Contextual Relevance based Significant Component Extraction from Contracts

Anonymous ACL submission

Abstract

both

Automatic extraction of “significant” components of a legal contract, has the potential to simplify the end user’s comprehension. Significance of a component may be defined at an individual contract level and at a contract-type level. A component, sentence or paragraph, may be considered significant at a contract level if it contains contract specific information (CSI), like names, dates or currency terms. At a contract-type level, components which deviate significantly from the norm for the type may be considered significant (type specific information(TSI)). In this paper, we present approaches to extract “significant” components from a contract at both these levels. We attempt to do this by identifying patterns in a pool of documents of the same kind. Here, the idea of “significance” is to include unique or novel details concerning the contract by looking at structural and semantic similarities among a pool of contracts belonging to a specific type. To the best of our knowledge, no such work is prior attempted. In our approach, the solution is formulated in two parts: *identifying CSI* using a BERT based contract-specific information extractor and *identifying TSI* by scoring sentences in a contract for their likelihood in the contract type.

extend

1 Introduction

Contracts are agreements between two or more parties, that govern what each party can or cannot do and are usually dense in information. Extracting contract elements and locating novel clauses and assignments from a legal contract is a desired feature by many as it will greatly simplify and accelerate user comprehension. Traditionally, it requires a domain expert as there are parts of a contract that can only be noticed by a reader experienced in reviewing contracts. For an untrained eye, it is often difficult and time consuming to identify rare and unique sentences. Therefore, to lessen the human effort, in this paper we introduce approaches for automatic identification and extraction of significant components of the contract.

Templatised Sentences
This Agreement shall be effective as of *November 5, 2014 (the Effective Date)*.",
CHOICE OF LAW.This Agreement shall be construed and interpreted in accordance with the internal laws of the *State of California*.

Table 1: Sentences with a Template Structure

In contracts, for the sake of ensuring legal unambiguity, the language used is very crisp and most of the contracts are created in a templatized format. On carefully examining the semantics and structure of diverse legal contracts *(employment, software license, purchase, severance)*, we observe that

i) *in contracts of same category*, although the wording and sentence structure differ, the information conveyed remains the same,

ii) *in an individual contract*, each contract has components, sentences or paragraphs that are remarkably distinct in their nature.

Components in an induvidual contract can be broadly classified as:

Templatised sentences: sentences that follow a template, where a phrase or only a part of the sentence may vary and the rest of the content is semantically same across similar contracts. Examples include contract elements (Chalkidis et al., 2017) like title of the contract, parties involved in the contract, dates, governing law. These sentences are of significance to the end-user ~~as they~~ are specific to a contract. As observed in Table 1, the sentences can be generated from a template by filling in relevant information in the slots where the effective date or the state governing the law will be different for different contracts. In templatised sentences the information changes rapidly in each document as the values are unique to each contract.

redundant

Boilerplate sentences[1]: sentences that are standard formulations, and are uniformly found in all contracts of a type. As we can see in Table 2,

they form the bulk of most contracts

Boilerplate Sentences
While employed by the Company hereunder, Executive shall be eligible to participate in the Company 's employee benefit plans as in effect from time to time pursuant to the terms of those employee benefit plans.
No waiver of any breach or condition of this Award Agreement shall be deemed to be a waiver of any other or subsequent breach or condition whether of like or different nature.

Table 2: Sentences with Standardized Clauses

these clauses are standard across contracts of a specific type. Business and technical documents often use boilerplate sentences to improve efficiency and standardize language and structure. The information divergence between contracts of a type is almost constant for boilerplate sentences.

Rare sentences: rare sentences in a contract include content not commonly found in contracts of that type and hence are conspicuous by their presence in the current contract. In Table 3, first

Rare Sentences
To achieve balance, your current tax withholdings may cease and a hypothetical rate of tax may be calculated and withheld from your wages.
No additional Stock Units granted as part of the Award may be earned following the Change in Control.

Table 3: Sentences with Rare Clauses

example refers to hypothetical tax which applies to employees who work at an onsite location. Similarly in the second example, change in control may or may not warrant changes in the stocks policy of the organization. Both these clauses are situational and may not appear in all contracts of that category. Intuitively, these sentences will be of interest to anyone examining the contract because they bring in novelty. Rare sentences in a contract are identified on the basis of contract type to which a contract belongs.

