Court Judgment Prediction and Explanation based on Transformers

Summary

Candidate:
Salvatore Junior Curello
Supervisors:
Prof. Luca Cagliero
Dr. Irene Benedetto

December 2023

Over the last few years, Artificial Intelligence (AI) has attracted much attention as it is becoming increasingly common in our lives. AI involves the development of automated systems capable of performing tasks that typically require humans. It can be applied in a wide range of applications such as healthcare, finance, education, social media, agriculture, etc.

Machine Learning (ML) plays a pivotal role in this context. ML algorithms have the main advantage that they can adapt and evolve based on the information that they analyze.

This thesis is dedicated to exploring the legal domain, focusing on using Machine Learning techniques, particularly those based on Natural Language Processing (NLP) methods. NLP is a sub-field of AI that tries to make an interaction between computers and human language in order to enable machines to understand, interpret, and generate human language.

However, it is important to consider some intrinsic characteristics of this field. The legal domain is complex, and rich in peculiar terms that are very technical and unique to this area. Then, a certain knowledge is required to understand them properly. Unfortunately, common Language Models (LM) face limitations in this context, since they are trained on general language patterns and do not have specific knowledge which is essential to understand and interpret legal concepts. Furthermore, legal documents are typically quite long, verbose, and noisy. The limitations of LM in terms of the amount of data they can process, is another significant challenge in this field.

Thesis Objectives

This thesis focuses on two different but closely related tasks. The first one is the *Court Judgment Prediction* task. It consists of predicting the outcomes of legal cases by analyzing patterns in past cases and identifying relevant factors that may impact the outcome of a court decision. This can be beneficial in many highly populated countries characterized by a vast number of pending legal cases that impede the judicial process due to multiple factors, including the unavailability of competent judges.

The second part of this thesis, instead, focuses on the *Court Judgment Prediction and Explanation* task. In the legal domain, the mere prediction of a legal case is not sufficient if not accompanied by the corresponding explanation. Then, the goal is to provide a model that not only gives as outcome the final decision of a legal case but also indicates the reasoning behind the predicted decision.

Contributions

The main contribution of this work is the implementation of an automated system that performs:

- A Court Judgment Prediction task in which the primary contribution is an extensive experimentation with various Transformers Models, making a comparison between Generic models and Domain-specific Transformers.
- A Court Judgment Prediction and Explanation task in which the application of the occlusions method and attention mechanisms can help to mitigate the disparity between the explanations generated by the machine and those manually annotated by legal experts.

Methods

All the experiments have been conducted using the ILDC (Indian Legal Documents Corpus) which is composed of a large set of Indian Supreme Court cases, expressed in the English language and annotated with original court decisions ("accepted" v/s "rejected").

Concerning the Court Judgment Prediction task, all the Transformers are trained on the last N tokens of the documents where N is equal to the maximum amount of tokens supported by each model.

In the Court Judgment Prediction and Explanation task, instead, the $ILDC_{expert}$ corpus is used. It consists of 56 documents that are extracted from the Test Set and given to five legal experts who were requested to predict the judgment and mark the sentences that they considered explanatory for their decisions. For the explanation generation, the occlusion method and attention mechanism have been exploited to extract the most relevant sentences from the case description that best justify the final decision. Performance evaluation used a battery of metrics that measure the overlap between the expert annotators' gold explanations and those generated by the machine. These metrics essentially quantify how well a system generates text compared to human standards.

Results

Regarding the Court Judgment Prediction task, domain-specific Transformer models such as LegalLSGBERT, CaseLawBERT, and LegalBERT achieved the best performance in terms of F1-score, outperforming generic models. This demonstrated their effectiveness, even though they were pre-trained on US/EU legal documents with legal systems differing significantly from the Indian context. Furthermore, the application of Hierarchical Transformer models led to improved performances. In particular, the results were slightly higher when using the [CLS] token instead of Mean and Max pooling, proving to be the most effective strategy in almost all models.

In the Court Judgment Prediction and Explanation task, the results obtained using the occlusions method show that, by considering the top 40% of sentences, the baseline is outperformed. Using the attention mechanism, instead, achieved similar performances to the baseline, but the explanations were shorter than the golden annotations of the legal experts. In some cases, this brevity can be considered an advantage, facilitating quick insights to expedite the decision-making process.

Conclusion and future works

Domain-specific Transformer models in Court Judgment Prediction task show effectiveness in different legal systems. Occlusions method and attention mechanisms are crucial for extracting relevant sentences and generate the explanations. Improvements could involve applying

Transformer models pretrained on Indian legal documents to capture cultural nuances. Extending hierarchical attention models for datasets with multi-modal information is a potential enhancement for the explanation generation.