Precedent vs. Politics? Case Similarity Predicts Supreme Court Decisions Better Than Ideology

Elliott Ash, Daniel Chen, Shivendra Panicker, and Akshay Trivedi

June 25, 2017

Abstract

Using the universe of U.S. Circuit Court cases appealed to the Supreme Court since 1946, we show that case similarity based on Circuit Court opinions achieves better prediction accuracy of Supreme Court decisions relative to the current best prediction model, which is based on ideology of judges and trends of how they vote. Relative to the benchmark prediction accuracy of 59%, textual measures of case similarity achieve prediction accuracy of 64%. We interpret this improvement to suggest that precedent matters more than politics alone. We also offer our model available as a web application.

1 Introduction

Previous work on predicting Supreme Court outcomes has chiefly involved estimating court, justice and case level attributes for a case. However, these models don't take into account outcomes of similar cases in the past. On the other hand, in a typical court-room argument it is a common practice of lawyers to cite past cases that could influence outcome of the current hearing. Judges also do the same while justifying the analysis of their judgment. Relying on this observation, we base our approach of predicting Supreme Court outcomes on identifying similar cases in the past.

Instead of using supreme court oral transcripts, we rely on an alternative data source of circuit court opinions to retrieve related cases. We chose circuit court opinions dataset since, when judges write an opinion, they justify their ruling in a case. Therefore each circuit court opinion is a rich data source containing concurring and dissenting opinion of the judges; statutory and factual background information about the case, analysis used to arrive at the conclusion and final conclusion of the case. In the section stated below, we describe in detail how this textual information helps us in retrieving prior cases.

2 Model

2.1 Features

We construct following features which help us capture the notion of similarity among cases.

1. Latent Dirichlet Allocation
   We create an *case topic model* using Latent Dirichlet Allocation(LDA) where each circuit court argument is considered to belong to a set of issues pertinent to the case. The topic distribution is assumed to have a sparse Dirichlet prior. The sparse Dirichlet priors encodes the intuition that each case covers only a small set of key issues and each of those issues are expressed in a fixed vocabulary of legal terms.

2. Citations
   The vast majority of prior cases are not explicitly cited in the opinion of the current case. This is partially due to the fact that some of the prior cases may not have been published at the time that the current opinion was written. On the average, explicit citations occur about 10% of the time in a typical sample of state and federal opinions. However, similar cases generally cite the same sources while presenting analysis of their judgment. So using citations helps us in retrieving similar cases.

3. Case Title
   Case titles can be composed of people (with or without their professional titles), government or private agencies, companies, country/state/city/town, and so on. The names in a case title may not be enough to retrieve prior cases, if any. For example, an instant case with the title "State v. Smith" would retrieve thousands of cases with the same or similar titles; though all of these cases may not be similar in terms of issues involved. In order to generalize the type of litigants involved in a case, four types of party entities are extracted- people, companies, places,

and government agencies. Thus, different phrases such as "David E. Smith" and "Antony Davis" map to an identical entity.

4. Issue
   This variable identifies the issue for each decision. The focus here is on the subject matter of the controversy (e.g., sex discrimination, state tax, affirmative action). The variable codes 260 issues, each of which has an identifying number.

5. Issue Area
   This variable simply separates the issues identified in the preceding variable (issue) into the following larger categories. It encodes 14 broad categories such as criminal procedure, civil rights, First Amendment etc.

6. Legal Provision Considered by the Court
   This variable identifies the constitutional provision(s), statute(s), or court rule(s) that the Court considered in the case. They identify the specific law, constitutional provision or rule at issue (e.g., Article I, Section 1; the Federal Election Campaign Act; the Federal Rules of Evidence). This variable takes on 206 categorical values each identifying a legal provision.