In summary, a contract has -

i) template sentences, which contain contract specific information (CSI) and are generic across contract types.

ii) rare sentences which deviate from other contracts of the same type, they convey type specific information (TSI) and can be recognised only if one has an in depth understanding of the content usually present in the contract type.

iii) common and well understood clauses that constitute boilerplate sentences. In terms of volume, they account for majority of the sentences in a contract.

This approach of extracting significant components does not really qualify as a standard summarization task because there is no merit in summarizing boilerplate sentences which are well understood. Abstractive summarization (Zhang et al., 2020a) techniques would inadvertently change the semantics of the contract. Even when compared to an extractive setting (Nallapati et al., 2017), in this study our main focus is to accentuate rare and contract elements based sentences as significant components in comparison to boilerplate sentences.

The output can be presented in either one of the formats *a) highlighted input document* - where sections of interest are highlighted within the overall contract, *b) a cover-page* - containing consolidated information of practical value. The effectiveness of the automated significant components identification model was further evaluated by conducting an experimental study that compares the performance between human and machine for the task.

2 Related Work

Since language is the core of law and legal contracts, there has been an increased interest in applying natural language processing techniques to a wide range of problems ranging from information extraction to sentence prediction in law (Zhong et al., 2020; Hendrycks et al., 2021; Kalamkar et al., 2021; Zheng et al., 2021). Considerable amount of work has been done in contract analysis and information extraction from contracts (Yang et al., 2013; Silva et al., 2020; Mittal et al., 2015).

The most obvious approach to automatic contract element extraction is to model it as sequence labeling task. Statistical methods like Conditional Random Fields (Finkel et al., 2005; Xu and

[1]The term boilerplate refers to standardized text, copy, documents, methods, or procedures that may be used over again without making major changes to the original.

Sarikaya, 2013) were popular for sequence labeling prior to neural networks. (Chalkidis et al., 2017) involved hand written rules along with hand crafted features to uniquely identify and extract the contract elements. Recently, neural networks (Huang and Xu, 2015; Ma and Hovy, 2016; Chalkidis and Androutsopoulos, 2017) and BERT (Devlin et al., 2019) based approaches (Zhang et al., 2020b; Chen et al., 2019) were developed for sequence labeling and slot with joint intent classification. Our work for CSI extraction closely resembles (Zhang et al., 2020b) where the contract elements are extracted from regulatory filings and property lease agreements using the standard BOI tagging scheme for the contract elements of interest. We include more categories of contracts (employment, incentive, purchase, severance, software-license) and the contract elements are majorly kept consistent.

Scope identification is another popular area of research in legal domain as it is tedious to read legal documents. Contracts or legal documents contain many key sentences, and it sometimes becomes necessary to have domain knowledge regarding that document to avoid missing any. Summarization (Andhale and Bewoor, 2016) is a reliable approach and summarizing legal contracts was attempted (Kubeka and Ade-Ibijola; Kore et al., 2020) by taking the document features and ordering the sentences according to their importance. These techniques do not differentiate between boilerplate sentences which forms the bulk of the contract and the remainder of the contract. Instead of summarizing well understood and accepted clauses, our study we intends to focus on contract specific information (CSI) and contract type specific information (TSI). Classification and hand crafted rules (Le et al., 2020) was another recent approach to precisely identify the scope and was applied to construction contracts to identify requirements automatically.

Regression (Ren et al., 2016; Zopf et al., 2018) is another technique where the sentences are scored on their importance and the model learns to include sentences in a summary based on the scores it predicts. Since our study aims at recognizing TSI, and based on our observations that legal contracts of same category have repetetive information, we devised an approach to we calculate sentence likelihood with respect to the contract type and use these scores to identify TSI. The likelihood scores were calculated using LaBSE(Feng et al., 2020) and BERT (Devlin et al., 2019) was adapted to learn and predict these likelihood scores.

