
Escalus: Overview

SangHyeon(Alex) Ahn
E-mail:alexahn@jhu.edu

Abstract

We construct a predictive model based on machine learning techniques to predict judicial outcomes on cases handled by United States Court of Appeals for the Federal Circuit. We explore various feature engineering methods and learning algorithms to aptly predict the judicial predilections on different cases. Using data from various sources including opinions and oral arguments made for previous cases, sources of origination and dates related with cases, and biographical information on the judges and the lawyers, we build a unique and novel method to forecast each cases' outcomes. We extract, transform and load relevant data from different sources (web, PDF, and MP3 files) to construct a baseline for our prediction model. We will select the best performing machine learning algorithm by measuring performances through various methods including cross validation and AUC analysis. We further explore uncovering patterns and predilections through PCA and CCA methods.

1 Introduction

Often, judicial decisions are constructively made through careful analysis on each case's circumstances and legal arguments from both parties. Usually, the court's decisions are very hard to predict due to its complexity regarding many different aspects such as past decisions, legal issues, social issues, political issues, personal preferences, individual interpretation and so on. Hence, traditionally, predicting judicial outcomes has been one of the most challenging problems for not only the non-professionals but also the law firms and lawyers.

However, using a large spectrum of data ranging from historical judicial outcomes to stakeholders' biographical information, we seek to construct an accurate, comprehensive and informative model predicting judicial outcomes even better than humans using technology: machine learning.

Especially, we begin by focusing on cases brought to United States Court of Appeals for the Federal Circuit because they are more technical, more geographically universal, less predictable by human intuition (i.e. having more weight on personal interpretation than legal), data available, and more specific in a broader sense.

We aim to uncover predictability of a judicial outcome, significance of specific feature dimension, patterns and predilections, and historical lessons on judicial decisions through multiple analysis.

1.1 Problem

The hardship of predicting judicial outcomes centers on degree of complexity in delivering decisions by the judges. Granted, the law plays the central role of a decision. However, different oral arguments made by different lawyers significantly effect an outcome, a past decision on similar case creates a strong bias, and each judge's personal preferences make an impact on favoring a decision over another. Therefore, we need to comprehend the problem through a variety of dimensions including case information, lawyer's and judges' biographical information (personal preferences), past decisions, text and oral arguments and more. For a baseline feature composition we include the

following aspects: Case Information, Case Opinions, Case Oral Arguments, Lawyer Biographical Information, and Judge Biographical Information.

1.2 Solution

We train a model using appropriate machine learning algorithm using constructed data instances of the past cases (labeled instances).

We predict an outcome of an (unlabeled) case instance using the trained model.

We provide specific feature dimensions that cause the most variance in terms of decision outcome (PCA).

1.3 Approach

ETL:

1. Extract: construct a relational database from various sources (web, PDF, and MP3 files) by extracting relevant features.
2. Transform: script an autonomous program to convert raw data into relevant feature spaces and combine into a full case instance.
3. Load: Sample Train/Development/Test data samples to train, construct, and test a model to predict judicial outcomes.

Attorney's Win Rate We calculate the win rate of an attorney's prior several years before a case is argued. For each case, we make an update to the case's attorney's number of wins/losses on both sides of appeal. Calculation efficiency: $O(N)$.

Trial Attorney We need to be able to identify which attorneys defended for both parties of appeal court at the trial court. This will be quite instructive for us to capture data on the lawyers and law firms who handled the underlying case at trial. We will utilize the resources for this particular information from websites (<https://www.plainsite.org>), and using Queries.

Rule 36 The most straightforward method we can use to determine if a Federal Circuit Rule 36 Judgment involves a patent dispute, and if so, what patents are involved, is to search for the underlying District Court opinion using one of several websites offering free, comprehensive access to federal court decisions, or summaries of such decisions, such as

1. Court Listener's RECAP service (<https://www.courtlistener/recap/>)
2. Justia (<https://dockets.justia.com>)
3. Law360 (https://www.law360.com/advanced_search/casesorhttps://www.law360.com/advanced_search)
4. Leagle (<http://www.leagle.com/casesearch?casename=&citation=&anyword=&dateon=&qsearchsubmit=1>)
5. Plainsite (<https://www.plainsite.org/cases/index.html>)

So, for example, if we wanted to find out if the Rule 36 Judgment in the Healtht trio v. Aetna decision referenced above involved a patent dispute and if so, which patents, we could search for the names of the parties (Healtht trio and Aetna), Federal Circuit case number (2016-1034) and/or District Court case number (1:12-cv-03229) using any of the 5 websites listed above, and we would quickly determine that this is a patent case, and the numbers of the 10 patents at issue in the case.

