LDSI S2021 Literature Survey

Name: Irtiza Chowdhury
Gloss ID: ldsi_s2_18

Summary

This survey attempts to briefly summarize the methodologies and findings of six papers which use sentence classification in some capacity on various legal texts. A brief discussion to compare and contrast them by different metrics follows afterwards.

Sentence Classification Experiments for Legal Text Summarisation (Hachey et al.)

The authors looked at several machine learning methods to build a classifier which determines the rhetorical status of sentences. A corpus of judgements of the UK House of Lords was utilized, consisting of 188 documents. Two annotators used guidelines developed by one of the authors, one of the annotators and a law professional to create a corpus of 40 documents, containing over 10000 sentences, which were annotated automatically except for manual annotation of sentences for their rhetorical status. The feature set comprised of sentence location, Thematic strength, sentence length, Cue phrases and features indicating initial part-of-speech, word features, etc. Four classifiers were initially used from the Weka package: C4.5 decision tree, Naive Bayes, Winnow and Support Vector Machines. Continuous features had to be discretized for the Winnow algorithm.

Both individual micro-averaged F-scores for each feature and scores incorporating features incrementally were presented for every classifier. C4.5 proved best with only location features but performed poorly with other features. Considering all features, Winnow had the lowest score while SVMs had the highest. Naive Bayes performed second best with all but thematic word features. However, a relatively high improvement was observed in Winnow and C4.5 upon using a more uniform distribution of rhetorical categories and considering the respective baselines. Maximum Entropy and Sequence Modelling were also carried out which showed interesting observations discussed later.

A Semi Supervised Approach for Catchphrase Classification in Legal Text Documents (Bajwa et al.)

A semi-supervised approach is used by the authors to extract catchphrases (important legal points), minimizing the understanding complexity of the reader. Few preprocessing steps were taken to create a corpus from software license agreement, including lexical processing, syntactic processing and filtering. The corpus of software license agreement was used as knowledge base containing 250 terms. Sentence splitting, tokenization and POS tagging was followed by thematic segmentation into License Grant, Ownership and Terms and Conditions. These segments contained catchphrases such as Fees, Limited Use, Warranties etc. Filtering was performed to remove text without loss of relevant information. Thirteen attributes, including TFIDF and features related to citing and occurrence of catchphrases were used. An Expectation-Maximization approach using Naive Bayes classifier was taken to achieve results with highest precision of 88.9% and recall of 82.8% across four case studies.

Recognizing Cited Facts and Principles in Legal Judgements (Shulayeva et al.)

The authors trained a classifier to automatically annotate cite facts and principles. The corpus was compiled from 50 common law case reports from British and Irish Legal Institute, comprising of issues covered by contract, trust and property law containing over 22000 sentences. Sentence splitting was performed using GATE while annotation guidelines were created by an expert annotator upon which another annotator was trained on. Features such as Part of Speech Tags, Unigrams, Dependency pairs, and others which could help to classify sentences were used for various reasons. Feature counts were normalised by TF and IDF and 10-fold cross-validation was performed at a sentence level. All of these were used to train a Naive Bayes Multinomial classifier.

The results showed precision and recall scores between 79% and 89%, despite the corpus being unbalanced, as well as a good agreement score. However, few issues still remained such as possible contamination of the test set, no independence assumption of Naive Bayes classifier and no other models being explored to compare results to.

Classifying Legal Norms with Active Machine Learning (Waltl et al.)

An active machine learning approach for classification with different strategies to decrease the required training data is discussed in the paper. A legal expert annotated every sentence of the tenancy law section in the German civil code published in 2017, resulting in 504 sentences with eight classes after removing labels with less than 1.2% support. A 75-25 train test split was performed, balancing classes as much as possible. Tokens and POS tags were used as features to represent norm instances. Nine combinations using AML queries and three combinations using conventional supervised learning (CSL) were experimented for Multinomial Naive Bayes, Logistic Regression and Multi-Layer Perceptron classifiers. Each round consisted of randomly querying instances from the unlabeled pool, labeling and then learning. Five most informative instances were used subsequently for query strategy with five-fold cross validation, and then applied on the test set.

Precision, recall, F1 and accuracy were calculated along with learning curves for visualization. Compared to CSL, AML proved superior when using NB and LR, increasing speed of learning and maximum accuracy. However, increasing AML rounds increased chances of overfitting. Some norms had a high F1 score and some were not easily classified. While initially thought to be as a result of low support, a high precision and recall scores indicated that worsening results were mostly due to overfitting.

Identification of Rhetorical Roles of Sentences in Indian Legal Judgments (Bhattacharya et al.)