3 Approach

Significant component extraction is accomplished in two stages:

(1) Identifying CSI - by ~~it is responsible for~~ processing each sentence of the document and identifying sentences with contract elements (Chalkidis et al., 2017)

(2) Identifying TSI - it looks at each sentence in the contract and assigns a likelihood score to the sentence. similar

These stages ~~models~~ contribute in effectively identifying the scope of significant components, by automating contract processing and extracting text relating to CSI and TSI from the contracts. we ~~Both models~~ use LEGAL-BERT-BASE (Chalkidis et al., 2020) which is fine-tuned on BERT (Devlin et al., 2019) for legal domain and has shown substantial improvement in challenging downstream tasks like multi-label-classification. Within the wide categories of legal contracts available, we ran our experiments on the contract types mentioned in Table 7.

The overall architecture is shown in Figure 1. The input to the model is a document D containing a set of sentences S. The output is a set of sentences P, that effectively highlight information unique and specific to the document D, such that $P \in S$.

3.1 Identifying CSI

3.1.1 Dataset

This section describes the process employed in creating corpus for identifying CSI through sequence labeling. We sampled 100 legal documents of each category mentioned above. These documents are then pre-processed into paragraphs, as a paragraph as a unit might be of a higher value than an isolated sentence. The documents are split into train , test and validation bins in the ratio 7:2:1. Commonly applicable contract elements are identified and selected as contract elements of interest. Most of the contract elements are phrases rather than a single token, therefore we followed the BIO tagging scheme. We manually annotated the contracts to mark the selected contract elements. <TBD the contract elements are kept consistent across the contract types to reduce the training and annotation effort and to increase robustness and generality in

Pose it as a seq labeling task

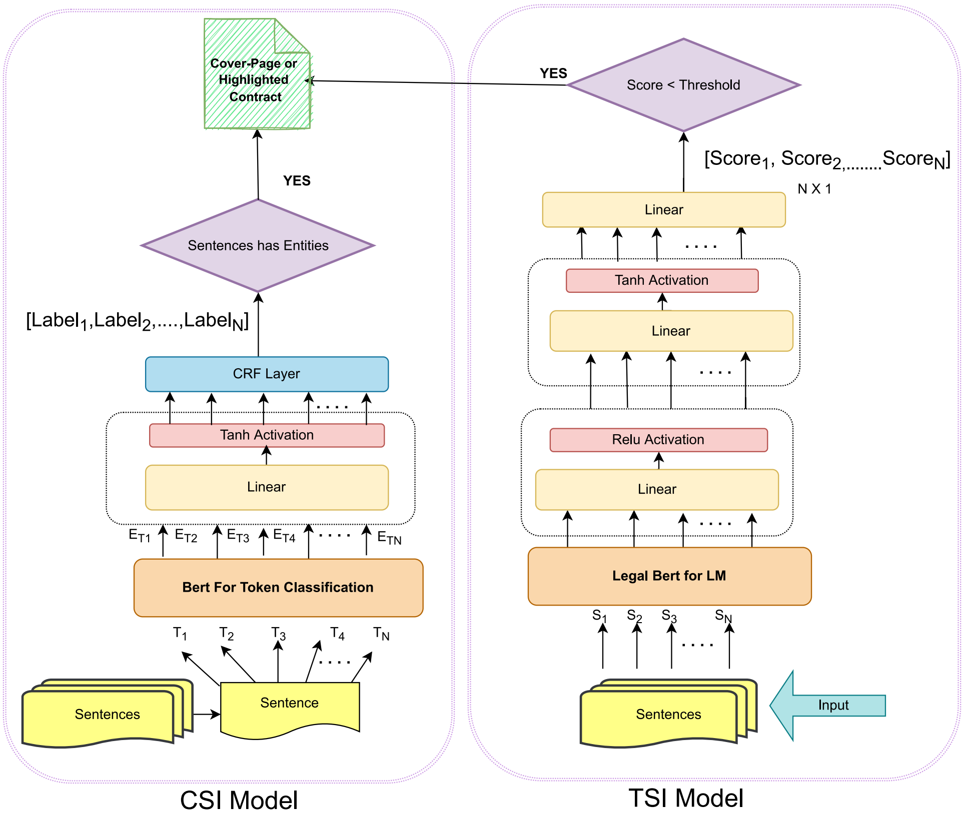

Figure 1: Architecture
<TBD Better decription>

the model.> The contract elements we annotated are listed in Table 4.