Once we determine the patent numbers involved in each of the the Federal Circuit's Opinions and Rule 36 Judgments on appeal from District Courts, we should export those patent numbers to our database. The task then becomes very simple to capture the remaining data on these patents that we are interested in analyzing, because the United States Patent & Trade-mark Office (USPTO) website has an excellent search engine (<http://patft.uspto.gov/>

metahtml/PTO/search-bool.html>) that allows you to search for patents using the patent number and then capture information about the patent from any of the 56 different data fields into which every patent is divided (<http://patft.uspto.gov/metahtml/PTO/help/helpflds.htm#Current_US_Class/SubClass>). For example, if we wanted to search for the 6,772,132 patent referenced above in the Trading Technologies, Intl v. CQG, Inc. opinion issued by the Federal Circuit on January 18, 2017, we would use the USPTO search engine and input the patent number 6,772,132 in the field "Term 1" and then select "Patent Number" from the drop down menu in "Field 1." This search will take you directly to the specified patent on the USPTO website, where we can then export all the data fields from the patent that we need.

PTAB We identify the patents that have been previously challenged and retrieve relevant information. <<https://ptab.uspto.gov/#/login.>>

2 Data Model

Machine Learning algorithm and prediction model description.

2.1 Feature Engineering

Judge	Lawyer
Judge.Identification.Number	Lawyer.Last.Name
Judge.Last.Name	Lawyer.First.Name
Judge.First.Name	Birth.year
Chief.Judge (Y/N)	Place.of.Birth..City.
Chief.Judge.Dates	Place.of.Birth..State.
Confirmation.Date	Birth.year
Commission.Date	Race.or.Ethnicity
Senior.Status.Start.Date	Gender
Termination.Date	Name.of.School
Termination.Reason	Degree
Prior.Professional.Career.1	Degree.year
Prior.Professional.Career.2	Name.of.School..2.
Prior.Professional.Career.3	Degree..2.
Birth.year	Degree.year..2.
Place.of.Birth..City.	Name.of.School..3.
Place.of.Birth..State.	Degree..3.
Race.or.Ethnicity	Degree.year..3.
Gender	Practice.Areas
President.name	State.Bars.Admitted
Party.Affiliation.of.President	State.Bar.Admission.Dates
Renominating.President.name	Total.wins
Party.Affiliation.of.Renominating.President	Total.wins.1.year
Nomination.Date.Senate.Executive.Journal	Total.losses
ABA.Rating	Total.losses.1.year
Name.of.School	Total.win.percentage
Degree	Total.win.percentage.1.year
Degree.year	
Name.of.School..2.	
Degree..2.	
Degree.year..2.	
Name.of.School..3.	
Degree..3.	
Degree.year..3.	

Opinion	Case Information
Law Type	Fed. Cir. Case No.
Tribute Type	Plaintiff
Date of appeal (filed)	Defendant
Writing Judge	Appellant
Writing Judge Law Clerk	Appellant.Attorney.lead
Harmless Error	Appellant.Attorney.team
Reversible Error	Appellee
Per Curiam	Appellee.Attorney.lead
Rule.36	Appellee.Attorney.team
District/Trial Court	Cross-Appellant
District Court	Cross-Appellee
District Court No.	Amicus Curiae
District Court Judge	Location
District Court Jury	Public or Private for each party above
Date of District Court Decision	Sales for each party above
Trial.Appellent.Attorney	Stock Ticker Symbol for each public party above
Trial.Appellent.Attorney.Total.wins	Market Cap for each public party above on date before decision
Trial.Appellent.Attorney.Total.wins.1.year	Market Cap for each public party above on date after decision
Trial.Appellent.Attorney.Total.losses	Market Impact of Decision
Trial.Appellent.Attorney.Total.losses.1.year	Law Firm for each party above
Trial.Appellent.Attorney.Total.win.percentage	Declaratory Judgment
Trial.Appellent.Attorney.Total.win.percentage.1.yr	Summary Judgment
Trial.Appellee.Attorney	Summary Judgment Grounds
Trial.Appellee.Attorney.Total.wins	Monetary Award
Trial.Appellee.Attorney.Total.wins.1.yr	Injunction Issued D.Ct.
Trial.Appellee.Attorney.Total.losses	Market Cap for each public party above on date before D.Ct. Decision
Trial.Appellee.Attorney.Total.losses.1.yr	Market Cap for each public party above on date after D.Ct. Decision
Trial.Appellee.Attorney.Total.win.percentage	Market Impact of D.Ct. Decision
Trial.Appellee.Attorney.Total.win.percentage.1.yr	Pendency of Case
Timeline	Previous Fed. Cir. Appeal
Date of Complaint	Previous Prevailing Party
Date of Notice of Appeal	Supreme Court remand
Date of Argument	Supreme Court Prevailing Party
Date of Decision	Circuit Judges
Time to Decision	Per Curium
	En Banc
	Recusals
	Opinion By
	Concur By
	Dissent By
	Patent Nos.
	Patent Law
	Judgment

