

A Semantic Analysis of a Comprehensive Corpus of Court Decisions for the Development of a Predictive Model of the Judicial Risk

Gildas Tagny Ngompe^{1,2}, Sébastien Harispe¹, Jacky Montmain¹, Stéphane Mussard², Guillaume Zambrano²

¹ École des mines d'Alès, LGI2P, 69 Rue Georges Besse, Nîmes, France

² Université de Nîmes, CHROME, Rue du Dr Georges Salan, Nîmes, France

Abstract. How can we observe and understand how judges usually make their decisions knowing that their subjective interpretation of legal rules make the application of the law uncertain? We are currently developing an approach to study comprehensively the jurisprudence by solving the challenge of the retrieval and analysis of the always growing big number of court decisions in France. This project addresses the tasks of extracting information from decisions in the aim of building a legal knowledge base (KB). It progresses to study how such a KB could be used for studying the judicial risk by being the cornerstone of descriptive statistical and predictive analyses. Our early extraction results show that CRF-based models work well for the segmentation of texts and the extraction of named entities.

Keywords: Court decisions analysis, information extraction

1 Introduction

The subject aims the extraction of law cases information from the textual contents of court decisions and their normalization to build a knowledge base of the jurisprudence. This knowledge base will enable the computer to structure and classify court decisions with regard to several specifications such as the applied rules and the obtained results. There are various application frameworks like better understanding how legal rules are applied, predicting court outcomes, search similar decisions ...

First, we focus on designing a system to extract information from past decisions. Then, we will build an analysis system based on some frequencies of results of judges given the demand and the norms applied in a selected set of decisions. These frequencies will give us access to foresee the probabilities that some judges make a certain decision in a given situation.

The expected results will not only help lawyers, but also individuals and businesses. The huge volume of decisions (over 4 million in France per year) make impossible an exhaustive manual analysis. Moreover, the justice is too complex and its language hardly understandable to allow non-lawyers to know the legal consequences of their actions without the help of an expert.

2 Analysis Approaches of Court Decisions

Ambiguities in the conditions of applying legal rules are left to the unique appreciation of judges in real cases. To study judges decisions, common automatic approaches are mostly predictive. The complexity in the application of the law with other reasons led to the fail of legal expert systems [1]. The only knowledge of legal rules is not enough to give advice. That is why recent approaches have been developed based on the analysis of past court decisions. Reasoning by analogy seems logical because we expect from a fair justice to give similar result for similar cases. Some of these approaches described a case with some metadata about the court, the judges(e.g. political party), or the case (e.g. the issue area of the case). Hence, statistical methods based on binary classification trees were designed to predict the votes of the judges of the U.S. Supreme Court. After poor performance with a tree per judge (66.7% for the tree with 67.9% for human experts [2]), an extremely randomized trees based approach [3] offers a better but still low accuracy (70.9%). The system "SMILE + IBP" [4], in contrary, relies on predefined "Factors" that model the facts to classify the cases. Hence, by combining rule-based and case-based reasoning, "SMILE + IBP" determines the favored party in a given case. This system presented a good accuracy (91.8%) but was restricted only to trade secret misappropriation cases. Moreover, the experiments were done on few cases (184 cases from the CATO's database).

Comparing our project to those works, we propose another approach to analyze semantically and statistically a huge set of court decisions. We can imagine many applications of these statistics. For instance, it will be easier and faster to get a clear and comprehensive insight on the application of the law over the time and in different locations. Moreover, a predictive analysis might be designed upon these statistics to forecast future decisions with other factors not available in the contents of decision (e.g. major events, their political ideology, their experience, change in the legal standards).

3 Information Extraction challenges and early results

Even if obtaining court decisions is not an easy task, we have gathered more than 600k decisions of various courts. Building the legal knowledge base requires a description of these decisions. The relevant information can be founded by analyzing the unstructured text of decisions. The forms of these information define various tasks of text analysis. Some elements(location, date, parties, judges) are directly recognizable in the text and their extraction is similar to the named entities recognition(NER). But to discover the demands, we suggest an unsupervised clustering of decisions. After identifying the demands in a decision, the computer must be able to identify and interpret the conclusion of judges to know if the demand was rejected or accepted.

Beginning by the NER task, we noticed that the entities are distributed between three main sections on which we chose to organize the extraction separately: header (case references), the middle text (norms invoked and demands

quantum), and the conclusion (norms applied, results quantum). It seemed logical at first that a rule-based system could easily detect these sections. We have implemented such a system that has shown its limits by suggesting multiple schema of segmentation for a document and a lot of wrong schema due to the detection of pattern occurrences at wrong location.

Many works demonstrated the effectiveness of probabilistic graphical model in some sequence tagging tasks like the segmentation of FAQs as questions and answers[5] or like for NER in the header of scientific papers (HMM and CRF[6]). Thus, we designed two approaches based on probabilistic models: Hidden Markov Model(HMM) and Conditional Random Fields (CRF). The two approaches determine the section of each line of a document. For the CRF-based model, we manually labeled and defined some features (line number, length ...) of the lines in a set of 543 decisions from the Court of Appeal of Nimes. The excellent performance (table 1) of CRF-based model demonstrates its effectiveness for decision segmentation.

