LLMS FOR STATISTIC LEGAL RESEARCH

Transition from manual labeling to automatic labeling of textual data through the implementation of language models, an effective sampling method, and an active learning labeling methodology.

AUTHORS

Poster Author:

Aviv Gelfand, Research Assistant at the Hebrew University of Jerusalem

Research Team Members:

Prof. J.J. Prescott, University of Michigan Law School

Dr. Adi Leibovitch, The Faculty of Law at the Hebrew University

Grady Bridges, Researcher at the University of Michigan Law School

Introduction

As part of a broad research project on the biases of judges during criminal law, 30k texts of defendants to the judge were collected and tagged manually, during the years 2019-2021. The way of labeling during that time was by alternating teams of students who manually went through each text and labeled the main types of arguments by the following list:

Label	Defenition
FINANCIAL	Personal financial difficulty to pay a fine or bear the financial consequences of the charge.
CIRCUMSTANCE	The specific circumstances of the commission of the offense (and not of the accused in general).
LICENSE	The need for a license in order to work, help a relative, etc.
REMORSE	Sincere remorse, acceptance of responsibility and intention to avoid committing the offense in the future.
GOOD_RECORD	A clean criminal record and the absence of previous offenses.
BAD_RECORD	A previous criminal record to which the offense in question may be acclimated to.
INNOCENCE	Claiming innocence and disclaiming responsibility for the offense in question.

Objective

Train language models that will replace and overcome manual taggers with high levels of accuracy to speed up research processes.

Sampling

Stratified Sampling by cardinality of the label, and by association with a TF-IDF (Term Frequency-Inverse Document Frequency) cluster.

Fine-Tuning LLM

Training language models (distillbert, llama-2 and Roberta) on a label classification task.

Inference

- Outputting the logits, the probabilistic predictions of the model (0 to 1).
- Learning and correcting wrong labeling according to the differences between the model and the real labeling.

Label	Priority	Accuracy	Auc-Roc	F1	Precision	Recall
GOOD_RECORD	1	0.973115	0.973546	0.951364	0.929316	0.974484
CIRCUMSTANCE	1	0.901148	0.909179	0.900312	0.910399	0.909037
LICENSE	1	0.991311	0.916055	0.888421	0.95045	0.833992
REMORSE	1	0.929157	0.885753	0.816359	0.824675	0.837592
FINANCIAL	1	0.954262	0.887627	0.802548	0.804255	0.800847
OUT_STATE	3	0.989344	0.838678	0.68599	0.68932	0.682692
BAD_RECORD	1	0.996129	0.960232	0.645161	0.8	0.540541
INNOCENCE	3	0.956557	0.769102	0.605067	0.663399	0.556164