Rhetorical role labelling for legal judgements and Legal document summarization

Siddhartha Rusiya[1], Aditya Sharma[1], Debajyoti Debbarma[1] and Samarjit Debbarma[1]

[1]National Institute of Technology,Agartala

Abstract

As legal documents are usually not well structured especially in case of judiciary it is a very difficult task for lawyers or legal advisors to proceed ahead with the case smoothly. Automatically classifying the sentences into different labels can help the cases to proceed more smoothly within less time as the facts and other factors of the case in the document will already be mentioned beforehand.It is the same in case of summarizing of legal case documents. Rhetorical role classification and summarizing of legal case documents is a very difficult task as the documents are not well structured as mentioned above. For summarizing, we need to know about significance of a sentence in the judgement. In this paper,we address both the tasks for the documents provided by AILA(Artificial Intelligence for Legal Assistance). In this paper for both the tasks of rhetorical role labelling and summarizing of legal documents we used BERT model to train and test the dataset provided.For this task researchers used various models like conditional random fields(CRF),GCN,Bi-LSTM,BERT,etc,in which we found that BERT is the most used and also consistently performing model among all the models that compels us to used BERT in our task.Many prior works also used handcrafted features to perform the tasks while in our work we used deep learning approaches as we found that deep learning approaches performs better and are more accurate than the traditional handcrafted features.

Keywords

Natural language processing, text classification, bidirectional encoder representations from transformer, neural networks, language mode, rhetorical role

1. Introduction and Background

Text classification is one of the most common problems of NLP, which targets to assign labels or tags to textual data such as sentences, queries, paragraphs, and documents. It has a wide range of applications including question answering, spam detection, sentiment analysis, news categorization, user intent classification, content moderation, and so on.

In task 1 we are to classify each sentence in the document in one of the 7 semantic segments rhetorical roles given below:-

- `Facts`: This refers to the occurrences of events that led to filing of the case.
- `Ruling by Lower Court`:Here,the documents given were from Indian courts.This refers to the judgements given by the previous courts before being presented in the current court.

Forum for Information Retrieval Evaluation, December 13-17, 2021, India

- `Arguments`: This refers to the sentences that denote the arguments of the contending parties.
- `statute`:This refers to the relevant statute cited in the documents. A statute is a formal written enactment of a legislative authority that governs the legal entities of a city, state, or country by way of consent.
- `Precedent`: This refers to a statement of law found in decision of the superior court. Such decisions are binding to that court and the inferior courts have to follow. The cases based on similar set of facts decided by a court may arise in any future case
- `Ratio of the decision`: This refers to the sentences that denote the rationale/reasoning given by the Supreme Court for the final judgement
- `Ruling by Present Court`: This refers to the sentences that denote the final decision given by the Supreme Court for that case document.

For task 2 the objective was to create a summary of the given judgements. Task2 is divided into two parts. First part is finding of significance of a sentences in a given judgement. Second part is to make a summary of judgements by considering significant sentences.

In solving one of our problem, we are focussed on labeling courts verdicts to corresponding labels, that comes under the question answering and for solving another problem, we are finding the significance of every sentence in generating the summary.

Approaches to automatic text classification can be grouped into two categories:
• Rule-based methods
• Machine learning (data-driven) based methods

Rule-based methods are used to classify text into different categories using a pre-defined set of rules, and require a deep learning knowledge. On the other hand, machine learning based approaches learn to classify text based on observations of dataset which is the training dataset. Using training data, a machine learning algorithm learns inherent relations between texts and their labels.

For solving both of our problems, we focused on rule based methods. After observing our dataset, we observe to make an close to accurate labels approximation, we have to use some deep learning based method. After careful observation, BERT(Bidirectional Encoder Representations from Transformers) seems to be the good option. BERT is based on Transformers, a deep learning model in which every input element makes a connection to every output element, and the weightings between them are dynamically calculated based upon their connection.