3.1.2 Model

(A)

In the CSI model, we trained Bert for ~~token classification model for~~ sequence labeling. It is common for contracts to follow a fixed structure with a certain number of prescribed elements (*contract title, contract parties, effective start date, termination|maturity date, governing law etc.*). <TBD ~~Though these elements are specific to the contract,~~ we model the task as a sequence labeling task as the sentences that contain them have template like structure >.

All the contracts are divided into paragraphs for both training and prediction. The input sequences are tokenized using BERT tokenizer and special tokens [CLS] and [SEP] are added at the beginning and end of the input sequence respectively. All the input sequences are padded to a maximum length of 256 tokens. After passing through BERT, we apply a linear layer and crf layer on top of the hidden states output of the last layer. The model is trained for 25 epochs with learning rate of 1e-05.

(A) In order to identify phrases of interest

3.2 Identifying TSI

3.2.1 Dataset

specific

In contracts that belong to a category, there are lot of similarities in structure as well as semantic content. To confirm this further we sampled 100 documents of each contract type. The average number of sentences for each category is 10,000. For each sentence in a contract we calculated its similarity with all the sentences of all the contracts belonging to that category. The similarity metrics used for analysis were: cosine similarity, dot product, BLEU (Papineni et al., 2002), ROUGE (Lin and Och, 2004) and LaBSE (Feng et al., 2020). Softmax is applied to all the similarity scores obtained and the sentences with maximum similarity scores across documents are grouped. Manual evaluation was performed on the sentences grouped to determine the metric that best captures the similarity among the sentences. After ~~meticulously~~ comparing all the grouped sentences, we concluded that the similarity of sentences captured by LaBSE and ROUGE is closer to how a human would group them.

these contracts

By calculating the similarity scores for a given sentence with respect to all the sentences, of all the documents of that category, we can safely assume that we are calculating the likelihood of occurrence of the sentence given a contract category. Calculating likelihood of all the sentences in a document given a contract type, will help us determine if the sentence is novel or familiar. We *mean pooled* similarity scores obtained by *LaBSE* for a sentence and assigned it as likelihood score for the corresponding sentence. Similarly, we compute likelihood scores for all the sentences in all the samples considered.

3.2.2 Model

In the TSI model we extend BERT (LEGAL-BERT-BASE) for regression. The pre-processing remains same, documents are segmented into paragraphs and tokenized using BERT tokenizer, adding special tokens [CLS] and [SEP] at the beginning and end of the input sequence respectively. The input sequences are padded to a maximum length of 128 tokens. The final hidden states output is passed through linear layers with an activation layer in between for non-linearity. From the last linear layer, a single output is retrieved which serves as the score for the sentence likelihood. The model is trained for 15 epochs with learning rate of 1e-05. The loss criteria is MSE(Means Squared Error) and the objective is to minimise the loss between the predicted scores and the training scores. Pearson Co-relation scores are calculated between the test scores generated by LaBSE and the trained BERT regression model.

4 Analysis and Evaluation

Commonly used named entity recognizers (NER systems) (Nadeau and Sekine, 2007) are different from CSI models described in the study. NER systems typically identifies persons, organizations, dates, locations, currency terms etc., but they cannot be directly applicable to contract elements extraction. For example, a NER system can identify dates and persons but will not be able to differentiate if the date is an effective start date or termination date. Similarly not all instances of persons, organization or location would be contract parties or governing law elements. The sentences that contain these CSI are almost in a template like schema, therefore training a sequence labeling model to understand and extract the specific sentences which contain contract elements yeilds better results.

move to approach

	F1	P	R
ContractParties	0.92	0.89	0.95
ContractTitle	0.81	0.72	0.94
EffectiveDate	0.84	0.80	0.89
GoverningLaw	0.55	0.40	0.86
EmploymentRole	0.42	0.42	0.42
SalaryCompensation	0.49	0.43	0.57
TerminationDate	0.40	0.60	0.30

Table 4: Evaluation of Contract Elements

Contract Element	Frequency in Train Data	Frequency in Test Data
ContractParties	218	62
EmploymentRole	179	52
EffectiveDate	131	32
GoverningLaw	83	22
ContractTitlle	80	15
TerminationDate	38	3
SalaryCompensation	12	2

