Critical Information Extraction from Terms of Services Document

Aditya Ashok Dave
Univ. of Southern California
daveadit@usc.edu

Akanksha Sanjay Nogaja
Univ. of Southern California
nogaja@usc.edu

Lavina Lavakumar Agarwal
Univ. of Southern California
llagarwa@usc.edu

Shreya Venkatesh Prabhu
Univ. of Southern California
prabhus@usc.edu

Sai Sree Yoshitha Akunuri
Univ. of Southern California
akunuri@usc.edu

Abstract

Terms of Services (ToS) are legal agreements between users and service providers. In order for the user to consume any service they must accept the terms. However, since ToS documents are very verbose and use a very opaque jargon, users tend to acknowledge them without fully understanding the agreement. This can lead to the user signing obligations which they might not be willing to in reality, or might be exposed to unfair terms and practices. The proposed idea is to make user more informed about the unfairness of the clauses in ToS and also present the obligations imposed by it.

1 Project Domain & Goals

Terms of Services (ToS) is a set of clauses that act as a contract between a service provider and a consumer. ToS are very common and needs to be accepted in order to proceed further and consume a service on the Internet. According to a study (McDonald AM and Cranor LF, 2008), the annual time estimate for a frequent online service user to skim through the terms was found to be 80 hours and, 200 hours to read through the entire document. The solution being proposed is to have an automated system that will extract critical information and present it to users thereby making the process of reading ToS easier.

2 Related Work

CLAUDETTE (Marco Lippi et al., 2018) is an experimental study to detect the potentially unfair clauses from ToS of online platforms using machine learning. 7 models (SVM, CNN, LSTM and HMM with various combinations of features) were tested on ToS dataset and the accuracy metrics (Precision, Recall and F1 score) were reported. While most of the models were sentence level, SVM with HMM was used to build a collective classification algorithm to include sequence of sentences. Top 5 models among these were picked to create an ensemble which reported the best F1 score of 0.806. Another research (Guarino A. et al., 2021) found promising results using Universal Sentence Encoder with SVM for classification with F1 score of 0.87.

In (L. Zheng et al., 2021), the authors have spoken about whether pre-training the neural network model is useful for law and legal dataset processing. For this, they compared BERT base model performance on ToS dataset with BERT Double, a BERT model which was trained for extra 1M iterations and Custom Legal-BERT, a BERT model trained from scratch on Harvard law case corpus. They found that BERT Double, with 77.3% F1 score and Custom Legal-BERT, with 78.7% F1 score, outperformed BERT base model and also the highest performing model from Claudette for the general setting of Terms of Service by 0.4% and 1.8% respectively.

(Vikas and Roshann, 2020), which forms the basis for finding obligatory terms, uses transfer learning to create a secondary dataset using Named Entity Recognition in English legal documents. For extracting entities, the authors have used libraries such as FlairNLP, AllenNLP, BERT, Deeppalvo on the Entity-Recognition-Datasets. They found that FlairNLP, being trained on multilingual text, outperforms AllenNLP followed by BERT.

3 Datasets

3.1 ToS Data

ToS dataset was created as a part of Claudette (Marco Lippi et al., 2018) experimental study. The ToS clauses in the dataset are annotated for 8 categories of unfairness using XML mark-up. The categories are documented in Table

1. Each of these categories are further subcategorized into 3 levels: clearly fair, potentially unfair and clearly unfair, giving us 24 classes. Of these 24, annotated data is available for 18 categories and the authors have made it publicly available: `http://claudette.eui.eu/ToS.zip`. All the clauses along with their categories must be extracted from these XML files.

Category	Annotation
Arbitration	<a>
Unilateral Change	<ch>
Content Removal	<cr>
Jurisdiction	<j>
Choice of Law	<law>
Limitation of Liability	<ltd>
Unilateral Termination	<ter>
Contract by Using	<use>

Table 1: ToS clauses categories

3.2 Data Preprocessing

The extracted ToS clauses require preprocessing and feature extraction in order to be fed to transformer models. The dataset will be tokenized into sub words using the WordPiece algorithm which is based on byte pair encoding (BPE). It is an iterative greedy algorithm which looks at the most commonly occurring pair of bytes and merges them into one. In addition to WordPiece embeddings, BERT (Jacob Devlin et al., 2019) and RoBERTa (Yinhan Liu et al., 2019) uses positional embedding to track the position of words in the input and segment embedding if the input clause is a collection of multiple sentences. The input to the model will be the sum of all of these embedding.

