AI for the legal domain: an explainability challenge

Alexandre Quemy
IBM Software Lab, Cracow, Poland
Poznań University of Technology, Poznan, Poland
aquemy@pl.ibm.com

1 Scientific and societal context

The legal environment is a *messy concept* [6] that intrinsically poses a certain number of difficulties to analyze: grey areas of interpretation, many exceptions, non-stationarity, deductive and inductive reasoning, non-classical logic, etc. In other words, it combines some of the most challenging elements for data scientists and mathematicians to study formally.

Problem statement: Is it possible to predict justice decisions and at the same time come up with an intelligible explanation based on legal arguments?

Problem importance: For some years and in several areas of the Law, some "quantitative" approaches have been developed, based on the use of more or less explicit mathematical models. With the availability of massive data, those trends have been accented and brand new opportunities are emerging at a sustained pace. Among the stakes of those studies, one can mention a better understanding of the legal system and the consequences some decisions on the economy, but also the possibility to decrease the mass of litigations in a context of cost rationalization. However, if statistical models provides better results at predicting justice decisions than expert knowledge systems, they often act as a black-box which is redhibitory for practical applications. Beyond the scope of the legal domain, explainability is very hot topic in Machine Learning (see e.g. XML contest[1]) and is of a particular importance to safely and ethically apply AI to the society.

Contribution: We elaborate a new machine learning algorithm based on hypergraph learning for classification with a huge potential for explainability. On top of a validation on standard datasets, we created the first large legal dataset based on real-data coming from the European Court of Human Right. With structured and unstructured features for several thousands of cases, we are able to extensively validate our approach, compare it to other existing methods and investigate explainability. The method being agnostic to the domain, we showed it can be used for any classification problem. On top of that, it offers some valuable properties that we detail briefly in Section 3.

[1]https://community.fico.com/community/xml

IFIP 2018, September 2018, Poznan, Poland
2018.

2 Previous work and selected problems

In [4], and after discussing with legal practitioners, we extracted four open-problems that received little cover:

- Predicting the outcome of a case given the legal environment. (Prediction)
- Building a legal justification, given some facts, a set of law texts with the jurisprudence and an outcome. (Justification)
- Taking the *best* decisions w.r.t. the legal environment dynamics and some criteria. (Decision)
- Modifying the legal environment dynamics to match some criteria. (Control)

The literature shows that most of the predictive power of the best forecasting methods holds in non-legal factors (e.g. estimated ideology of the judge). However and by definition, building a legal justification requires to use only legal arguments. As far as we know, the legal domain is the only field where the prediction problem is separated from the justification one. This doctoral project focuses on the challenge to solve both conjointly.

The Prediction problem is challenging by itself, even for the best legal experts: for the Supreme Court of the United States (SCOTUS), 58% accuracy has been reached in [7] . Using crowds, the Fantasy Scotus[2] project reached 84,85% correct predictions. No similar results exist in Europe. In general, the previous approaches can be broken down into three groups, namely: the statistical models, the case-based reasoning (CBR) and the abstract argumentation (AA). If the statistical methods provide interesting results for the prediction problem [1–3, 7], they cannot handle the justification problem. On the opposite, CBRs do not integrate non-legal factors and thus are unable to handle the prediction problem while they do (partially) answer the justification problem. In AA, two kinds of opposed approaches emerged: a *positive* one that intends to model real-life decision processes, and a *normative* one that tries to elaborate methods to select among the best alternatives and discuss arguments. The first approach may handle the prediction problem and the second one the justification problem. They both heavily rely on expert knowledge to construct the so-called “arguments”, which limit the applicability of AA. For a more comprehensive view on the state-of-art, we refer the reader to [4], and in particular to Table 1.

[2]https://fantasyscotus.lexpredict.com/

3 Results and contributions

This doctoral project provides several results and contributions at different levels. In this section, we briefly present the main contributions, work in progress and further plans.

Fundamental problems: The first contribution of this doctoral project consists in formalizing fundamental problems of legal analytics and study the relation between them [4]. If this project focuses on some in particular, all of them raise new challenges to the machine learning community.