2. Related works

In this section we discuss prior works related to rhetorical role labelling,summarization of documents and use of machine learning or deep learning in legal domain. Many of the rhetorical role labelling of sentences were previously mostly done with handcrafted features,while in today's world mostly used of deep learning is preferred incase of task like rhetorical role labelling of sentences[1],where human annotators were also used to annotate the documents.A better and deep understanding of annotation study and curation of a gold standard corpus for

the task of sentence labelling can be found in legal-case-annotation [2]. There were already numerous attempts on automatic role labelling of sentences on legal documents.The initial idea behind the rhetorical role labelling of sentences were to summarize the documents from legal domains to make it easier for the legal person without reading the whole document. various technologies were used for rhetorical role labelling of sentences like Condition Random Fields(CRF),BERT,etc.One work where CRF features were used was segmenting u.s court decisions into functional and issue secific parts[3],handcrafted feature was used in segmenting the us court documents along with CRF. In this paper we uses deep learning approaches where no handcrafted is needed.Deep learning approaches are widely being used in legal domain with the progress of time. Here are some more prior works related to the rhetorical role labelling,summarization of documents and use of machine learning or deep learning in legal domain:- AILA 2021 Overview Paper and Extended abstract[4][5] Semi supervised Training[6],Textual legal case elements [7],AILA 2021 [8],canadian immigration case [9],conditional random fields [10],legal document clustering [11],Dynamic pairwise attention [12],Automatic classification [13],crime classification [14], citation[15],Text Classification[16]

3. Dataset

The datasets used for both the tasks were provided by the AILA(Artificial Intelligence for Legal Assistance).The documents provided were based on supreme court of India. References for Task 1 Dataset[17][18] & Task 2 Dataset[19] Different sets of training and test datasets were provided for both the tasks which consisted of multiple documents.The documents were divided into multiple sentences in both task 1 and 2.The sentences were later classified into the seven rhetorical roles in the task 1 while in task two the important sentences representing the important facts and arguments of the cases were merge back inorder to create a summary.The datasets were later balanced as to give equal priority to each class in laymen terms. After balancing the datasets the sentences classified were as follows: Ratio of the decision-2500, Facts-2500, Precedent-1764, Argument-939, Statute-902, Ruling by Lower Court-483, Ruling by Present Court-341.

3.1. Preprocessing

As we know preprocessing a large document is a very challenging task as it contains many gratuitous words.For preprocessing of the given documents in both task 1 and task 2 we used various libraries. In both the task the documents were split into sentences. After that, remove all the stopwords, commas, name of months, numbers etc. After that do tokenization followed by lemmatization of words. There were total of 11285 sentences in task 1 which in turn becomes 9429 sentences after balancing the dataset in which every sentences were classified into the seven rhetorical roles as required by the task. The number of sentences were counted using value count from pandas library.

4. Experiments

4.1. BERT and GCN model

To perform both the tasks given by the Artificial Intelligence for Legal Assistace(AILA) team,we used BERT and GCN model. As per the research by our members BERT is one of the best performing pre-trained models for NLP learning representation. BERT has also been used in many tasks which are similar to the tasks we are addressing in this paper i.e Rhetorical Role Labeling for Legal Judgements and Summarization of Legal Documents.with the above fact we expect that through pre-trained BERT we will achieve a high performance in both the tasks. legal data such as papers also contains various metadata,in addition to textual data. The GCN model was used for representing the various rhetorical roles in the documents provided and also to extract a learning representation of them.

A textual encoder was constructed to extract textual embeddings from the documents,using BERT and also an encoder was made to checks the importants sentences and match the rhetorical roles. The encoders were pre-trained with contextual data,and important datas were extracted from the documents.The data is inserted into the pre-trained models and concatenated embeddings are calculated by each encoder. Later the softmax output layer is generated and cross entropy is adopted(function loss for training) after passing the concatenatedvectors to a feedforward neural net.

The proposed model structure was referenced from the baseline of CACR[20].Both the encoders mentioned above were present in CACR.CACR demonstrates the performance of SOTA as the most recent context-aware citation recommendation model using AAN dataset and LSTM model.Our model constructed the encoders with GCN solely using the given documents information.

Using the citation relationships between papers as in put values linking a prediction with the GCN-based variational Graph Auto Encoders model VGAE[21],the citation encoder conducts unsupervised learning for extracting the sentences. The model returns the relational learning representation as the embedding vector whenever the document information is used as input to a pre-trained GCN.

