Assistive Technology To Expedite Judicial Procedure Using Machine Learning And Natural Language Processing Techniques

Prof. Hemalatha KN[1], Ms. Saarah Asad[2], Ms. Aishwarya S[3], Ms. Payal Makhija[4]
[1, 2, 3, 4] Dept of Computer Science
[1, 2, 3, 4] Atria Institute of Technology

Abstract- *The Indian Judiciary stacks a pile-up of more than 30 million cases, some in backlog for more than 10 years. Renowned cases such as the 1984 Sikh riots still await a final judgement with many of the accused and those seeking justice having died in the extensive 32-year lengthy wait. This paper aims to present an application that would assist the jury and expedite the process of delivering a judgement. Given required documents at hand, judges take a minimum of one month to an infinite number of years to deliver the judgement. This paper proposes a system performing the same task in less than a minute. The application aims to summarize judicial support documents constructing an overview of them using NLP methods and categorize documents into categories as identified by experts with machine learning techniques.*

Keywords- Machine learning (ML), Natural language processing (NLP) , Information retrieval (IR) , Indian Judicial System, Document summarization

I. INTRODUCTION

The Indian judicial system stands as a prime pillar of the Indian democracy. But for all its sovereignty, it remains volatile, deficient in staff and above all slow. A person above 50 years of age, shouldn't even think of getting his case resolved through courts in his lifetime! The Judicial system is in crucial need of reforms if we are to continue to strengthen our society that dwells within a framework of fair justice for all.

Acquiring justice is riddled with obstacles from the very start for an average Indian, from filing an FIR to escalating it to lower courts. After the court proceedings end, a judge could take from a minimum of one month to more than 10 years to read documents relevant to the case, research and write the judgement. Additionally, the great number of vacancies existing in the courts makes the scenario truly distressing.

We propose a system that uses NLP, IR and machine learning techniques to perform the judge's tasks in less than a minute, originally taking a minimum of one month. The system aims to categorize judicial support documents constructing an overview of them and categorize sentences within accelerating the judicial process and decreasing the backlog of cases.

This paper is organized as follows. Section II describes the challenges faced by the judicial system and reason for backlog of cases. Technique for summarization of judicial support documents is given in section III. Categorization of text in a case is discussed in Section IV. Finally, conclusions are presented in section V

II. MODERNISATION OF COURTS

The Indian justice system is a principal pillar of the Indian democracy. India's judiciary follows an integrated three-tiered system. The Supreme Court, an all-india forum at the apex. High Courts crown each state's judiciary. Lower courts or district level courts form the third tier [1]. This section describes the judicial process after formulation of arguments and documents, reasons for delay in delivery of judgements and the methods to accelerate judgement delivery.

A. The Judicial Procedure

In this paper, we consider only those cases that come under the civil law sector and are under the district court jurisdiction. The process a case undergoes can be very briefly summarised as: Once the FIR is filed, the police commence investigation and collect evidence pertaining to the case.

1. If the case is not resolved, it is documented and submitted to the lower court and proceedings commence. All proceedings are documented.

2. After the case has been heard in depth and arguments are advanced by both the parties, the judge usually

reserves judgement for a period that could be minimum of three weeks to an infinite number of years, in order to examine the arguments, do research and write the judgement before delivering it.

A. Backlog of Cases

According to official figures, more than 27 million legal cases are pending in various courts in India, 6 million of which have been stuck in courts for 5 years or more. India has a mere 16,000 courtrooms and the bare minimum number of judges to preside over and carry out hearings. For every 73,000 people in India, there is only one judge.

A present-day survey of litigants conducted by Daksh, "a civil society organization that undertakes research and activities to promote accountability and better governance in India", showed that greater than 60% of the respondents found that the set-back in their own cases was due to the not delay of judges in making judgements [2].

It is the delay on step 3, from the previous subsection, that this paper aims to rectify. The figure above represents statistics of time taken for judges to deliver orders on cases.

B. The Fix

The figure summarises that it would approximately require 2 months for a judge to research and write the judgement. We propose a system performing the same task in less than a minute!