Table 5: Frequency of Contract Elements in Train and Test data

Table 4 shows micro-averaged metrics F_1, precision and recall across the selected contract elements. By examining these results, we can infer that common elements like ContractTitle, ContractParties, EffectiveDate which occur in all documents are well generalised by the BERT model and so have higher precision and recall values. The precision and recall scores are very low for contract elements like TerminationDate, SalaryCompensation which have not commonly occurred in the test contracts sampled. The primary reason contributing to these low values is that the legal contracts dataset contains contracts as well as amendments made to the contracts. Though we are filtering amendments by considering the document length (considering contracts that have more than 100 sentences), few amendment contracts are still sampled and these may not have all the contract elements that a new contract would mention. The positives from this result is BERT is able to generalise commonly occurring contract elements with samples as low as 100 contracts. For rare contract elements, it requires more data. Table 5 shows the frequencies of the contract elements in both train and test bins after de- duplication. The low representation of TerminationDate and SalaryCompensation samples in the train and test data could also contribute to low precision and accuracy values.

explains

4.1 TSI model

TSI model is used for assigning likelihood scores to the sentences and identifying rare sentences that marks novelty in the contract. We computed likelihood score of a sentence occurring in contracts of a particular contract type. Fig 2 shows the plot for sorted likelihood scores of sentences given a contract type for a sample of 10000 sentences. We observed that the plot remained same across contract types for the contract mentioned in Table 7 under contract types. From the plot we notice two bends in the data and the data can be classified into three classes:

i) likelihood score < 0.5: these sentences map to rare sentences, not normally present in all the contracts of that category.

ii) likelihood score in between 0.5 - 0.7: these sentences map to boiler-plate sentences which uniformly occur in all the contracts with a minor change in wordings or expression.

iii) likelihood score > 0.7: these sentences would map to core sentences in the contracts which contain named entity mentions. Likelihood of templatised sentences was computed using TSI model in two ways-

i) comparing the original unmasked sentences across contracts.

ii) masking all the named entities to their entity type (people's names to PERSON, organization names to ORG) using ner model of spaCy(Honnibal and Montani, 2017). Masking also helps in normalizing the data.

The likelihood score for sentences increased for masked sentences, which proves that semantically the sentence structures are common, but the entity values make them less probable and therefore specific to the contract. Table 6 where 'lscore' refers to the likelihood score, identifies few examples where masking entities has shown impact on the sentence likelihood scores.

<TBD Should the graph have title>

Since masking the sentences normalizes the data, all the named entities in the contracts were masked. We also observed that contract elements are different from named entities, therefore we can divide the data into two classes instead of three- i)Rare sentences which have likelihood score less than 0.5 ii)familiar sentences which have likelihood score above 0.5. On unseen data the TSI model will score the sentences given the contract type, and sentences having having score below the threshold

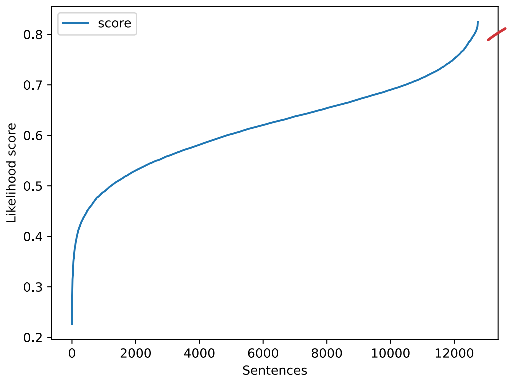

Figure 2: Sorted Likelihood Scores of Sentences

are highlighted in the contract.

4.1.1 Evaluation

On the 100 contracts sampled for each contract type, we preformed K-fold validation, with k=3. The contracts were split into train,test and validation bins in the ratio 7:2:1. The TSI model was trained to predict the likelihood score of the sentences in the test bin.

$$C = \frac{(N\sum_{i=1}^{N} T_i P_i - (\sum_{i=1}^{N} T_i)(\sum_{i=1}^{N} P_i))}{\sqrt{N\sum_{i=1}^{N} {T_i}^2 - (\sum_{i=1}^{N} T_i)^2}\sqrt{N\sum_{i=1}^{N} {P_i}^2 - (\sum_{i=1}^{N} P_i)^2}} \quad (1)$$

To measure the performance of our proposed model in predicting the likelihood score, we compute the Pearson product-moment correlation (C) (Benesty et al., 2009) between likelihood scores computed by maean pooling LaBSE similarity scores (T) and likelihood scores generated by the TSI model (P), for a sample of 10000 sentences (N) using (Equation 1). Pearson correlation estimates the degree of statistical relationship between two independent variables.