Hypergraph Case-Based Reasoning: Keeping in mind the desired properties and inherent drawbacks of each partial solution analyzed in [4], we developed a new supervised algorithm for classification called Hypergraph Case-Based Reasoning (HCBR) [5]. As suggested by the name, it represents a training set as a hypergraph and uses the partition induced by the subhypergraphs to estimate, for a given subset of features, the support toward a specific class. It has been shown to perform as good as the state-of-art methods on some well-known datasets (see Table 1).

The method offers several interesting properties, not only useful for the application to the legal domain. In particular, the model space and data representation as hypergraph provides a convenient and promising way to explain not only the model but also and mostly each decision separately based on the interactions with past decision (e.g. seen as "counter-examples" or "analogies" in case of a trial, like in CBR systems). Additionally, the sensitivity to hyperparameters is neglictable s.t. time-consumming tuning is not mandatory for the end-user. A "online" version exists. Last but not least, HCBR does not assume any metric on the feature space, is agnostic to the feature representation and can work with incomplete or unstructured datasets.

For the experiments, a fast, scalable and open-source[3] modern C++ implementation of the different versions of HCBR has been developed. A highly parallel version is planned before the end of the doctoral project to handle massive datasets.

Current on-going work focuses on theoretical properties of "Hypergraph sequences" and HCBR in general, as well as model space extension to be able to represent more complex functions. Explainability is also a main axis of development to handle the justification problem mentioned above. Finally, larger experiments in the legal domain are also being conducted.

ECHR Datasets and Open Database: To be able to apply HCBR to the legal domain, we needed to create an open database using real-life data. The European Court of Human Rights publishes all documents related to cases in natural language. This court is very important for all Europeans and provides over 50k decisions. In a work to be released soon, we extracted standard descriptive features (very structure database with several columns like dates, parties, court members, article in discussion, etc.) and complex bag-of-words representation from the court judgments (structured by paragraphs), including entity matching using IBM Watson Services (semi-structured representation). It will also offer several datasets: for specific law articles, for binary classification, multiclass and multilabel classification (and probably other usage not intended at first). The purpose of this work is twofold. First, we expect it to draw the attention of researchers on a very important subject for society that offers new challenges to the Machine Learning community. Secondly, finding large and open dataset based on real life data which is usually a problem in Machine Learning. Indeed, people tend not to share their data and/or keep reusing small synthetic datasets that are not reflecting the real-life difficulties.

[3]https://github.com/aquemy/HCBR

Table 1. Matthew Correlation Coefficient and rank obtained with several methods (average over 7 various datasets). Implementation provided by Scikit-Learn.

Method	MCC	Rank
Neural Network	0.8914	1
HCBR	0.8435	2
RBF SVM	0.8267	3
Decision Tree	0.8066	4
AdaBoost	0.8063	5
k-NN	0.7859	6
Linear SVM	0.7858	7
QDA	0.7358	8
Random Forest	0.7237	9
Naive Bayes	0.6953	10

References

[1] R. Guimerà and M. Sales-Pardo. 2011. Justice Blocks and Predictability of U.S. Supreme Court Votes. *PLOS ONE* 6, 11 (2011), 1–8.

[2] D. M. Katz, M. J. Bommarito, and J. Blackman. 2014. Predicting the Behavior of the Supreme Court of the United States: A General Approach. *SSRN Electronic Journal* (2014).

[3] A. D. Martin, K. M. Quinn, P. T. Kim, and T. W. Ruger. 2004. Competing approaches to predicting supreme court decision making. *Perspectives on Politics* (2004), 761–767.

[4] A. Quemy. 2017. Data Science Techniques for Law and Justice: Current State of Research and Open Problems. In *ADBIS Short Papers and Workshops, AMSD, BigNovelTI, DAS, SW4CH, DC, 2017*. 302–312.

[5] A. Quemy. 2018. Binary Classification With Hypergraph Case-Based Reasoning. In *International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data (DOLAP), 2018.*

[6] E. L. Rissland. 2006. AI and Similarity. *IEEE Intelligent Systems* 21, 3 (2006), 39–49.

[7] T. W. Ruger, P. T. Kim, A. D. Martin, and K. M. Quinn. 2004. The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking. *Columbia Law Review* 104, 4 (2004), 1150–1210.