

In order for legal practitioners and scholars to have a correct understanding and application of existing case-law, their mapping of existing precedent ought to be comprehensive. Interdisciplinary endeavors towards creating such a mapping have involved the study of citation networks of court judgments from the U.S. Supreme Court, the World Trade Organization and the European Court of Human Rights, among others. Such quantitative approaches can reveal at once the motivations behind the citation of previous judgments and potential areas of the jurisprudence that have gone understudied in legal textbooks that attempt to catalogue important case-law.

Existing work on citation networks, either legal or academic, have focused on the possibilities of performing statistical modeling on the network. A dataset of connected texts, in the hands of a computer scientist, invites link prediction experiments. Can we predict which texts ought to be cited by future legal cases and academic articles? Or given a social graph, can we predict which users ought to connect? Literature on this task is plentiful and useful; lawyers or applicants to a court can use a recommender system to inform their search of existing case-law; researchers can contextualize their work or discover previous related articles.

Modeling a citation network can also divulge valuable information about the data, beyond what links ought to exist. The approach I take involves modeling the citation network of judgments from the European Court of Human Rights in order to reveal which textual features of the associated judgment texts are highly predictive of citation. The intuition behind this method relies on an assumption in which ngrams that are stronger predictors of citation are also descriptive of the court's jurisprudence. A more extensive explanation of how this method and how it is evaluated is fleshed out below.

The Court and the data:

Officially inaugurated in 1959, the establishment of the European Court of Human Rights is meant to apply the articles of the European Convention for the Protection of Human Rights and Fundamental Freedoms, entered into force in 1953. Today, the combined instruments of the convention and the court, understood as a regime, is considered the most effective in the world. The legal dynamics of the regime draws

comparisons to a bill of rights (the convention) interpreted and enforced by a nation's constitutional court (Strasbourg).

The entirety of the data used for my experiments consist of a corpus of judgments texts from the European Court of Human Rights and a network representing these judgments and the citations between them. These two datasets were generously provided by the iCourts research centre at the University of Copenhagen. An additional dataset employed in my analyses is a corpus of publications covering the case law related to particular articles of the European Convention of Human Rights.

The judgment texts total ___ documents, mostly in HTML format. A look at how these judgments are distributed across time reveals a much more active court in recent years: ___ percent of the data consists of judgments from after 2010, ___ percent from after 2000 while the first 20 years of the court's history produced ___ judgments, representing ___ percent of the corpus.

A typical judgment contains sections on the circumstances surrounding the case and covers relevant jurisprudence from national courts. A following section discusses the purported right violation and how similar cases have been dealt with in the past. Dissenting opinions have been filtered out. Sections that contain more standardized language and legal boilerplate have not been preprocessed, assuming that the model will handle their prevalence adequately.

I have collected a corpus of relevant publications that are affiliated to the court without being binding on its judges. One series, Guides on the Convention, is prepared by the Research and Library Division within the Directorate of the Jurisconsult and is made available on the ECHR website. A second series, Human Rights Handbooks, is made available by the Directorate General Human Rights Council of Europe. Both series cover important cases, discuss how the convention article has been interpreted, and provide insights on how the article ought to be implemented.

Methodology

I use stability selection combined with L1 logistic regression models to identify variables that are highly indicative of citation.

The estimation of unknown parameters in a logistic regression model is done with conditional maximum likelihood estimation. The weights for features will attempt to perfectly fit details of the training set, in fact too perfectly, modeling noisy factors that just accidentally correlate with the class. So regularized or penalized methods are needed to fit the regression model to the data, and to keep the variance of the regression coefficients under control. To avoid overfitting a regularization term is added.

L2 regularization is easier to optimize because of its simple derivative (the derivative of w^2 is just $2w$), while L1 regularization is more complex (the derivative of $|w|$ is noncontinuous at zero). But where L2 prefers weight vectors with many small weights, L1 prefers sparse solutions with some larger weights but many more weights set to zero. Thus L1 regularization leads to much sparser weight vectors, that is, far fewer features. A limitation of the Lasso method occurs when there is a group of highly correlated variables. In this case, the Lasso tends to select one variable from a group, ignoring the others. This is overcome with stability selection, as the model is reestimated many times, resulting in a set of stable variables

Stability selection is essentially a randomization technique involving the application of a selection algorithm, such as logistic regression. Rather than being applied to the whole dataset, stability selection re-estimates the model many times over different random subsets of the data, drawn without replacement. After many iterations, the results are aggregated by detecting which variables are selected at a higher frequency over the subsamples.

Features

The estimator upon which stability selection is performed in order to return ranked predictors is limited to training on features whose interpretation can help explain the court's case-law. Thus, features pertaining to textual similarity between judgment pairs, citation network metadata and node importance measures are left out. These additional features are used in experiments conducted to gauge the performance of the logistic regression models involving only ngram feature vectors. As a sanity check for my central experiment, I've

included any feature that may also be predictive of citation. These include:

- The textual similarity between citing and cited judgment, based on the cosine similarity of either document's tf-idf weighted vector representation.
- Hub score of the citing judgment in the network and the authority score, eigenvector centrality and indegree centrality of the cited judgment.
- Metadata from the citation network: number of days between cited and citing judgments, match between respondent countries of either judgment, number of shared article violations involved in either case, and the importance rating, provided by the court.

