into two matrices, one for the distribution of topics in a document and the other for the distribution of words in a subject, as illustrated in

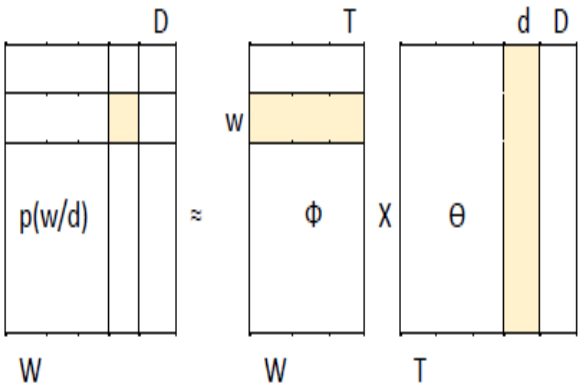

Figure 2. LDA Computation

Training Model

To Extract best features we need to find optimal number of topics for modelling on the Legal Documents Collection. To determine this number automatically from the data set, Hierarchical Dirichlet Process) is implemented. HDP is a non parametric Bayesian approach that learns the distribution of latent topics from the documents. Then the model is trained using LDA for the inferred no. of topics. The relevant words for a specific topic generated are stored.

Query Document

Each of the 50 questions illustrates the situation that led to the filing of a case in an Indian court of law (in the original English language). The data collection includes documents of 2,914 previous cases heard in the Supreme Court of India. For each question, the aim is to return the most significant / appropriate case documents in relation to the situation in the given question. The term of `compliance 'can be understood in this context as follows : the previous issue is considered important in the case when the case discusses a similar situation in the question, as judged by legal experts. The model is tested on unknown legal case to determine the prior cases. The query document is pre-processed, applied topic modelling to determine the topics and retrieved the top 5 relevant case documents. To determine the similarity between the query and the documents in the document collection, the cosine similarity measure is applied on the features obtained from the trained document and the query document. The Query Documents and Prior cases Relevant to a query are provided in two files. The parsed Query document is provided for testing the model. The sample Queries and Relevant documents are shown in table below

Table 2
Sample Query Relevant Prior Cases

Slno	Query	Relevant
1	AILA_Q1	C9,C14
2	AILA_Q2	C22,C27
3	AILA_Q3	C1
4	AILA_Q4	C182
5	AILA_Q5	C36,C54,C121,C144,C155
6	AILA_Q6	C19,C26,C99,C152
7	AILA_Q7	C130
8	AILA_Q8	C32,C60,C125
9	AILA_Q9	C42,C90
10	AILA_Q10	C86,C180,C185

Evaluation

To evaluate the model the Precision and Recall measures are used. Since no. of query documents are 50, the model performance is demonstrated with Mean

Precision and Mean Recall computed by considering 1-10 query documents, then 1-20 query documents and so on and so forth till 1-50 query docs.

Recall : The system's ability to deliver all the relevant items.

$$Recall = \frac{no.of\ relevant\ itemes\ retrieved}{no.of\ relevant\ items\ in\ collection}$$

Precision **:** The system's ability to deliver only relevant items.

$$Precision = \frac{no.of\ relevant\ items\ retrieved}{total\ no.of\ items\ retrieved}$$

Mean un-interpolated average precision (MAP) and Mean average Recall were used as the metric to analyse the overall performance of the model.

$$MAP = \sum_{i=1}^{q} Pr_i$$

where q is the no. of queries and Pr is precision and similarly, Mean average Recall is given by

$$MAR = \sum_{i=1}^{q} Rc_i$$

where Rc is recall IV Results

Results

In the previous works the best performing MAP Score is 0.149 demonstrated by BM25 algorithm. This task is difficult, as evidenced by the comparatively low MAP Score. The present work implemented proposed LDA model and BM25. BM25 (Best Match 25) improves upon TFIDF measures in search retrieval. It performs very well in many ad-hoc retrieval tasks, especially those designed by TREC. BM25 is considered as baseline model. For this experiment 200 prior cases were taken for training data. 50 Queries provided by FIRE are taken as test data. LDA the no. of topics in the dataset is determined by Hierarchal Dirichlet Processes and we encountered 7. Around 30 words are considered as the topic words.

Table 3
Retrieved Query Relevant Prior Cases by BM25, LDA

Sl no	Query	Relevant	Retrieved BM25	Retrieved LDA
1	AILA_Q1	C9,C14	-	-
2	AILA_Q2	C22, C27	C21	C22
3	AILA_	C1	-	C1,C10

	Q3			
4	AILA_ Q4	C182	C21,C92,C191	C26,C92, C182

For the query document the topic words are determined. LDA is a clustering approach. The cosine similarity measure is used to determine the similarity between the topics and the query document. The top 10 prior cases were extracted.

Table 4
Average Precision, Recall Comparison

	BM25	LDA
Mean Average Recall	0.182	0.424
Mean Average Precision	0.151	0.309

Table 4 depicts the Average precision and Recall values when the model is run over all queries in the test data. It can be observed that the precision is .309 for LDA much better than previous work 0.149 [14] using BM25.