For the transformer model XLNet (Zhilin Yang et al., 2019), the input will be generated by using the SentencePiece algorithm. SentencePiece, unlike WordPiece, does not assume that the words in sentences are separated by white spaces. It instead treats white spaces as characters to process and uses BPE and unigram language model. It can be implemented using XLNetTokenizer.

In order to perform named entity recognition to identify clauses related to user obligations, we will be performing the following data preprocessing steps - tokenization, lemmatization and parts of speech tagging. Word2Vec will be used for input features generation.

4 Technical Challenges

4.1 Design

In (L. Zheng et al., 2021), authors claimed that BERT double and Custom Legal-BERT outperformed the baseline model CLAUDETTE for classifying terms by fairness. Inspired by this, we would like to perform more research on advanced transformer models such as RoBERTa and XLNet. One of the technical challenges would be to pre-train and fine tune these models using Masked Language Modelling and Next Sentence Prediction techniques. This will be done on Harvard law corpus (`https://case.law/`) to get our models familiarized with vocabulary used in legal documents. The performance of these models will then be tested on the ToS dataset for any significant gains.

While prior work exists for fairness classification of ToS, it has not been explicitly taken up yet to return obligatory clauses from ToS for the individuals involved. For this purpose, we will extract important obligatory entities from ToS using named entity recognition, using spaCy and NLTK. To identify obligatory clauses, we will use Word2Vec features and find clauses that express similarity to the word ”obligation” like responsible, required, commit, oblige etc. These clauses will then be mapped to obligatory entities and will be returned as critical information. After examining the performance of this approach, we would like to explore complex transfer learning based models as mentioned in (Vikas and Roshann, 2020).

4.2 Evaluation Metric

F1 score will be used for benchmarking the above-mentioned models with the baseline models.

5 Division of Labor

Various tasks involved in the proposed project and division of labour among the team members are documented in Table 2.

References

[McDonald AM and Cranor LF2008] McDonald AM, Cranor LF. 2008. *The cost of reading privacy policies*. Isjlp 4:543.

[Marco Lippi et al.2018] Marco Lippi, Przemyslaw Palka, Giuseppe Contissa, Francesca Lagioia, Hans-Wolfgang Micklitz, Giovanni Sartor, Paolo

Task	Assignees
Try baseline benchmarks	All
Preprocessing-Obligations	Yoshitha, Aditya
Preprocessing-Fairness	Shreya, Akanksha
Named Entity Recognition for Obligations	Shreya, Lavina, Aditya
Fairness Classification Models	Yoshitha, Lavina, Akanksha
Documentation and Presentation	All

Table 2: Division of labor among Team members

Torroni. 2018. *CLAUDETTE: an Automated Detector of Potentially Unfair Clauses in Online Terms of Service*. Artificial Intelligence and Law.

[Guarino A. et al.2021] Guarino A., Lettieri N., Malandrino D. et al. 2021. *A machine learning-based approach to identify unlawful practices in online terms of service: analysis, implementation and evaluation.* Neural Comput & Applic 33, 17569–17587

[L. Zheng et al.2021] L. Zheng, N. Guha, B. R. Anderson, P. Henderson, and D. E. Ho 2021. *When does pretraining help? Assessing Self-Supervised Learning for Law and the CaseHOLD Dataset.* Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law

[Vikas and Roshann2020] Vikas Mastud and Roshan Jaiswaln. 2020. *Named Entity Recognition on legal text for secondary dataset*

[Jacob Devlin et al.2019] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova 2019. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.*. https://aclanthology.org/N19-1423

[Yinhan Liu et al.2019] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. *RoBERTa: A Robustly Optimized BERT Pretraining Approach.*. arXiv:1907.11692 [cs.CL]

[Zhilin Yang et al.2019] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. *XLNet: Generalized autoregressive pretraining for language understanding.* . arXiv preprint arXiv:1906.08237.