A high positive correlation between the actual and predicted values implies that the model can be trusted to work reasonably well on new unseen contracts of that category. Table 7 has the Pearson correlation scores averaged for K-fold data sets on contract types considered. The high Pearson correlation values instills confidence that the model can identify rare sentences with a reasonable accuracy.

To further access the effectiveness of the TSI model we conducted an experimental study that compares the performance of the machine and hu-

Original Sentence	lscore	Entities Masked Sentence	lscore
EX-10.8 4 a17-1046_1 EX-10.8 EXHIBIT 10.8 EMPLOYMENT AGREEMENT This EMPLOYMENT AGREEMENT (the Agreement) is entered into and effective as of this 3rd day of March	0.75	EX-10.8 4 a17-1046_1 EX-10.8 EXHIBIT10.8 EMPLOYMENT AGREEMENT This EMPLOYMENT AGREEMENT (the Agreement) is entered into and effective as of this DATE DATE DATE (the Effective Date)	0.86
Term of this Agreement. The Term of this Agreement shall mean the period commencing on the Effective Date and ending on March 31	0.64	Term of this Agreement . The Term of this Agreement shall mean the period commencing on the Effective Date and ending on DATE DATE.	0.88

Table 6: Sentence Likelihood Scores (lscore) with and without Masking Entities

Contract Type	Pearson Correlation on Kfold
Employment	0.996
Incentive	0.998
Severance	0.990
Software License	0.997
Purchase	0.987

Table 7: Averaged K-Fold Validation for Pearson Correlation of test and predicted likelihood scores

	Annotator 1			**Annotator 2**		
Contract Type	P	R	F1	P	R	F1
Employment	0.94	0.94	0.94	0.88	0.93	0.91
Incentive	0.	0.	0.	0.	0.	0.
Severance	0.90	0.99	0.95	0.84	0.99	0.90
Software License	0.89	0.99	0.93	0.83	0.92	0.87
Purchase	0.	0.	0.	0.	0.	0.

Table 8: Human Evaluation Statistics

man. Assigning likelihood score will be difficult for a human, therefore labels were assigned to the sentences in the test set based on their likelihood scores. If the sentence score is below the threshold set for rare sentences, then the sentences are labeled rare (0) . If the sentence score is above the threshold (set for rare sentences), then it is labeled familiar (1) . We had two annotators, the first annotator was provided with 15 contracts of each contract type and the respective test sentences set. The second annotator was provided with 5 contracts of each contract type and the test sentences set. Both the annotators were asked to read the contract set provided to them and then label the test sentences as rare or familiar based on their inference. Table 8 details the precision, recall and f1 scores of both the annotators on selected contract types. <TBD Inference from the table>

5 Conclusion and Future Work

Our study aims at capturing significant components of a legal contract with an emphasis on identifying information that is specific and unique to a contract. While the approach successfully highlights and identifies significant components, there were a few limitations. The dataset contains new contracts as well as amendments. We screen amendments with fewer sentences, but filtering out amendments completely did not feel right as they are additions or alterations made to an existing legal agreement and can add value to the agreements. These amendments contribute to low coverage of contract elements. Increasing the data for each contract type might yield in better coverage and results. We show that our models can achieve reasonable accuracy with relatively low training data. <TBD add future work>

References

Narendra Andhale and LA Bewoor. 2016. An overview of text summarization techniques. In *2016 international conference on computing communication control and automation (ICCUBEA)*, pages 1–7. IEEE.

Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. *Pearson Correlation Coefficient*, pages 1–4. Springer Berlin Heidelberg, Berlin, Heidelberg.

Ilias Chalkidis and Ion Androutsopoulos. 2017. A deep learning approach to contract element extraction. In *JURIX*, pages 155–164.

Ilias Chalkidis, Ion Androutsopoulos, and Achilleas Michos. 2017. Extracting contract elements. In *Proceedings of the 16th edition of the International*

Conference on Artificial Intelligence and Law*, pages 19–28.

Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. Legal-bert: The muppets straight out of law school. *arXiv preprint arXiv:2010.02559*.

