LDSI W2021 Literature Review

Name: Md Siyam Sajeeb Khan (03720299)
Gloss ID: tum_ldsi_61

Accounting for Sentence Position and Legal Domain Sentence Embedding in Learning to Classify Case Sentences (Huihui Xu, Jaromir Savelka and Kevin D. Ashley)
Main idea: Analyzing the influence of sentence start position information and a legal domain-based sentence embedding scheme on sentence classification of legal texts.
Corpora and data annotation method: The type system contained three categories: issue, conclusion, and reason. Two third-year law students were deployed for annotating a randomly chosen pair of 1049 case and summary documents from a total of 28,733 case/summary documents. The annotation resulted in 11496 sentences from the summary documents and 215,800 sentences from the full case documents.
Learning methods and features: Three types of sentence embedding techniques, Sentence-BERT, Universal Sentence Encoder (USE), and Legal-BERT were used. The USE produced a 512-dimensional sentence vector embedding whereas the BERT-based methods produced a sentence embedding vector of size 768. They also experimented with meta-sentence embeddings such as averaging the embeddings of Legal-BERT and USE, and Legal-BERT and Sentence-BERT. A bi-directional LSTM was used on top of these features to test the performance of classification with and without the sentence position information.
Results and discussion: From the experiments, it was concluded that different embedding schemes produced varying levels of performances with and without the sentence start position. Overall, the sentence start position improved performance on full cases compared to the summaries and Legal-BERT was the best performing model for both cases. The authors investigated that in full texts, the models became more robust in detecting the 'Issue' type sentences when fed with the sentence start position. However, this did not apply to the summary documents. One strange phenomenon was the decrease in the model's performance when trained with a larger train set which requires further investigation.

Classifying Semantic Types of Legal Sentences: Portability of Machine Learning Models (Glaser, Ingo, Elena Scepankova, and Florian Matthes)
Main idea: Semantic type classification of sentences from legal contracts with a focus on transferability of ML models.
Corpora and data annotation: The authors used three different datasets for this task, the tenancy law of the German Civil Code (§535-§597) (BGB) containing 601 sentences, a basic version of a rental agreement dataset containing 169 sentences, and an extended version of it with 312 sentences created by the authors They used a classic rule-based technique for sentence segmentation. The 913 sentences in total were then manually categorized by two human legal experts and a third legal expert had a role in resolving disagreements over annotations. Due to the limited nature of the dataset, they proposed three different taxonomies.
Learning methods and features: The authors experimented with six different classifiers, namely, MNB, LR, SVM, MLP, RF, and finally DT. They leveraged two different feature representations, a TFIDF transformed feature vector and another feature vector based on word embeddings with word2vec. They utilized these models for calculating the mean embedding of a sentence.
Results and discussion: SVM and decision trees were the best performing classifiers. It was found that a combination of a simple TF-based vector instead of TF-IDF with the POS tag and the lemmatized token produced the best performance with an SVM classifier. The performance of the models varied across different classes, for instance among the 9 classes, it performed the worst while classifying sentences labeled as 'Consequence' in the rental agreements. However, the performance of the models fluctuated considerably in

different datasets. It was suspected that the disparity in sizes of the datasets, very low support for some classes, and different preprocessing steps might cause the portability issues of the ML models.

Sentence Classification Experiments for Legal Text Summarization (Ben Hachey, Claire Grover)
Main idea: The paper uses classifiers and maximum entropy models for the task of rhetorical role classification of sentences in legal documents for text summarization.
Corpora and data annotation method: The authors accumulated a corpus of judgments of the House of Lords. Along with some structural information, each document consists of mostly five Law Lord's judgments, a free-running text with at least one significant speech. There are a total of 188 documents. Manual summaries were available for 153 of these 188 documents. In total there are 98,645 sentences. Apart from the automatic annotation of sentences after applying the tokenization component, all sentences were manually labeled for their rhetorical status. The manual annotation is still an ongoing process and currently, 70 documents are annotated. For research purposes, the authors used 40 annotated documents which were manually annotated by two annotators.
Learning methods and features: C4.5, Naive Bayes, Winnow, SVM, Maximum Entropy Classification, and Sequence Modeling are the general classifiers employed in this paper. Location, Thematic Words, Sentence Length, Quotation, Entities, and Cue Phrases were all utilized as characteristics by the authors.
Results and discussion: SVMs and Maximum Entropy Models outperformed all other approaches. In comparison to the basis classifier, the sequence modeling technique enhanced performance. The authors are working on a system-level sentence extraction component and experimenting with active learning.

