Simple Transformers in Rhetoric Role Labelling for Legal Judgements

B Sai Shridhar[1], S Kayalvizhi[2] and D Thenmozhi[3]

[1]SSN College Of Engineering, Chennai
[2]SSN College Of Engineering, Chennai
[3]SSN College Of Engineering, Chennai

Abstract

Legal case documents follow a common thematic structure with implicit sections like "Facts of the Case", "Issues being discussed", "Arguments given by the parties", etc. These sections are popularly termed as "rhetoric roles". Knowledge of such semantic segments or roles will not only enhance the readability of the documents but also help in downstream tasks like computing document similarity, summarization, etc. The task is, given a legal document, to classify each sentence into 7 rhetoric roles. We compare ALBERT, BERT, RoBERTa and LaBSE for this task. The results show that BERT had the best accuracy at predicting the labels.

Keywords

Legal documents, Rhetoric labels, BERT, ALBERT, RoBERTa, LaBSE

1. Introduction

In countries that follow the common law system (e.g., UK, USA, Canada, Australia, India) there are two primary sources of law- Statutes (established laws, such as the Constitution of a country) and Precedents (prior cases decided in courts of law). Precedents or prior cases help a lawyer understand how the Court has dealt with similar scenarios in the past, and prepare the legal reasoning accordingly. When a lawyer is presented with a situation (that will potentially lead to filing of a case), it will be very beneficial to him/her if there is an automatic system that identifies a set of related prior cases involving similar situations as well as statutes/acts that can be most suited to the purpose in the given situation. Most of the legal case documents follow a common structure with different sections like "Details of the Case", "Issues being discussed", "Arguments given by the parties", etc. These sections are popularly termed as "rhetoric roles". Acquiring such semantic roles will not only improve the readability of the documents but also needed to compute document similarity, summarization, etc. However, this information is generally not specified explicitly in case documents, which are usually just free-flowing text. The task is to semantically label the sentence into one among the seven roles.

2. Related Work

In AILA 2020, the task of labelling rhetoric roles for legal judgments was solved by many authors. In [1] RoBERTa along with Bi-LSTM was used. In [2] they combine TF-IDF features and deep semantic features using BERT. Logistic regression, linear kernel SVM and AdaBoost are used as classifiers. In [3] RoBERTa and a fully connected layer for classification was used. In [4] they experiment with both TF-IDF features and BERT-based features for the task and in [5] they experiment with FastText and TF-IDF from the feature engineering aspect, and Multi-layer perpceptron and Random Forest from the classifier aspect. From [6] we can see that the RoBERTa and BERT transformers seem to be giving better performance.

3. Task & Dataset Description

The AILA 2021 Task 1 training data consists of 60 case documents. In each document the sentences are labelled by one of the 7 categories:

1. Facts: sentences that denote the chronology of events that led to filing the case
2. Ruling by Lower Court: the cases in the dataset were given a preliminary ruling by the lower courts (Tribunal, High Court etc.). These sentences correspond to the ruling/decision given by these lower courts.
3. Argument: sentences that denote the arguments of the contending parties
4. Statute: relevant statute cited
5. Precedent: relevant precedent cited
6. Ratio of the decision: sentences that denote the rationale/reasoning given by the Supreme Court for the final judgement
7. Ruling by Present Court: sentences that denote the final decision given by the Supreme Court for that case document

These documents were manually annotated by legal experts. In addition, we also included the rhetoric labels of the AILA 2021 Task 2 dataset and the AILA 2020 Task 2 dataset. The labels of these datasets were categorical variables so first we converted these variables to ordinal encoded variables. In total we had 12170 text-label pairs to train the model.

4. Proposed Methodology

The task involves the classification of sentences from legal case documents into 7 rhetoric roles. We use the simpletransformers library to import the BERT, RoBERTa, LaBSE and ALBERT transformers which we use as classifiers. We then select the 3 best models and measure the macro F1, precision, and recall values to find the best suited classifier for the task.

