

Predicting Judgment decisions using Natural Language Processing

CS671 - Natural Language Processing

Abhishek (14017)
Bhaskar (14178)
Mayank (14378)

Akhilesh (14053)
Devendra (14214)
Shreyash (14666)

- Legal institutions are typically overburdened with cases.
- New Delhi High court estimates that courts would take **466 years** to rule all pending cases assuming same rate.
- There is significant scope for applying NLP techniques to predict judgments.

Objective of this study

- Predicting violations of Articles 3,5,6,8 for cases ruled by European Court of Human Rights.

- 4178 cases from European Court of Human Rights
- Regarding violations of Human rights Article 3,5,6,8
- Prediction is made using court judgments as lodged complaints are inaccessible
- It is assumed that judgments are sufficiently similar to complaints and are unbiased.

Number of Cases verses Article of Convention

Article No.	Article	No. of Cases
3	Prohibits torturous & degrading treatment	1151
5	Right to liberty & security	946
6	Right to a fair trial	1319
8	Right to privacy	762

Table: No. of cases for each article

	Violations	Not Violations
Article 3	591	560
Article 5	509	437
Article 6	754	565
Article 8	411	351

Table: Violation and Not Violation cases for each article

- **Procedure:** from individual complaint to court verdict
- **Facts:** Content formulated by court besides written law
 - **Circumstances of the case**
Factual background and procedure by lower courts (by ECHR)
 - **Relevant Law**
Relevant legal provisions of the domestic courts.

- **The law:** Merits and legal arguments of the case.
 - **Alleged violation of article x**
 - **Parties' submissions**
Articulates primary arguments by the plaintiff and state.
 - **Merits**
Legal reasons to justify an outcome
- **Operative provisions:** Outcome of the case regarding violation of the article

Selection of **feature representations**

- **Tfidf**

Features for Procedure, Facts, The Law, Full case

- **N-gram**

Features for Procedure

- **Important words and topics**

- Use cluster (topics) weights as given by Aletras et al. [1] (2016)
- Construct weights for each word, such that higher weight indicate better predictive performance in specific article violation.

Selection of Classification models

- **Neural network**

We train a FF-NN to learn labels for tfidf and N-gram features.

- **Support vector Machines**

A linear kernel is used to identify tfidf and N-gram features.

- **Prediction Scores**

We formulate a predictive score for each paragraph. +ve score indicates violation and vice versa.

- **Min-max** $\rightarrow \max S_N + \min S_N$
- **Aggregate Sum** $\rightarrow S = \sum_N S_{para}$
- **Weighted Sum** $\rightarrow S = \sum_N S_{para} * (word_size)$

Example

Score = 2.477

7. The sentence was imposed on him by the Criminal Court on 14 March 2012 for drug related offences. The Criminal Court also ordered him to pay a fine of 23,500 euros (EUR) as well as EUR 1,462.94 in fees payable to the experts, which were to be converted into a further term of imprisonment if not paid. On 13 April 2012 the aforementioned sums were converted into 490 days' imprisonment. Following an amnesty and payment of the fee for the experts, the applicant was released on 14 April 2016 and flown back to Costa Rica after being given the requisite vaccinations.

Score = -0.841

Moreover, it is essential that the prevention of suicide, including the identification of those at risk, should not rest with the health-care service alone. All prison staff coming into contact with inmates – and, as a priority, staff who work in the reception and admissions units – should be trained in recognising indications of suicidal risk. The sharing of information concerning suicidal tendencies with prison staff can be considered as an ethical necessity in light of the possible consequences that inaction may entail. In this connection, it should be noted that the periods immediately following admission to prison as well as before and after trial and, in some cases, the pre-release period, are associated with an increased risk of suicide.

- We calculate accuracy is measured as:

$$Accuracy = \frac{TV + TNV}{V + NV}$$

where TV and TNV are the number of cases classified correctly regarding article **violation** and **no violation**.

and V and NV are the total number of cases with a **violation** and **no violation** respectively.

Results

Case Structure	Metric	Article 3	Article 6	Article 8
Procedure	Max/min	65.7	58.32	53.05
	Sum Compare	66.78	57.18	50.13
	Size Weighted Sum	66.51	55.74	49.73
Facts	Max/min	60.64	56.92	54.88
	Sum Compare	60.99	56.54	54.19
	Size Weighted Sum	60.18	56.39	53.92
The Law	Max/min	54.59	56.62	51.85
	Sum Compare	53.87	57.22	48.83
	Size Weighted Sum	53.96	56.92	48.84
Full Doc	Max/min	57.16	56.86	53.14
	Sum Compare	54.12	57.24	50.26
	Size Weighted Sum	55.25	57.31	49.88

Table: Accuracies for SVM weights extracted from [1]

Features extracted from the paper [1] for

- **Article 8:**

son, body, result, russian, department, death, group, relative, head, described, men, deceased, town, attack, died etc.