Sentence Embeddings and High-Speed Similarity Search for Fast Computer Assisted Annotation of Legal Documents (Hannes Westermann, Jaromır Savelka, Yunyao L, Siddhartha Brahma, Matthias Boehm, Laura Chiticariu, Rajasekar Krishnamurthy)
Main idea: The paper renders the sentence classification task in a different way to speed up the annotation process by associating a sentence with the label of its closest neighbors in terms of meaning.
Corpora and data annotation method: Three datasets are used in this paper: 1) 50 fact-finding decisions issued by the U.S. Board of Veterans' Appeals (BVA) from 2013 through 2017, 2) Statutory Interpretation Data Set, 3) 50 opinions of the Supreme Court of India.
Learning methods and features: To identify semantically similar sentences, the authors use 3 pre-existing models. They are 1) The Google Universal Sentence Encoder (GUSE), 2) Sentence Transformers built on top of BERT and RoBERTa, 3) InferSent, a BiLSTM network with max-pooling trained with fastText word embeddings.
Results and discussion: The neural models are quite good at recognizing similar sentences. In the worst-case scenario, 40% of the 20 closest sentences had the same classification, while the average was 70%. When applied to datasets including sentences from multiple areas, such as the Supreme Court of India dataset, the algorithm used has a lot of room for improvement.

Exploring the Use of Text Classification in the Legal Domain (Octavia-Maria Șulea, Marcos Zampieri, Shervin Malmasi, Mihaela Vela, Liviu P. Dinu, Josef van Genabith)
Main idea: In this paper, the authors investigate the application of text classification methods in the legal domain utilizing cases and rulings of a supreme court.
Corpora and data annotation: The authors use the large collection of court rulings from the French Supreme Court (Court de Cassation) with 126,865 unique court rulings. In most

documents, the common metadata available are: law area, timestamp, case ruling (e.g. cassation, rejet, non-lieu, etc.), case description, and cited laws.
Learning methods and features: As a learning method, the authors use a mean probability ensemble system combining the output of multiple SVM classifiers. As features, the authors use the word unigrams and the word bigrams.
Results and discussion: The experimental results show that provided the case description, rulings, and time span of cases, SVM ensembles can achieve high scores in predicting the law area and the ruling of a case. The ensemble method used in this paper outscored a previously suggested method by Șulea et al. (2017) in the paper "Predicting the Law Area and Decisions of French Supreme Court Cases" that used the same dataset.

Learning Explainable Linguistic Expressions with Neural Inductive Logic Programming for Sentence Classification (Prithviraj Sen, Marina Danilevsky, Hannes Westermann, Jaromır Savelka, Vern R. Walker, Kevin D. Ashley, and Karim Benyekhlef)
Main idea: The authors aim to learn explainable and modifiable models for sentence classification. They present RuleNN, a neural network architecture to learn linguistic expressions.
Corpora and data annotation: Datasets used in this paper are: 1) TREC (consists of questions), 2) Contracts (consists of sentences from legal contracts among enterprises).
Learning methods and features: Each sentence in the dataset is parsed using SystemT's SRL and dependency parser. The tense, aspect, mood, modal class, voice, and polarity of each action are extracted, together with semantic arguments. Predicates are then built for each label using hand-crafted dictionaries. RuleNN then learns 50 Linguistic Expressions comprising up to four predicates. RuleNN comprises two sorts of modules: the Predicate Generation Module and the Clause Generation Module.
Results and discussion: The experiments show that RuleNN performs better both in terms of efficiency and quality of rules learned. RuleNN is contrasted against other rule induction techniques such as NeuralLP and @ILP from neuro-symbolic AI, LSM, and BoostSRL (BSRL) from StarAI, MITI, and MIRI from multiple instance learning, and top-down ILP system metagol (MG) and its noise-tolerant counterpart. A more in-depth comparison is made using MINet and BiLSTM. In all circumstances, RuleNN outperforms or provides almost similar performances.

Summary
Sentence classification task in the legal domain has its own challenges and almost all the literature discussed above adopt different interesting approaches for solving this problem. All these methods treat this classification task as a supervised learning problem and try to build annotated data for training different machine learning models. No universally best-performing model can be generalized for all types of datasets and models are not always portable from one dataset to other as shown by Glaser et al. Also, the model struggles in classifying sentences that have semantic similarities. The models' performances depend heavily on the quality of the annotation data. However, many of these annotation tasks require significant human effort and due to the error proneness of humans, often there are instances of incorrect annotations. Westermann's approach of assisting the annotators by retrieving semantically similar sentences can be highly useful in this regard and speed up the annotation process. Additionally, sentence classification tasks heavily depend on the features used and it has been observed in the survey that different BERT-based word embedding techniques exhibit impressive performance with much lower feature dimensions than TFIDF features. Developing law-specific meta-sentence embeddings as suggested by Xu et al. is a promising idea. The domain of law is inherently complex due to the nature of its ambiguity and sometimes it is even difficult for a human to distinguish between different types. This domain mostly suffers from the scarcity of high-quality annotated data and with carefully curated annotation data the performance of the models can be improved and better generalized for different legal cases.