Experimental set up

For every article of the convention and its associated subgraph of the citation network and judgment texts, stability selection is performed in order to retrieve a ranked list of terms (unigrams, bigrams or trigrams) corresponding to predictors of citation. In order to validate the scores attributed to text features, this ranking is compared to the appearance of predictors in the literature on the case law of the court. To get an idea of why citation happens, the ranking is also compared their appearance in either the 'Fact' section or 'Law' section of the judgments.

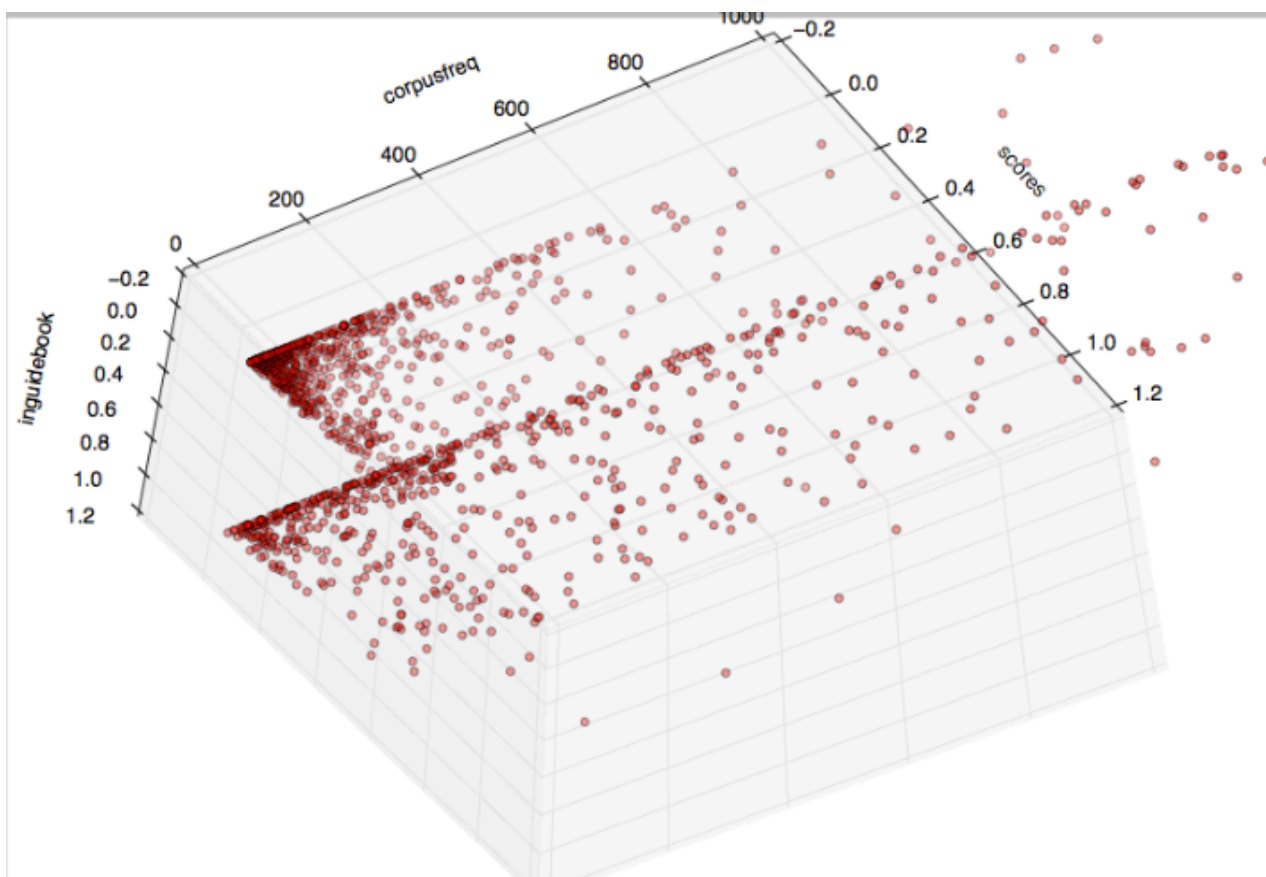
The performance of the predictors trained on ngrams only is also compared to predictors trained on textual content, metadata and network structure features.

Following the advice of experts in the field of European human rights law, I've given particular attention to my results for article 10, "freedom of expression". Textbooks have apparently done a better job of describing the case law for this area comprehensively. Whereas I have automated the comparison of predictor rankings and their occurrence in the literature for all other articles, a manual comparison, with a more careful consideration for redundancy in the rankings and synonymy between ngrams is included below.

Results

Freedom of assembly	0.4 . 0.16, 0.13, 0.08, 0.01
Private life	0.55, 0.25, 0.14, 0.11, 0.08, 0.04, 0.02
Conscience and religion	0.41, 0.34, 0.28, 0.2
Democratic society	0.38, 0.33, 0.29, 0.28, 0.27, 0.20, 0.12, 0.1, 0.05
Progress	--
Individuals self fulfillment	--
Notwithstanding its content	'content' 0.05, 0.01
Racism, nazi ideology	--
Hatred and racial discrimination	'hatred' 0.47, 0.02
Theory of the paradox of tolerance	'tolerance' 0.44, 0.08, 0.07
...	...

Key concepts from guidebook on article 10 (left) and the stability selection score of matching or similar trigram predictors (right)



Scatter plot of 1-predictor strength, 2-predictor's frequency in the judgment corpus and 3-Whether (1) or not (0) the predictor shows up in guidebook