Qian Chen, Zhu Zhuo, and Wen Wang. 2019. Bert for joint intent classification and slot filling. *arXiv preprint arXiv:1902.10909*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2020. Language-agnostic bert sentence embedding. *arXiv preprint arXiv:2007.01852*.

Jenny Rose Finkel, Trond Grenager, and Christopher D Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL'05)*, pages 363–370.

Dan Hendrycks, Collin Burns, Anya Chen, and Spencer Ball. 2021. Cuad: An expert-annotated nlp dataset for legal contract review. *arXiv preprint arXiv:2103.06268*.

Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.

Zhiheng Huang and Wei Xu. 2015. Kai yu. *Bidirectional LSTM-CRF models for sequence tagging. CoRR, abs/1508.01991*.

Prathamesh Kalamkar et al. 2021. Indian legal nlp benchmarks: A survey. *arXiv preprint arXiv:2107.06056*.

Rahul C Kore, Prachi Ray, Priyanka Lade, and Amit Nerurkar. 2020. Legal document summarization using nlp and ml techniques. *Int. J. Eng. Comput. Sci*, 9:25039–25046.

Sibusiso Kubeka and Abejide Ade-Ibijola. Automatic comprehension and summarisation of legal contracts. *contract*, 9:10.

Tuyen Le et al. 2020. Automated requirements identification from construction contract documents using natural language processing. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, 12(2):04520009.

Chin-Yew Lin and FJ Och. 2004. Looking for a few good metrics: Rouge and its evaluation. In *Ntcir workshop*.

Xuezhe Ma and Eduard Hovy. 2016. End-to-end sequence labeling via bi-directional lstm-cnns-crf. *arXiv preprint arXiv:1603.01354*.

Sudip Mittal, Karuna P Joshi, Claudia Pearce, and Anupam Joshi. 2015. Parallelizing natural language techniques for knowledge extraction from cloud service level agreements. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 2831–2833. IEEE.

David Nadeau and Satoshi Sekine. 2007. A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26.

Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2017. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Thirty-first AAAI conference on artificial intelligence*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Pengjie Ren, Furu Wei, Zhumin Chen, Jun Ma, and Ming Zhou. 2016. A redundancy-aware sentence regression framework for extractive summarization. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 33–43, Osaka, Japan. The COLING 2016 Organizing Committee.

Paulo Silva, Carolina Gonçalves, Carolina Godinho, Nuno Antunes, and Marilia Curado. 2020. Using natural language processing to detect privacy violations in online contracts. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, pages 1305–1307.

Puyang Xu and Ruhi Sarikaya. 2013. Convolutional neural network based triangular crf for joint intent detection and slot filling. In *2013 ieee workshop on automatic speech recognition and understanding*, pages 78–83. IEEE.

Dan Yang, Christina Leber, Luis Tari, Aravind Chandramouli, Andrew Crapo, Richard Messmer, and Steven Gustafson. 2013. A natural language processing and semantic-based system for contract analysis. In *2013 IEEE 25th International Conference on Tools with Artificial Intelligence*, pages 707–712. IEEE.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. 2020a. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In *International Conference on Machine Learning*, pages 11328–11339. PMLR.

Ruixue Zhang, Wei Yang, Luyun Lin, Zhengkai Tu, Yuqing Xie, Zihang Fu, Yuhao Xie, Luchen Tan, Kun Xiong, and Jimmy Lin. 2020b. Rapid adaptation of bert for information extraction on domain-specific business documents. *arXiv preprint arXiv:2002.01861*.

Lucia Zheng, Neel Guha, Brandon R Anderson, Peter Henderson, and Daniel E Ho. 2021. When does pretraining help? assessing self-supervised learning for law and the casehold dataset of 53,000+ legal holdings. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, pages 159–168.

Haoxi Zhong, Chaojun Xiao, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. 2020. How does nlp benefit legal system: A summary of legal artificial intelligence. *arXiv preprint arXiv:2004.12158*.

Markus Zopf, Eneldo Loza Mencía, and Johannes Fürnkranz. 2018. Which scores to predict in sentence regression for text summarization? In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1782–1791, New Orleans, Louisiana. Association for Computational Linguistics.

A Example Appendix

This is an appendix.