LDSI S2021 Literature Survey

Name: Quazi Fahim Faisal Dhruba (03722194)
Gloss ID: ldsi_s21_13

Paper: Recognizing cited facts and principles in legal judgements
Authors: Olga Shulayeva, Advaith Siddharthan, Adam Wyner
General Remark
The paper addresses the recognition of facts and principles in an automatic, machine learning driven fashion which is comparable to the same task being done by manual annotators.
Corpora
The paper compiles a corpus from **50 common law reports taken from the British and Irish Legal Institute (BAILII)** website in RTF format. This data is first used for a manual annotation study using two annotators and then an automated annotation study.
Learning Methods and Features
The features used by the authors are the following:
1) Parts of Speech Tags, 2) Unigrams, 3) Dependency Pairs, 4) Length of Sentence, 5) Position in the text, 6) Cit (which indicates the presence of a citation instance in the sentence)
The following preprocessing steps and learning methods were used:

1) Feature counts were normalized by tf and idf.
2) Attribute Selection was performed over the entire dataset
3) The **Naïve Bayes Multinomial classifier** was used
4) Results were reported for tenfold cross-validation

Results Obtained and Remaining Problems
The automated classifier developed overall performed better than the annotator with no legal training (Accuracy of 0.85). Dependency Pairs were the most important features, which, when used by themselves achieved an accuracy of 0.81. The study overall address classifying over only 3 classes (Facts, Principles and Neither), thus it remains to be seen if the approach scales to more complex multi-class classification tasks.

Paper: Sentence Classification Experiments for Legal Text Summarisation
Authors: Ben Hachey, Claire Grover
General Remark
The paper addresses the task of rhetorical role classification using classifiers and maximum entropy models.
Corpora
The paper utilizes a corpus of **judgements of the House of Lords**. It consists of 188 documents. 153 out of these 188 documents have manually created summaries available which the authors used for system evaluation. The authors used 40 manually annotated documents for their experiments.
Learning Methods and Features
The following features were used by the authors:
1) Location relative to the encoded LORD element and the paragraph, 2) Thematic Words, 3) Sentence Length, 4) Quotation, 5) Entities, 6) Cue Phrases.

Learning Methods used by the authors:
General Classifiers: C4.5, Naïve Bayes, Winnow, SVM
And additionally, Maximum Entropy Classification and Sequence Modelling

Results Obtained and Remaining Problems

All methods used perform quite well compared to the baseline. SVMs and Maximum Entropy Models perform the best. The authors continue to improve their sentence extraction component and explore the use of active learning

Paper: Sentence Embeddings and High-Speed Similarity Search for Fast Computer Assisted Annotation of Legal Documents

Authors: Hannes Westermann, Jaromır Savelka, Yunyao L, Siddhartha Brahma, Matthias Boehm, Laura Chiticariu, Rajasekar Krishnamurthy

General Remark

The paper explores a unique approach of speeding up annotation, by attempting to automate the recognition of semantically similar sentences which may be laterally annotated by an annotator.

Corpora

The authors utilize three separate datasets:
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

The authors use 3 pre existing models to identify semantically similar sentences:
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 Obtained and Remaining Problems

The neural models perform quite well in identifying similar sentences. At its worst 40% of the 20 closest sentences had the same label while in general it was 70%. The method utilized has large scope for improvement when applied to datasets comprised of sentences from various domains, as was the case for the dataset from the Supreme Court of India.

Paper: Learning Explainable Linguistic Expressions with Neural Inductive Logic Programming for Sentence Classification

Authors: Prithviraj Sen, Marina Danilevsky, Hannes Westermann, Jaromır Savelka, Vern R. Walker, Kevin D. Ashley and Karim Benyekhlef

General Remark

The authors aim to learn explainable and modifiable models for sentence classification. They present RuleNN, a neural network architecture to learn linguistic expressions.

Corpora

Two Datasets were used:
TREC, which comprised of questions and Contracts which comprised of sentences from legal contracts among enterprises.