Patent

Applicant City
Applicant Country
Applicant Name
Applicant State
Applicant Type
Application Date
Application Type
Assignee City
Assignee Country
Assignee Name
Assignee State
Assistant Examiner
Attorney or Agent
Cooperative Patent Classification
Current CPC Classification Class
Current US Classification
Government Interest
International Classification
Inventor City
Inventor Country
Inventor Name
Inventor State
Issue Date
Patent Number
Primary Examiner
Priority Filing Date
PTAB Trial Certificate
Re-Examination Certificate
Re-Examination Certificate Date
Referenced By
Patent References Cited
Non-Patent References Cited
Total References Cited
Novelty Ratio
Patent Maturity in Months
Title
Prior Challenge (Y/N)

PTAB

Trial Number
Prosecution Status
Petitioner Party
Patent Owner Name
Inventor Name
Filing Date
Accorded Filing Date
Institution Decision Date
Last Modified Date
Claims Cancelled
Claims Confirmed
Number of Claims Cancelled
Number of Claims Confirmed
Percentage of Claims Cancelled

Oral Argument

Time in mins. of Appellant's Opening Argument
Time in mins. of Appellee's Argument
Time in mins. of Appellant's Rebuttal Argument
Questions asked of Appellant in Opening Argument
Questions asked of Appellee
Questions asked of Appellant in Rebuttal Argument
Questions/min. of Appellant in Opening Argument
Questions/min. of Appellee
Questions/min. of Appellant in Rebuttal Argument
Total Questions of Appellant
Total Questions of Appellee
Appellant Counsel Word Count
Appellant Counsel Words/Min.
Appellee Counsel Word Count
Appellee Counsel Words/Min.
Presiding Judge Word Count
Judge.2
Judge.3
Relative Heat Index
Relative Nervousness Index
Appellant Word Frequency Vector
Appellant Number of judge interruptions
Appellee Word Frequency Vector
Appellee Number of judge interruptions
Number of questions asked by judge
Judge Voice Amplitude
Judge Voice Tone
Judge Voice Decibel

3 Prediction Model

Machine learning algorithms and prediction models descriptions.

3.1 Candidates

Binary Classification Algorithm candidates.

1. SVM Pegasos
2. SVM Kernel Method
3. Boosting Method (AdaBoost)
4. Decision Tree
5. Random Forest

3.2 Comparisons

Algorithms results.

Relevancy:

Efficiency:

3.3 Performance Evaluations

Cross Validation.

Accuracy Comparisons.

ROC AUC analysis.

4 Citations

Citations.

4.1 Sources

1. US Court of Appeals for the Federal Circuit
<<http://www.cafc.uscourts.gov/>>
2. Case Opinions (PDF files)
<<http://www.cafc.uscourts.gov/opinions-orders>>
3. Case Oral Arguments (MP3 files)
<<http://www.cafc.uscourts.gov/oral-argument-recordings>>
4. Case Oral Arguments (enhanced MP3 files)
<<https://www.courtlistener.com/audio/>>
5. Judge Biographical Information
Web: <<http://www.fjc.gov/history/home.nsf/page/judges.html>>
Exportable: <<http://www.fjc.gov/history/export/jb.txt>>
6. American Bar Association
<http://www.americanbar.org/groups/committees/federal_judiciary/ratings.html>
7. Attorney Biographical & Rating Information
<<http://www.martindale.com/Find-Lawyers-and-Law-Firms.aspx>>

5 Footnotes

Footnotes.

Acknowledgments

Acknowledgements.

References

References.