4.1. BERT

BERT[7], which stands for Bidirectional Encoder Representations from Transformers, is based on Transformers, a deep learning model in which every output element is connected to every

input element, and the weightings between them are dynamically calculated based upon their connection. BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications.

4.2. ALBERT

The backbone of the ALBERT architecture is similar to BERT in that it uses a transformer encoder (Vaswani et al., 2017) with GELU nonlinearities (Hendrycks & Gimpel, 2016). The three main differences are:

- Splitting the embedding matrix into two smaller matrices using parameterized factorization embeddings
- Cross layer parameter sharing
- Inter sentence Coherence Loss

4.3. RoBERTa

RoBERTa stands for Robustly Optimized BERT Pre-training Approach. optimize the training of BERT architecture in order to take lesser time during pre-training. It has almost similar architecture as compare to BERT, but in order to improve the results on BERT architecture, the authors made some simple design changes in its architecture and training procedure. These changes are:

- Removing the Next Sentence Prediction Objective
- Training with bigger batch sizes and longer sequences
- Dynamically changing the masking pattern

4.4. LaBSE

LaBSE [10] stands for Language-Agnostic BERT Sentence Embedding. The architecture is based on a Bidirectional Dual-Encoder (Guo et. al.) with Additive Margin Softmax (Yang et al.) with improvements.It produces language-agnostic sentence embeddings for more than 100 languages in a single model. The model is trained to generate similar embeddings for bilingual sentence pairs that are translations of each other.

We use these models as Multiclass Classification Models with the default parameters given by simple transformers.

5. Results

The BERT, ALBERT, RoBERTa and LaBSE classifiers are compared using their mcc scores and evaluation loss on an evaluation data set. The results are shown in Table 1. The 3 best runs are

chosen for predicting the rhetoric values for the given test data. The Precision, Recall and Macro F1-Score are calculated class wise and shown in Table 2. The overall score is then calculated and shown in Table 3. From Table 2, we can see that LaBSE has the best precision scores for

Classifier	Epochs	mcc	Evaluation Loss
BERT	5	0.5333	1.5881
ALBERT	5	0.4954	1.1785
RoBERTa	5	0.5323	1.2877
LaBSE	5	0.5396	1.0932

Table 1
Results of Evaluation Set

Classifier	Metric	Argument	Facts	Precedent	Ratio Of Decision	Ruling by Lower Court	Ruling by Current Court	Statute
LaBSE	P	0.2153	0.599	0.3922	0.694	0.07407	0.1667	0.7333
	R	0.7949	0.4812	0.2985	0.4462	0.1333	0.8846	0.7333
	F	0.3388	0.5336	0.339	0.5432	0.09524	0.2805	0.7333
BERT	P	0.5849	0.6122	0.3148	0.6852	0.06383	0.4222	0.4762
	R	0.7949	0.5021	0.5075	0.5927	0.2000	0.7308	0.6667
	F	0.6739	0.5517	0.3886	0.6356	0.09677	0.5352	0.5556
RoBERTa	P	0.5172	0.6243	0.3200	0.6642	0.1034	0.2857	0.5500
	R	0.7692	0.4728	0.3582	0.6201	0.2000	0.8462	0.7333
	F	0.6186	0.5381	0.3380	0.6416	0.1364	0.4272	0.6286

Table 2
Class wise Results

Classifier	Precision	Recall	Macro F1-Score
BERT (Run 2)	0.451	0.571	0.491
RoBERTa (Run 1)	0.438	0.571	0.475
LaBSE (Run 3)	0.411	0.539	0.409

Table 3
Results of Test data

precedent and ratio of decision and has the best scores for predicting statutes. The LaBSE model does well in classifying statutes but poorly in classifying arguments whereas BERT was the best at classifying Arguments and the worst at Statutes. RoBERTa performs the best at ratio of decision and ruling by lower court classes. It also has highest precision with Facts. BERT performs consistently in most class predictions and has the best overall scores as shown in Table 3. All the classifiers struggle in the task of classifying decisions by lower courts with the highest F-Score being only 0.1346.