2.2 Similarity score

Given any two cases, we generate their d-dimensional vector representations $\mathbf{x}$, $\mathbf{y} \in \mathcal{R}^d$ based on the features stated above. Let $\mathbf{W} \in \mathcal{R}^d$ be the corresponding set of weights for each of the d-features. Assigning weights to features allows flexibility in modeling relative feature importance among each of these features. We assign weight to a feature by measuring its discriminatory power or the amount of information it contains.
Now the similarity score $\mathbf{S}$ simply computes the weighted cosine similarity between two vectors as follows:

$$\text{Similarity Score } \mathbf{S}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^{n} W_i x_i y_i}{||\boldsymbol{x}|| \cdot ||\boldsymbol{y}||}$$

2.3 Predicting supreme court outcomes

This module takes a list of ranked prior cases based on cosine similarity and decides which of these candidates, if any, is a true prior. To achieve this, it applies two thresholds. The first threshold, is an absolute threshold on the similarity score. The second threshold, is a relative threshold from the highest scoring candidate. For example, a prior case candidate with a similarity score of, say, +3 is not very likely to be one of the true priors if another candidate for the same case scored a 10, even though it may be the next most similar case. Thus the first threshold limits the number of cases the system suggests, while the second limits the number of suggestions made per case. The values of both of these thresholds were set empirically based on cross validation. On the cases that are suggested, we use a Majority Voting Classifier of their outcomes to predict the current outcome.

3 Experimental Setup

3.1 Data

Supreme Court Database(SCDB) features more than two hundred years of high-quality, expertly-coded data on the Court's behavior. Each case contains as many as two hundred and forty variables, including chronological variables, case background variables, justice-specific variables, and outcome variables. Since we are interested in retrieving prior cases and using them to predict outcomes we make use of case background variables and outcome variables from SCDB. These include: Issue, Issue Area, Legal Provision and Case outcome Disposition.

We restrict our analysis to those supreme court cases that have been heard in federal circuit courts due to our choice of using circuit court opinions. This amounts to 4339 cases from 1946-2013. We utilized docket number as the crosswalk to map a supreme court case to its corresponding circuit court opinion.

3.2 Data cleansing

The statutory and factual background information in a circuit court opinion contains details too specific to a case. Training text-based models on this results in over fitting them as most of these details are generally unwarranted in considering similarity to prior cases. We extract 'Abstract', 'Opinion' and 'Conclusion' from each opinion as these present a broad overview of the case.

To further enhance data quality, we filter out nouns and adjectives from the text as these generally pertain to the issue involved in the case and their severity. This was done using NLTK's Stanford POS tagger. Additionally, to discard multiple occurrences of the same root word, we use NLTK's Snowball Stemmer.

3.3 Model

1. LDA To choose the number of topics on which LDA should be trained, we used 5-fold cross validation set and selected the number of topics that minimized the perplexity measure. Topics proportions are extracted for each case from LDA model and these are used as features in judging similarity among documents. On our corpus of circuit court opinions, selecting number of topics to be 100 minimized perplexity.

   <Photos of LDA.>

2. Citations Typical citations generally include references to cases from Supreme Court, Court of Appeals, District Courts and United States Codes. We further extracted citations of the above four types discussed above. This accounted for a total of 51k citations out of which 6k citations were cited in more than two cases. All citations within a case belonging to this subset were included as one-hot encoded categorical features in the vector representation of a case.

3. Title All word sequences delimited by the string “v.” were grouped together for parsing purposes. Based on heuristics, all sequences of up to four capitalized unknown words are parsed as names. These were then mapped to respective generic types- people, companies, places, and government agencies. After feature construction and categorization, each case was represented as a 6660 dimensional feature vector.

4. Feature Encoding(Issue, Issue Area..)

3.4 Baseline

We compare our results to two different baselines. The first, REVERSE, is a majority class baseline which assumes all votes to be reversed by supreme court. The second, KATZ, is Katz et al's model which reported predictive accuracy of 69.7% on average from 1953 to 2013, which is to the extent of our knowledge the best predictive accuracy on SCOTUS prediction yet achieved in the literature. Katz et al report on the entire corpus of supreme court cases from 1953-2013, however we restrict the data set to those supreme court cases that have been appealed in the circuit courts. On this dataset, their model gets an accuracy of 59%. The third, PRIOR, is our model based on the setup described above. For the sake of comparing the three models, we also split out the results into individual terms.

	REVERSE	KATZ	PRIOR
2008	74.46	59.57	72.91
2009	82.29	56.02	80.64
2010	56.60	56.60	55.55
2011	56.73	56.73	54.71
2012	71.67	67.86	70.37
2013	68.01	68.01	72.91
Overall Accuracy	-	59.01	64.91

4 Observations

We also developed a web application that shows 5 most-similar cases for any case as ranked by our model, PRIOR. In order to visualize the quality of the retrieved results, we present each case with its word cloud containing the most important words in its circuit court opinion.