It has always been difficult to extract proper and important sentences from a document using Natural language processing.As in this task we are using the documents from supreme court it is harder to extract the sentences into different rhetorical roles and also to summarized it based on important sentences in the documents due to the unstructured nature of the legal documents.

In our tasks, we tested our model with a series of different hyperparameters and found that our best NN systems use 128 units for the RNNs and 128 filters for each of the convolutional layers in the CNN. For both these settings, we tried values of 32, 64, 128 and 256 and 128 gave the best results. Basically, the 128 gave better results than the lower settings and it turned out that the 256 setting could not be run effectively when training with an Nvidia GPU. It seems that additional GPU memory would be required (or a more efficient algorithm) to use 256 units. It is probable that 128 is simply the largest (power of 2) setting that is practical to use given the available equipment. This seems to be supported by the fact that many other NN systems use a value around 100. Additionally, each of the models are regularized with a dropout, which works

by ”dropping out” a proportion p of hidden units during training. We found that a dropout of 0.5 before the final dense layer and batch size of 32 worked best for the LSTM, GRU, and CNN. We also found that the Adam optimizer worked best for both the for CNN and RNN networks.

5. Result and Discussions

For classification in both the tasks we use BERT with some optimization like adding hyperparameters for building our classifier model and also do class weight balancing. The overall precision, recall and F-score for task 1 are 0.192, 0.220 and 0.179 respectively. The overall precision, recall and F-score for task 2a are 0.38, 0.5 and 0.43 respectively.The overall rouge scores AVERAGE R, AVERAGE P and AVERAGE F for task 2b are 0.176, 0.15 and 0.15 respectively. The category wise precision, recall and F-score for Task 1 are given below.

Label	Precision	Recall	F-Score
Arguments	0.119	0.183	0.143
Facts	0.369	0.444	0.403
Precedent	0.128	0.299	0.179
Ratio of the decision	0.590	0.224	0.325
Ruling by Lower Court	0.000	0.000	0.000
Ruling by Present Court	0.089	0.192	0.122
Statue	0.049	0.200	0.079
Overall	0.192	0.220	0.179

6. Conclusion

In this paper we discussed about the automatic role labelling of sentences and summarization of legal documents.The documents provided by the Artificial Intelligence for Legal Assistance team were from Supreme Court of India.The main goal of the first task was to classify the sentences into seven rhetorical roles which are mainly facts(sentences that denote the chronology of events that led to filing of the case),Ruling by lower court,Arguments,Statute,Precedence,Ratio of decision,Ruling by the present court.
The second task is mainly about the summarization of the legal documents which consisted of 2 substasks where task 2a is to ”Identify ’summary-worthy’ sentences in a court judgement i.e we are required to extract only the sentences that marks important decisions,facts argument,etc in the documents,on the other hand for task 2b we are to automatically generate summary either extractive or abstractive.

Automatic classification of sentences and summarization of documents are both difficult tasks as indian legal documents are not very well structured.The main goal of the given tasks are to

make it easier for the legally engaged individual to understand the court documents for the cases the person is handling which also saves a lot of time. For both the tasks given we firstly divided the documents into multiple sentences and perform basic preprocessing operations such as stemming,lemmatization,tokenization,etc.We used BERT and GCN model to train and tests our datasets.Prior works related to the tasks were mostly using traditional handcrafted features while in this paper we used advanced deep learning model.Deep learning models performs better than the traditional handcrafted features.Among all the deep learning features we found that BERT is the most consistent performing and mostly used model for the automatic classification of tasks.By using BERT we have achieved the required result and has successfully completed the tasks. Lastly,we surveyed many deep learning models, which are developed in the past and have significantly improved state of the art on various classification tasks. We provide description of both the tasks, and present a quantitative analysis of the performance of these models on several public benchmarks.

Acknowledgments

The authors would like to thank all the anonymous reviewers for reviewing this work and also AILA for providing this opportunity.

References

[1] Paheli Bhattacharya, Shounak Paul, Kripabandhu Ghosh,Saptarshi Ghosh, and Adam Wyner, "identification of rhetorical roles of sentences in indian legal judgments", 2019. Https://arxiv.org/abs/1911.05405.