6. Conclusion

The task given was to semantically label the sentences in a legal document into 7 rhetoric roles. Previously, the BERT and RoBERTa transformers, used as classifiers, produced the best results. In this paper we use the simpletransformers library and import the classifiers BERT, ALBERT, RoBERTa and LaBSE. First, we compare the runs of all 4 models and the 3 best performing models were chosen to predict the test data set. In the test data set the BERT classifier performed the best with RoBERTa being a close second. LaBSE outperformed the other two in predicting statutes but performed significantly worse in classifying arguments and Ruling by Current Court. The proposed method can be further improved by trying DynaBERT and ConvBERT transformers [11].

Acknowledgments

References

[1] Majumder, S. B., & Das, D. (2020). rhetoric Role Labelling for Legal Judgements Using RoBERTa. In FIRE (Working Notes) (pp. 22-25).

[2] Gaoa, J., Ninga, H., Han, Z., Kongb, L., & Qib, H. (2020). Legal text classification model based on text statistical features and deep semantic features.

[3] Jain, R., Agarwal, A., & Sharma, Y. (2020). Spectre@ AILA-FIRE2020: Supervised rhetoric Role Labeling for Legal Judgments using Transformers. In FIRE (Working Notes) (pp. 66-70).

[4] Wu, M., Wu, Z., Wang, X., & Han, Z. (2020). Retrieval Model and Classification Model for AILA2020. In FIRE (Working Notes) (pp. 82-86).

[5] Balaji, N. N. A., Bharathi, B., & Bhuvana, J. (2020). Legal Information Retrieval and rhetoric Role Labelling for Legal Judgements. In FIRE (Working Notes) (pp. 26-30).

[6] Bhattacharya, P., Ghosh, K., Ghosh, S., Pal, A., Mehta, P., Bhattacharya, A., & Majumder, P. (2020). Overview of the FIRE 2020 AILA Track: Artificial Intelligence for Legal Assistance. In FIRE (Working Notes) (pp. 1-11).

[7] https://searchenterpriseai.techtarget.com/definition/BERT-language-model

[8] https://huggingface.co/transformers/model_doc/albert.html

[9] https://huggingface.co/transformers/model_doc/RoBERTa.html

[10] https://towardsdatascience.com/labse-language-agnostic-bert-sentence-embedding-by-google-ai-531f677d775f

[11] https://towardsdatascience.com/advancing-over-bert-bigbird-convbert-dynabert-bca78a45629c

[12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Advances in neural information processing systems, pages 5998–6008.

[13] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy,

Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized BERT pretraining approach. CoRR, abs/1907.11692, 2019.

[14] V. Parikh, U. Bhattacharya, P. Mehta, A. Bandyopadhyay, P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, A. Bhattacharya., P. Majumder, Overview of the third shared task on Artificial Intelligence for Legal Assistance at Fire 2021. In FIRE (Working Notes) - Forum for Information Retrieval Evaluation, India, December 13-17, 2021.

V. Parikh, U. Bhattacharya, P. Mehta, A. Bandyopadhyay, P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, A. Bhattacharya., P. Majumder, FIRE 2021 AILA track: Artificial intelligence for legal assistance. In Proc. of FIRE 2021 - 13th Forum for Information Retrieval Evaluation, India, December 13-17, 2021.

[15] P. Bhattacharya, P. Mehta, K. Ghosh, S. Ghosh, A. Pal, A. Bhattacharya., P. Majumder, Overview of the Fire 2020 AILA track: Artificial Intelligence for Legal Assistance. In FIRE (Working Notes) - Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020.

P. Bhattacharya, S. Paul, K. Ghosh, S. Ghosh, A. Wyner, Identification of Rhetorical Roles of Sentences in Indian Legal Judgments. In Proc. of JURIX 2019 - International Conference on Legal Knowledge and Information Systems, 2019.

[16] V. Parikh, V. Mathur, P. Mehta, N. Mittal, P. Majumder, LawSum: A weakly supervised approach for Indian Legal Document Summarization. arXiv preprint arXiv:2110.01188v3.