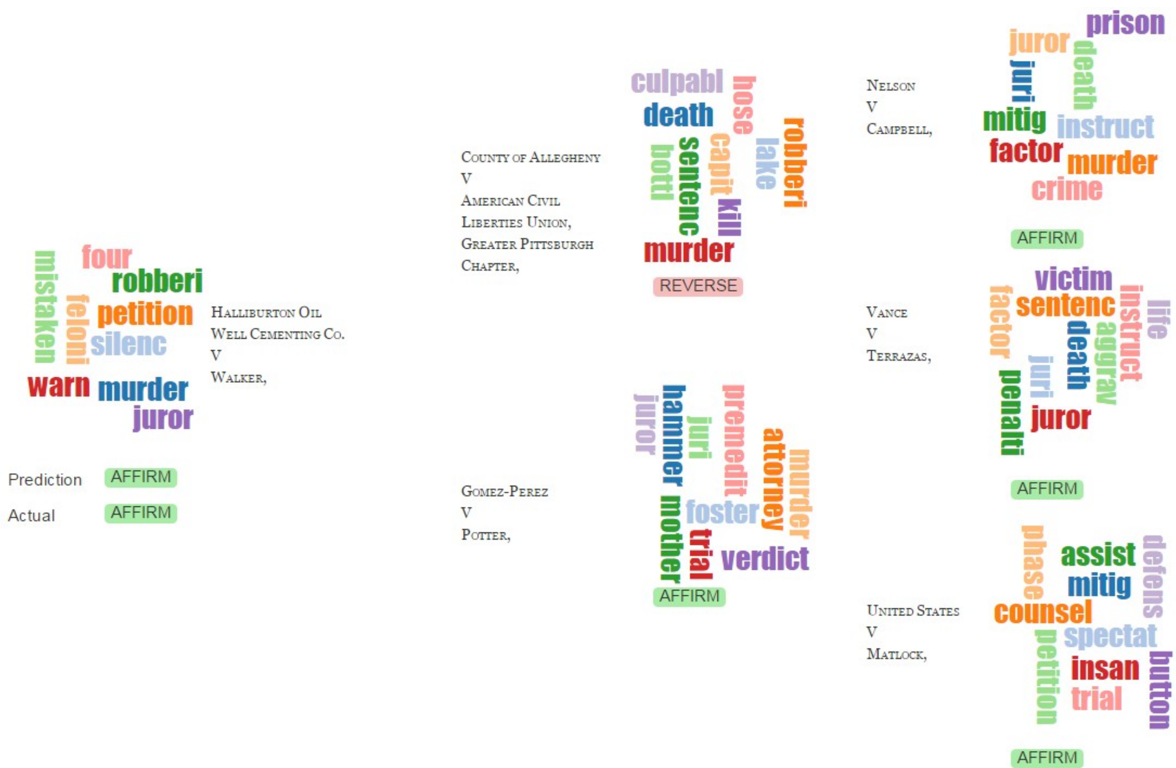

Figure 1: Word clouds of 5-similar cases where outcome was in direction of prior cases

Our model works well in retrieving similar cases in the past. As we can see, all the cases in the example above involve 'murder' and 'robbery'. We also show the decision given by the Supreme Court outcome for each case as 'AFFIRM' or 'REVERSE'. In this example, Supreme court follows the precedent set by similar cases in the past and affirms the lower court's decision. Our model predicts it correctly as we take Majority Vote among prior similar cases.

To interpret where our model goes wrong, we present another example. Again our model works well in retrieving prior similar cases as can be seen from word clouds of the retrieved cases. All the cases in the example below, center around 'racial discrimination of employee'.