Learning Methods and Features

Each sentence in the dataset is processed using SystemT's SRL and dependency parser. Each action's tense, aspect, mood, modalclass, voice and polarity is extracted along with semantic arguments. Predicates are then constructed using hand-crafted dictionaries for each label. RuleNN then learns 50 Les (Linguistic Expressions) containing upto 4 predicates.
RuleNN consists of two types of modules, the Predicate Generation Module and the Clause Generation Module.

Results Obtained and Remaining Problems

RuleNN is compared against NeuralLP and @ILP, from neuro-symbolic AI; LSM and BoostSRL (BSRL), from StarAI; MITI and MIRI, from multiple instance learning; and topdown ILP system metagol (MG) and its noise-tolerant variant, these being rule-driven models. Further comparison is done against MINet and BiLSTM. RuleNN has comparable performance if not better in all cases. Further user study to determine explainable, denotes possible model compression by 96% through human driven modifications.

Paper: Machine Learning versus Knowledge Based Classification of Legal Texts
Authors: Emile de Maat, Kai Krabben and Radboud Winkels

General Remark

The authors compare the performance of a pattern-based classifier against a machine learning approach, to classify sentences in Dutch legislation.

Corpora

The Dutch Legislation

Learning Methods and Features

The authors analyze a sentence by first selecting the words it contains in accordance with the following features: 1) stop list, 2) stemming, 3) grouping of numbers, 4) all words are converted to lowercase, 5) a minimal term frequency. The selected words are then analyzed further using the following: 1) binary weight, 2) term frequency, 3) inverse document frequency weight.
An SVM classifier is used against the pattern-based classifier.

Results Obtained and Remaining Problems

The SVM classifier has comparable performance to the pattern-based classifier, in some cases even doing worse. The authors recognize the following issues when analyzing the classifiers: 1) presence of Keywords in subordinate sentences, 2) missing standard phrases, 3) Keywords being linked to different classes, 4) "Statement of fact" label being largely misclassified by the ML model, 5) in general not enough data, 6) sparse data, 7) Focus on wrong keywords, 8) presence of keywords outside of a standard phrase and 9) skewed data.

Paper: An unsupervised approach to sentence classification
Authors: Shailesh S. Deshpande and Girish Keshav Palshikar and G. Athiappan

General Remark

The authors explore a knowledge-based unsupervised approach to classify sentences based on a calculated specificity score derived from several features.

Corpora

1) Employee Satisfaction Survey responses
2) Product Reviews by Kelty

Learning Methods and Features

The paper utilizes the following word level features based on the ontology WordNet (T) and general document factors:
1) Average semantic depth (distance in terms of edges from the root of T), 2) Average Semantic Height (as the length of the longest path from the word w to a leaf node under w in the given ontology T), 3) Total Occurrence Count, 4) Number of Named Entities, 5) Sentence Length, 6) Number of Proper Nouns. These features are then used to calculate a specificity score.
These features are then used to classify sentence into suggestions and complaint classes.

Results Obtained and Remaining Problems

In two separate case studies the clustering of sentences based on the features and specificity score calculated prove useful. Further improvements in narrowing down more specific sentences can be pursued.

Overall Discussion Points

- Results are often not reproducible on data of different legal contexts as illustrated by the work of Hannes et al. when analyzing results on the dataset from the Supreme Court of India.
- Interpretability and modifiability are of significant concern as predictions and classifications by black box models may not be applicable in a legal setting. This is demonstrated by the prevalence of work such as RuleNN.
- Classes which have obvious structural consistency are often easier to classify, such as citations as denoted by Olga et al.
- Models often struggle to capture nuances and subtle differences in classes such as "Statement of Fact" and "Obligations" as denoted by Emile et al.
- Overall, the domain seems to suffer from a lack of high quality labeled data, which has led to avenues being explored to increase efficiency in the manual annotation process such as the one proposed by Hannes et al. even if complete automation is hard to achieve.