III. DOCUMENT SUMMARIZATION

Legal documents are documents that state some contractual relationship or grant some rights or could be documentation of court proceedings or could be documents submitted by lawyers or advocates. These documents provide information that is official in nature. Using text summarization, we aim to summarise these documents to save the time it would take a judge to read and research on them. We incorporate a technique which results in a high percentage accuracy in including crucial, case deciding statements of paramount importance in the overview. Text summarization addresses both the problem of selecting the most important portions of text (extractive text summarization)and the problem of generating coherent summaries. Extractive text summarization methods work by identifying important sections of the text. The extractive text summarizer performs these three relatively independent tasks : Intermediate representation, Score sentences and Select summary sentences [3].

A. Intermediate Representation

Almost every text summarization system creates some form of intermediate representation of the text it intends to summarize. There are two approaches : "topic representation" and "indicator representation". "Topic representation approaches transform the text into an intermediate representation and interpret the topic(s) discussed in the text." It includes frequency, TF-IDF and topic word approaches, lexical chain approaches, latent semantic analysis and full blown Bayesian topic models. "Indicator representation approaches describe every sentence as a list \ features (indicators) of importance such as sentence length, position in the document, having certain phrases, etc."

B. Score Sentences

On obtaining the intermediate representation, each sentence is assigned a score to indicate its importance. For "topic representation approaches", the score indicates how well a sentences expresses the important topics in the document or how well it combines information about different topics. "In most of the indicator representation methods, the score is computed by aggregating the evidence from different indicators".

C. Select Summary Sentences

The summarizer selects the k most important sentences to generate a summary. The factors that need to be taken into account could include, for example, the context in which the summary is being created. The type of the document can be another factor that could impact the sentences that are selected.

IV. CATEGORIZATION OF DOCUMENTS

Automatic text classification has always been an important application and research topic since the inception of digital documents. Today, text classification is a necessity due to the very large amount of text documents that we have to deal with daily. In general, text classification includes topic based text classification and text genre-based classification. Text classification can be done using CRF.

Conditional random fields (CRFs) are a class of statistical modeling method often applied in pattern recognition and machine learning and used for structured prediction. CRFs fall into the sequence modeling family.

Whereas a discrete classifier predicts a label for a single sample without considering "neighboring" samples, a CRF can take context into account; e.g., the linear chain CRF (which is popular in natural language processing) predicts sequences of labels for sequences of input samples. CRFs are a type of discriminative undirected probabilistic graphical model. They are used to encode known relationships between observations and construct consistent interpretations and are often used for labeling or parsing of sequential data, such as natural language processing or biological sequences and in computer vision. Specifically, CRFs find applications in POS Tagging, shallow parsing, named entity recognition, gene finding and peptide critical functional region finding, among other tasks, being an alternative to the related hidden Markov models (HMMs).

V. CONCLUSION

The Indian judicial system forms the foundations of India's democracy. As the saying goes, justice delayed is justice denied. The judicial system has been failing to serve justice to millions with the pile-up of backlog cases that have been on the stack for years and years. Thus, we propose a system to accelerate the judicial process and decrease the backlog of cases.

The application we propose summarizes judicial support documents constructing an overview of them using NLP methods and categorizes it using Machine Learning.

REFERENCES

[1] Bhuwania, Anuj. Courting the People: Public Interest Litigation in Post-Emergency India. Cambridge University Press, 2017.
[2] Kumar, Alok Prasanna. "Delayed Justice: When Judgement Day Arrives Too Late." Https://Www.livemint.com/, Livemint, 6 June 2016.
[3] ALLahyari, Mehdi, et al. "Text Summarization Techniques: A Brief Survey." International Journal of Advanced Computer Science and Applications, vol. 8, no. 10, 2017.
[4] Rose, Stuart, et al. "Automatic Keyword Extraction from Individual Documents." Text Mining, Apr. 2010.
[5] Siddiqui, Tanveer, and U. S. Tiwary. Natural Language Processing and Information Retrieval. Oxford University Press, 2008.
[6] SITE, www.site.uottawa.ca/~diana/.
[7] Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. "Legal judgment prediction via topological learning". In Proceedings of EMNLP, 2018.