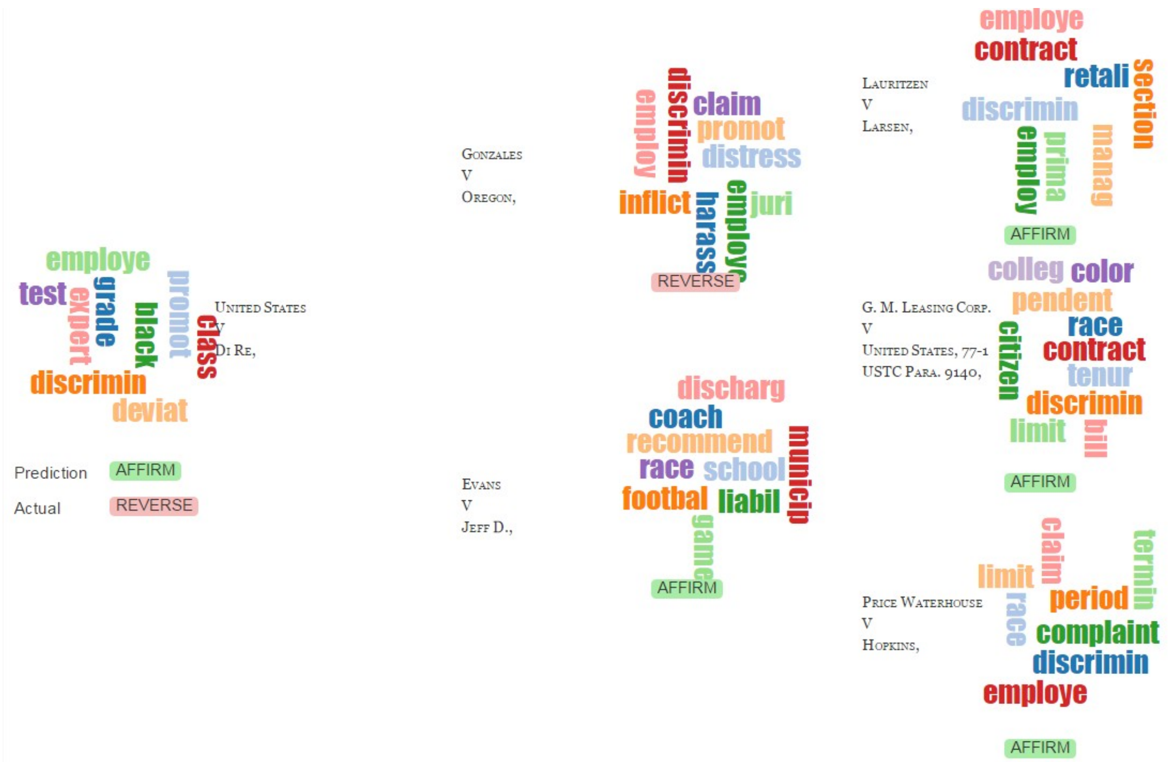

Figure 2: Word clouds of 5-similar cases where outcome was in opposite direction of prior cases

Interestingly, the Supreme Court reverses the judgment of lower court in this case rather than affirming the judgment as had been its standpoint for similar cases in the past. Since our approach assumes precedent influences current decision making our model gets this wrong. Inspite of this, we note that such cases should be of increased interest to Supreme Court watchers and legal pundits as they imply a drift in Supreme Court's earlier stand-point.

5 Conclusion

We have developed a novel model for predicting Supreme Court cases. This new model provides a new and interesting way of analyzing Supreme court outcomes. Our approach not only predicts Supreme court outcomes but also lends itself very intuitively in interpreting its results by offering insights into the behavior of Supreme Court compared to its historical standpoint. This is our unique contribution as previous attempts on Supreme court prediction offer black-box predictions of Supreme court outcomes.

In doing so, we also present an efficient system that retrieves past cases that are most similar to a case. Our code is available online at <> and our web application is hosted at <>.

Bibliography

Khalid Al-Kofahi, Alex Tyrrell, Arun Vachher, and Peter Jackson. A machine learning approach to prior case retrieval. In *Proceedings of the 8th international conference on Artificial intelligence and law*, pages 88–93. ACM, 2001.

Tara Balakrishnan, Paula Kusumaputri, and Luda Zhao. Predicting supreme court votes through conversational dynamic features.

Daniel L Chen, Manoj Kumar, Vishal Prabhu Motwani, and Philip Yeres. Is justice really blind? and is it also deaf? 2016.

Daniel Martin Katz, Michael James Bommarito, and Josh Blackman. Predicting the behavior of the supreme court of the united states: A general approach. 2014.