The authors took a deep learning approach to automatically identify rhetorical roles of sentences across five legal domains. They used legal judgements from the Supreme Court of India, sampling 50 documents from the top 5 domains proportionally. GATE Teamware tool was used by 3 annotators to annotate the documents with 7 rhetorical roles, such as Facts, Argument etc. Precision, Recall and F-score were calculated instead of Kappa. Each document was split into sentences using SpaCy with accuracy close to 90%. A baseline with CRF and features such as POS tags, layout features, cue phrases, etc. was carried out. Two deep learning approaches were considered. A Hierarcichal Bidirectional LSTM classifier was

used with initialization of sentence embeddings using sent2vec from 53K court case documents. The model was enriched further with a CRF on top of the architecture using feature vectors generated by top-level BiLSTM.

Using 5-fold cross validation and 10 documents for testing, the macro averaged Precision, Recall and F-score were calculated. Neural models performed significantly better and using pretrained embeddings showed high improvement. The model was satisfactory across all labels except one which was interleaved with other labels and contained only 9% of total sentences. Performance across domains was consistent to the inter-annotator agreement, showing the advantage of neural models in requiring no hand-crafting of features to get impressive results.

Sentence Embeddings and High-Speed Similarity Search for Fast Computer Assisted Annotation of Legal Documents (Westermann et al.)

An approach called “lateral annotation” is presented in this paper where sentence embeddings are used to capture meaning of sentences to deliver semantically similar ones in assisting manual annotation. Three data sets comprising of 50 BVA decisions with over 6000 sentences, Statutory Interpretation database of sentences from case law mentioning three terms from provisions of the US Code, and 50 opinions of the Supreme Court of India containing over 9000 sentences. The different sentences in the datasets were classified based on the rhetorical roles played by the sentences in the decision. The system retrieved closest sentences to a query sentence and observed the precision. The longest chain and average chain length are considered for each dataset and embedding method.

The sentence embeddings proved to capture enough linguistic information and the neural models performed much better than random and TF-IDF baselines. The BVA dataset seemed to benefit most from the lateral annotation method while it was least effective in the Supreme Court dataset as the sentence embeddings did not separate the classes clearly. A direct classification difficulty was observed by other researchers working on this dataset which indicates the importance of sentence embeddings in classification.

Discussion

The language used in most legal documents is greatly specialized with long sentences and interleaving of different classes of sentences which create vast possibilities, making it harder for classifiers to recognize patterns. Shulayeva et al [3]. pointed out that courts embed facts within legal reasoning, creating a wide variety of grammatical patterns and few instances, leading to difficulty in classification. The large and noisy nature of legal texts makes it suitable for Maximum Entropy approaches as noted by Hachey et al [1]. where they showed significant improvement over Naive Bayes where features are highly dependent. Furthermore, a sequence modelling approach improved performance greatly as it helped to consider the context of boundary. The importance of context is further emphasized by Westermann et al [6]. as they incorporated sentence embeddings to maximize accuracy by capturing context. According to them, “Citation” label seemed by far the easiest because of the special tokens, e.g. parentheses and the word ‘See’, as well as “Rule” because of being cited multiple times. Waltl et al [4]. noticed types such as “Definition” were hard to classify as the kind of definitions could linguistically vary from the one in the test set. The complex structure of legal documents,

such as reason behind judgment interleaving with Precedents and Statutes, was pointed out by Bhattacharya et al [5]. Having a higher fraction of sentences in the corpus as well as always having a fixed a position in a document also lead to easier classification, but a lower fraction of total sentences in general was an issue for neural models. Despite being a comparatively newer approach, the considerably better results shown by deep learning models with no hand-crafted features indicate that it is a viable option. How well it can generalize across different categories still remains to be seen.

References

1. Hachey, B., Grover, C.: *Sentence classification experiments for legal text summarisation*. In: Proceedings of the 17th Annual Conference on Legal Knowledge and Information Systems (Jurix). (2004)
2. Bajwa, I.S., Karim, F., Naeem, M.A., ul Amin, R.: *A semi supervised approach for catchphrase classification in legal text documents*. JCP 12(5), 451–461. (2017)
3. Shulayeva O, Siddharthan A, Wyner A (2017) Recognizing cited facts and principles in legal judgements. Artificial Intelligence and Law 25(1):107–126
4. B. Waltl, J. Muhr, I. Glaser, G. Bonczek, E. Scepankova, and F. Matthes. *Classifying legal norms with active machine learning*. In Proceedings of the 30th International Conference on Legal Knowledge and Information Systems, pages 11–20, Luxembourg City Luxembourg. (2017)
5. P. Bhattacharya, S. Paul, K. Ghosh, S. Ghosh, and A. Wyner, *"Identification of rhetorical roles of sentences in indian legal judgments,"* in Proc. International Conference on Legal Knowledge and Information Systems (JURIX). (2019)
6. Hannes Westermann, Jaromír Šavelka, Vern R Walker, Kevin D Ashley, and Karim Benyekhlef. *Sentence Embeddings and High-Speed Similarity Search for Fast Computer Assisted Annotation of Legal Documents*. In JURIX 2020, Vol. 334. IOS Press, 164. (2020)