TAG (meaning)	Precision	Recall	F1 score	TAG	Precision	Recall	F1 score
E (header)	0.9625	0.9433	0.9528	E	1	1	1
T (middle text)	0.7679	0.703	0.734	T	1	1	1
D (conclusions)	0.3272	0.4486	0.3784	D	1	1	1
F1 Average = 0.6884				F1 Average = 1			

(a) HMM

(b) CRF

Table 1. Evaluation of HMM-based and CRF-based segmentation on 109 decisions

Then, we implemented and compared two NER approaches based respectively on HMM and CRF. The objective of the systems is to tag each token (word or punctuation) in the headers with the label of the entity the token belongs to. After labeling manually the entities in 543 headers from the precedent dataset, we defined some tokens features for the CRF (e.g. position in line, number of characters). With the two models trained with 80% of the data, we observed that CRF-based system outperforms the HMM-based system on the 109 remaining headers (table 2). However, the CRF-based system detects quite accurately some entities (e.g. city, jurisdiction) but remains ineffective on the name of parties (plaintiffs and respondents). Some more discriminating features are needed to improve the performance of the CRF-based system.

4 Conclusion

During our work, we gathered more than 600 thousands decisions from various courts. We labeled a training dataset of 543 documents for segmentation and NER in the headers. The labeling took a lot of time since we did it manually. Moreover, some features were defined and computed for the CRF-based systems.

TAG (meaning)	Precision	Recall	F1 score	TAG	Precision	Recall	F1 score
O (Not An Entity)	0.9717	0.9352	0.9531	O	0.9932	0.9946	0.9939
RG (<i>numéro</i>)	-	0	-	RG	1	0.9908	0.9954
JR (<i>jurisdiction</i>)	0.6149	0.8349	0.7082	JR	1	0.9908	0.9954
VL (<i>ville</i>)	0.876	0.9725	0.9217	VL	1	0.9908	0.9954
FM (<i>formation</i>)	0.8428	0.9872	0.9093	FM	1	1	1
DT (<i>date</i>)	0.9729	0.9818	0.9773	DT	1	0.9848	0.9923
APL (<i>appellant</i>)	0.2551	0.1725	0.2058	APL	0.8491	0.7392	0.7904
AV (<i>avocat</i>)	0.8493	0.9573	0.9	AV	0.9681	0.954	0.961
ITM (<i>intimé</i>)	0.1974	0.4162	0.2678	ITM	0.7982	0.864	0.8298
ITV (<i>intervenant</i>)	-	0	-	ITV	-	0	-
JG (<i>juge</i>)	0.5732	0.9719	0.7212	JG	0.9445	0.9414	0.943
RL (<i>rôle</i>)	0.9048	0.0435	0.083	RL	0.8758	0.8879	0.8818
F1 Average = 0.5539				F1 Average = 0.8649			

(a) HMM

(b) CRF

Table 2. Evaluation of HMM-based and CRF-based NER on 109 headers

However, we obtained some excellent results on our testing dataset even if some entities remain hardly recognizable.

Future works will focus first on the recognition of the others entities with a priority on norms. Next, we intend to use the norms to discover automatically the demands available in our corpus because similar demands are generally based on similar norms. So we want to cluster the decisions in such a way that each cluster represent a demand. The clusters should overlap since a decision can contain multiple demands.

References

1. Philip Leith. The rise and fall of the legal expert system. *European Journal of Law and Technology*, 1(1):179–201, 2010. <http://ejlt.org/article/view/14/1>.
2. Andrew D Martin, Kevin M Quinn, Theodore W Ruger, and Pauline T Kim. Competing approaches to predicting supreme court decision making. *Perspectives on Politics*, 2(04):761–767, 2004. <http://wusct.wustl.edu/media/man1.pdf>.
3. Daniel Martin Katz, Michael James Bommarito, and Josh Blackman. Predicting the behavior of the supreme court of the united states: A general approach. *Available at SSRN 2463244*, 2014.
4. Kevin D. Ashley and Stefanie Brüninghaus. Automatically classifying case texts and predicting outcomes. *Artificial Intelligence and Law*, 17(2):125–165, 2009. <http://link.springer.com/content/pdf/10.1007%2Fs10506-009-9077-9.pdf>.
5. Andrew McCallum, Dayne Freitag, and Fernando CN Pereira. Maximum entropy markov models for information extraction and segmentation. In *ICML*, volume 17, pages 591–598, 2000.
6. Fuchun Peng and Andrew McCallum. Accurate information extraction from research papers using conditional random fields. In Daniel Marcu Susan Dumais and Salim Roukos, editors, *HLT-NAACL 2004: Main Proceedings*, pages 329–336,

Boston, Massachusetts, USA, May 2 - May 7 2004. Association for Computational Linguistics.