[2] A. Z. Wyner, W. Peters, and D. Katz, "a case study on legal case annotation,", 2013. Http://jurix2013.cirsfid.unibo.it/wp-content/uploads/2013/05/WynerJURIX2013.pdf.

[3] Jaromír Šavelka, Kevin D. Ashley, "segmenting u.s. court decisions into functional and issue specific parts", 2018. Https://ebooks.iospress.nl/volumearticle/50840.

[4] V. Parikh, U. Bhattacharya, P. Mehta, B. Ayan, 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, 2021.

[5] V. Parikh, U. Bhattacharya, P. Mehta, B. Ayan, P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, A. Bhattacharya, P. Majumder, Fire 2021 aila track: Artificial intelligence for legal assistance, 2021.

[6] Isar Nejadgholi, Renaud Bougueng, Samuel Witherspoon, "a semi-supervised training method for semantic search of legal facts in canadian immigration cases", 2017. Https://ebooks.iospress.nl/volumearticle/50840.

[7] Adam Wyner, "towards annotating and extracting textual legal case elements", 2010. Http://ceur-ws.org/Vol-605/paper1.pdf.

[8] Paheli Bhattacharya, Parth Mehta, Kripabandhu Ghosh, Saptarshi Ghosh,Arindam Pal , Arnab Bhattacharya and Prasenjit Majumder, "overview of the fire 2020 aila track: Artificial intelligence for legal assistance", 2020. Http://ceur-ws.org/Vol-2826/T1-1.pdf.

[9] Isar Nejadgholi, Renaud Bougueng, Samuel Witherspoon, ”a semi-supervised training method for semantic search of legal facts in canadian immigration cases”, 2017. Https://ebooks.iospress.nl/volumearticle/48054.

[10] John Lafferty, Andrew McCallum, Fernando C.N. Pereira, ”conditional random fields: Probabilistic models for segmenting and labeling sequence data”, 2001. Https://dl.acm.org/doi/10.5555/645530.655813.

[11] Qiang Lu, William Keenan, Jack G. Conrad, Khalid Al-Kofahi, ”legal document clustering with built-in topic segmentation”, 2011. Https://dl.acm.org/doi/pdf/10.1145/2063576.2063636.

[12] Jaromír Šavelka, Kevin D. Ashley, ”modeling dynamic pairwise attention for crime classification over legal articles”, 2018. Https://dl.acm.org/doi/10.1145/3209978.3210057.

[13] Vern R. Walker,Krishnan Pillaipakkamnatt,Alexandra M. Davidson,Marysa Linares,Domenick J.Pesce, ”automatic classification of rhetorical roles for sentences”, 2019. Http://ceur-ws.org/Vol-2385/paper1.pdf.

[14] Pengfei Wang,Yu Fan ,Shuzi Niu ,Ze Yang ,Yongfeng Zhang ,Jiafeng Guo, ”hierarchical matching network for crime classification”, 2019. Http://www.bigdatalab.ac.cn/ gjf/papers/2019/SIGIR-crime.pdf.

[15] Chanwoo Jeong , Sion Jang , Hyuna Shin , Eunjeong Park , Sungchul Choi, ”a context-aware citation recommendation model with bert and graph convolutional networks”, 2019. Https://arxiv.org/abs/1903.06464.

[16] Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, Jianfeng Gao, ”deep learning based text classification: A comprehensive review”, 2021. Https://arxiv.org/pdf/2004.03705.pdf.

[17] 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, 2020.

[18] P. Bhattacharya, S. Paul, K. Ghosh, S. Ghosh, A. Wyner, Identification of rhetorical roles of sentences in indian legal judgments, 2019.

[19] V. Parikh, V. Mathur, P. Mehta, N. Mittal, P. Majumder, Lawsum: A weakly supervised approach for indian legal document summarization, 2021.

[20] L. Yang, Y. Zheng, X. Cai, H. Dai, D. Mu, L. Guo, and T. Dai., ”cacr”, 2018. Https://arxiv.org/pdf/1903.06464.

[21] Thomas N Kipf, and Max Welling., ”variational graph auto encoders”, 2016. Https://arxiv.org/abs/1611.07308.