- **Article 3:**

injury, protection, ordered, damage, civil, caused, failed, claim, course, connection, region, effective, quashed, prison, detainee, visit, well, regard etc.

- **Article 6:**

instance court, caused, time limit, individual, responsible, receipt, court decision, copy, employee, time, article, convention, January, human, lodged, domestic, February, September etc.

Results(Contd...)

	Article →	3	5	6	8
Procedure	Max/min	81.22	80.86	81.67	71.07
	Sum Compare	81.57	81.08	81.90	71.27
	Size Weighted Sum	81.40	80.65	81.52	72.07
Facts	Max/min	66.87	71.80	68.34	71.11
	Sum Compare	68.59	74.33	71.23	75.37
	Size Weighted Sum	68.86	74.10	71.68	73.31
The Law	Max/min	73.12	71.47	72.07	70.79
	Sum Compare	78.35	76.74	78.38	76.44
	Size Weighted Sum	75.02	73.76	77.54	74.49
Full Doc	Max/min	72.98	75.89	79.22	70.99
	Sum Compare	75.15	73.28	80.00	74.14
	Size Weighted Sum	72.54	79.06	79.57	74.27

Table: Accuracies for calculated SVM Weights

Features corresponding to calculated SVM weights for

- **Article 8:**
breach, disclosed, legal, allocated, granted, informed, requested, respect, assigned, domestic, information, contrary etc.
- **Article 3:**
abduction, diagnosed, sexual, non pecuniary, HIV, confession, nationals, expulsion, drug, ethnic, disappearance, illness, regional, etc.
- **Article 6:**
republic, rules, appeal, representative, pursuant, inadmissible, informed, public, observations, etc.
- **Article 5:**
investigators, removal, suspension, abduction, damages, exhaustion, injuries, administrative, compulsory, hospitalization etc.

Results(Contd...)

Classifier	Metric	Article 3	Article 5	Article 6	Article 8
NN	Procedure	92.48	92.92	93.41	92.97
	Facts	90.92	90.05	91.69	91.00
	The Law	91.80	91.23	92.77	92.04
	Full Doc	92.16	92.24	93.60	93.12
SVM	Procedure	96.31	96.41	95.74	94.28
	Facts	92.96	92.63	94.14	91.76
	The Law	94.86	94.04	95.05	93.41
	Full Doc	94.17	94.82	95.67	94.48

Table: Binary classification for Tfidf representations

Results(Contd...)

	No. of epochs	Accuracy
Article 8	2	63.54
	5	66.43
Article 3	3	61.67
Article 5	3	62.17
Article 6	3	64.10

Table: Binary classification for N-Gram representations

- N-grams representation used with $N = 5$
- Only **Procedure** sub-section was analysed because this section had performed better in previous models

- The section **Procedure** has higher density of informative features than other sub-sections.
- Thus it has better accuracy than other sections for the Paragraph scoring Classifier.
- For **Tfidf** repn. the accuracy of **full** case representation was comparable to **Procedure** due to more uniform representation (it takes into account of the frequency of words.)
- One important reason for the poor performance of **Law** sub-section is that a large number of cases does not include this subsection

Comparative Analysis

- In contrast with Aletras et al. [1] a larger data set was used in our study hence the features obtained were more rich and informative.
- Aletras et al. [1] had equal weights for each feature within cluster, we have different weights for each feature (gives better scores)
- The section **Procedure** shows the best predictive performance. Aletras et al. [1] had **Circumstances**. Thus the significance of a section changes with size of dataset, choice of word representation and classifier model

Conclusions and Extensions

- This approach involves binary classification for article violation. Such an analysis can be extended to include sentence prediction.
- Significant accuracy was achieved. This analysis can indicate qualitative patterns like performance of the section **Procedure**.
- This kind of approach can be significantly improved as per the availability of inaccessible data, as we got improved results with a larger labelled dataset.

References



Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoțiuc-Pietro, and Vasileios Lamos.

Predicting judicial decisions of the european court of human rights: A natural language processing perspective.

PeerJ Computer Science, 2:e93, 2016.



Reed C Lawlor.

What computers can do: Analysis and prediction of judicial decisions.

American Bar Association Journal, pages 337–344, 1963.



Benjamin E Lauderdale and Tom S Clark.

The supreme court's many median justices.

American Political Science Review, 106(4):847–866, 2012.

[1] [2] [3]

Thank You