CS6101-10

Legal Text Classifier with Universal Language Model Fine-tuning

LIM How Khang (LL.B. First Class Hons) (NUS), Advocate & Solicitor, Singapore

Overview

Recent work has shown that pre-training a neural language model on large public text datasets improves the accuracy of a neural text classifier while requiring fewer labelled training samples.

We use the Universal Language Model Fine-tuning method introduced by Howard and Ruder (2018)* to train a legal text classifier to perform 19-way legal topic classification on a dataset of 3,588 legal judgments issued by the Singapore Court of Appeal and High Court.

***Universal Language Model Fine-tuning for Text Classification (https://arxiv.org/abs/1801.06146)**

***https://github.com/fastai/fastai**

Dataset

The dataset was created by parsing the PDF Singapore Supreme Court judgments between 2000 and 2018 found on Singapore Law Watch (SLW), a free daily legal news website.

URL: https://www.singaporelawwatch.sg/Judgments

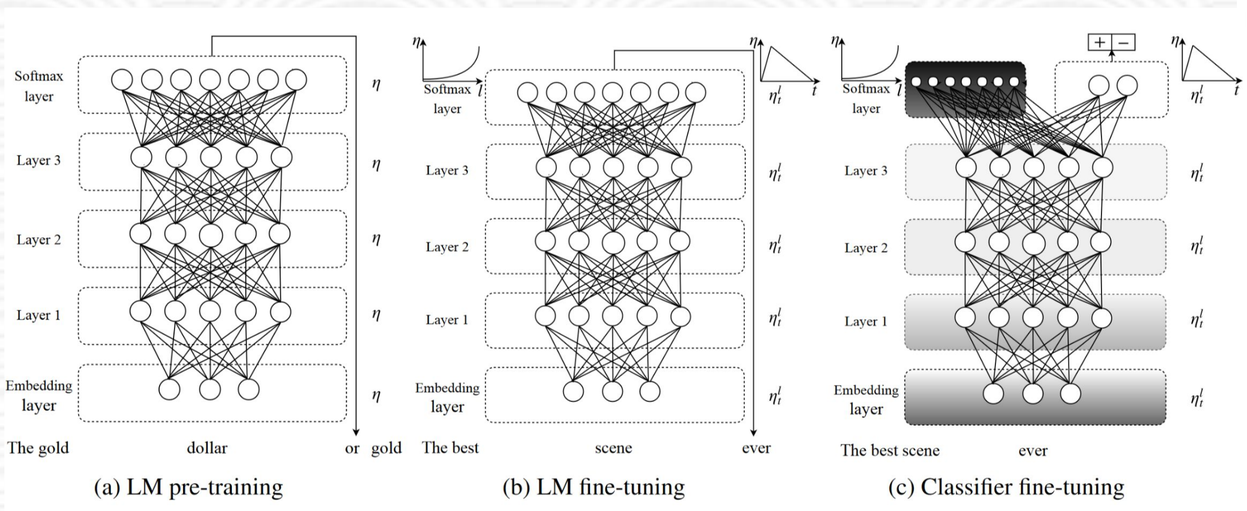
Size: 3,588 judgments (with only one listed topic)

Topics: Civil Procedure, Criminal Law, Contract Law, Family Law, Companies, Tort, Arbitration, Legal Profession, Damages, Land, Insolvency Law, Building and Construction, Conflict of Laws, Administrative Law, Revenue Law, Trade Marks, Admiralty Law, Employment Law, Trusts.

Universal Language Model Fine-tuning by Howard and Ruder (2018)

Wikitext-103 Language Model → **Specific Language Model** → **Specific Text Classifier**

- Pre-trained language model on Wikitext-103
- Discriminative fine-tuning
- Slanted triangular learning rates
- Target task classifier fine tuning with additional linear blocks and gradual unfreezing



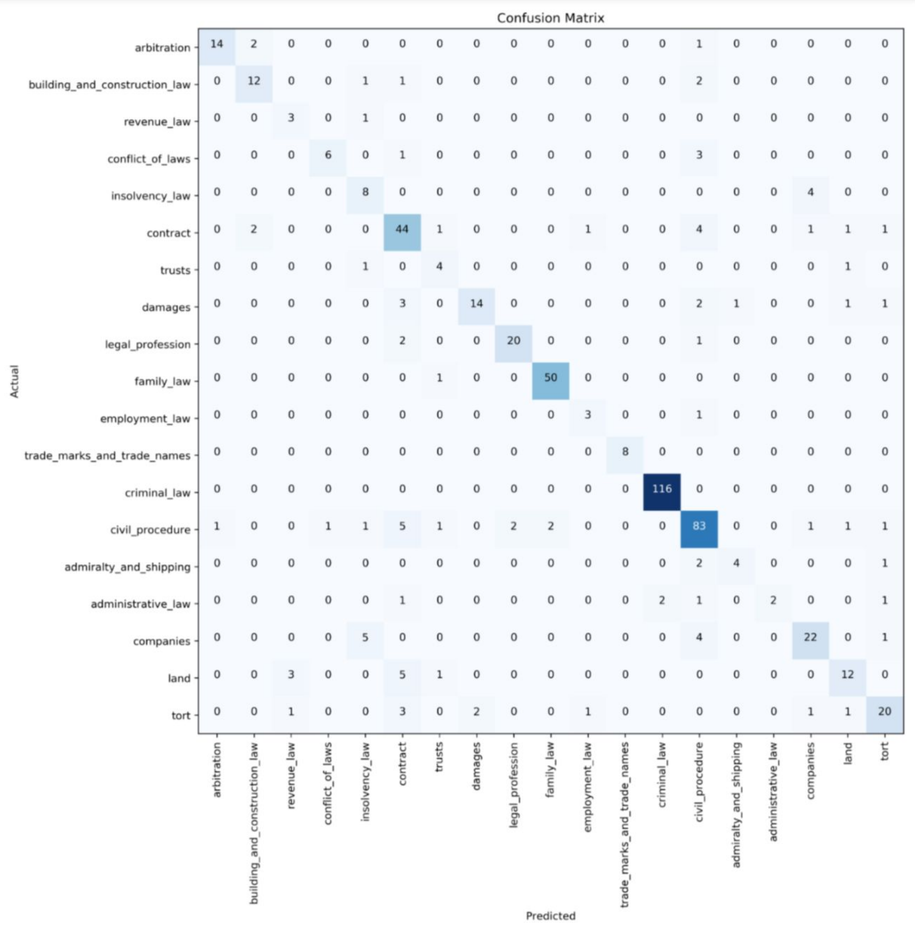
(a) LM pre-training (b) LM fine-tuning (c) Classifier fine-tuning

Results

Accuracy: 82.56% accuracy on 19-way classification task
Time taken to fine-tune LM (GTX 1080 Ti): approx. 3h 40m
Time taken to train classifier (GTX 1080 Ti): approx. 1h+

The model achieved strong results using the default settings in the *fastai* library. The incorrect test set predictions were reasonable errors due to an overlap in subject matter (e.g. Companies vs Insolvency Law, Revenue Law vs Land) and the generality of topics like Civil Procedure and Contract.

It would be useful to explore different levels of fine-tuning and whether multilabel classifier can achieve better results.



NUS National University of Singapore
INFOCOMM MEDIA DEVELOPMENT AUTHORITY
sea connecting the dots
indeed
SGINNOVATE
WING Web Information Retrieval / Natural Language Processing Group