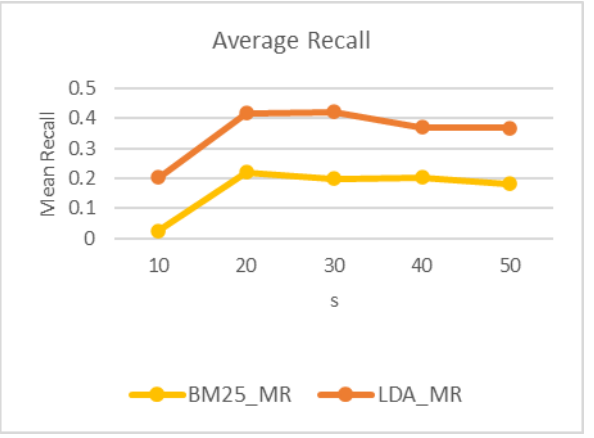

Figure 1 & 3. Average Recall Over all queries

Figure 3 depicts the Recall rate considering the no. of queries is step count of 10. It can be clearly observed that as the no. of queries increased there is a fall in recall, by both the models. Hence the models are performing well, but the fraction of recall rate is better with LDA when compared to BM25. Figure 4 depicts the rate of Precision considering the no. of queries in the increasing order of step count 10. The precision is improved consistently with the increasing size of query topics with both the models. The mean average precision of LDA is much better than the BM25 model.

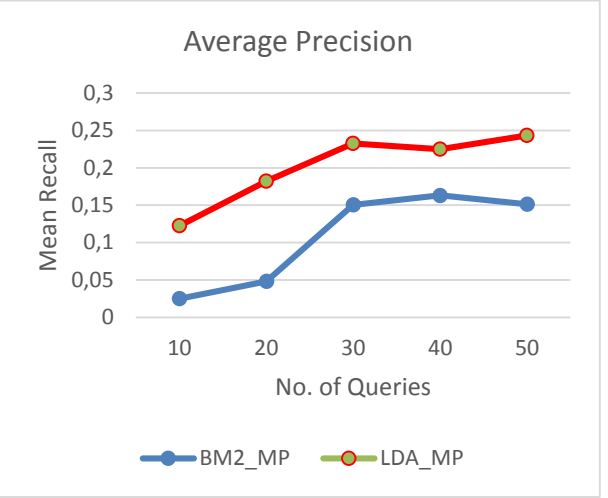

Figure 4. Average Precision Over all Queries

Another observation is that with both the models the first 10 queries performance evaluation is poor. The queries may not be relevant to the dataset considered. Since the dataset is first 200 legal cases. We experimented with the whole dataset of 2194 documents and trained both the models. It is observed for some queries the most relevant prior cases were fetched. Since its time consuming process to retrieve all relevant documents for a given query, it is claimed to be sufficient if the retrieved prior case is from first 'N' prior cases, where 'N' is a user defined value. This procedure showed better results especially for queries which could not retrieve atleast one prior case.

Conclusions and Future Work

The work presented here is a case-study to improve the accuracy of prior cases retrieval for legal. The proposed model demonstrated better accuracy, but is not satisfactory. In our future work we should see the impact of clustering the prior cases and perform classification and study its result. Another option that can be explored is application of deep learning models. Especially sequence to sequence model building may give better results. We will work on improving accuracy of result and try to print accurate solution for particular case study as well.

References

1. Geist A, "Using citation analysis techniques for computer-assisted legal research in continental jurisdictions", Ph.D. thesis, University of Edinburgh 2009.
2. Elliott C, et.al., "English legal system", in 14th edition. Pearson Education Inc, Karnataka (2013)
3. Ashley K. D, et.al "Toward constructing evidence-based legal arguments using legal decision documents and machine learning". in Proceedings of the fourteenth international conference on artificial intelligence and law, ACM, pp 176–180, 2013.

4. Zhang P, et.al., “Semantics-based legal citation network”, in proceedings of the 11th international conference on artificial intelligence and law, ACM, pp 123–130 2007.
5. Galgani F, Compton P, et.al “Lexa: building knowledge bases for automatic legal citation classification”. Expert System Application 42(17):6391–6407, 2015.
6. Zhang P, et.al., “Knowledge network based on legal issues”, in Winkels R (ed) Proceedings of workshop on network analysis in law (NAilL2013), pp 23–28, 2014.
7. Gain, B., Bandyopadhyay, et.al., “System Report for Artificial Intelligence for Legal Assistance Shared Task”, in Proceedings of FIRE 2019, December.
8. Zhao, Z., et.al.,”Legal information retrieval using improved BM25”, in Proceedings of FIRE 2019, December..
9. N. N. A. Balaji, B. et.al., “Legal information retrieval and rhetorical role labelling for legal judgements”, in Proceedings of FIRE 2020.
10. M. Wu, et.al. “Retrieval model and classification model for aila2020”, in: Proceedings of FIRE 2020.
11. Gao, J, et.al., “Legal retrieval based on information retrieval model”, in: proceedings of FIRE 2019.
12. I. Almuslim, D. Inkpen, “Document level embedding for identifying similar legal cases and laws”, in: Proceedings of FIRE 2020.
13. Y. Xu, T. Li, Z. Han, et.al., “The language model for legal retrieval and bert-based model for rhetorical role labelling for legal judgments”, in: Proceedings of FIRE 2020.
14. Paheli Bhattacharya1, Kripabandhu Ghosh2 et.al. “Overview of the FIRE 2019 AILA Track: Artificial Intelligence for Legal Assistance”, in Proceedings of the 11th Forum for Information Retrieval Evaluation, December 2019,, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In AAAI (pp. 4278- 4284).