

2/6/2019

The purpose of the study is to analyze court documents, specifically judge decisions, then to identify clusters.

I arrived at this topic because I was primarily interested in NLP and text analysis. In order to choose a corpus, I considered topics that I am interested in. I chose the judicial system and decided to look at Supreme court texts.

I remembered last fall when President Trump criticized "Obama judges". Chief Justice John Roberts responded that there is no such thing. But is there? This is an area where I can do more research. First, I have to understand what an "Obama judge" looks like. These are presumably judges that were nominated into either the Supreme Court, Appellate Courts, or District Courts by President Obama. So other than this characteristic ("i.e. the President who nominated these judges"), how else are they unique? Do they share certain biases? Do they rule in favor of certain policies? ~~Are they more likely to take more specifically~~

① What is an "Obama judge"? ~~\* Ideology, Time~~

A concern is that clusters may identify generations (time-clusters) i.e. judges who ruled on "issues that were prevalent in the 60's". So we may see clusters of ~~just decisions that were made~~ judges who lived in a certain era, instead of judges who share an ideology.

SQ the word, "ideology" is important. When I say "Obama judge", I'm not talking about judges who were active in the same time period. I'm interested in whether these judges during this time period shared a common ideology, which is common amongst other judges who were nominated by President Obama.

This highlights the need to identify ideological clusters within the same topic. That is, text clustering decisions will identify different genres of the law, different legal topics. This tells us little about political ideology. I need to ask whether within decisions that rule on a legal topic, such as abortion, do justices that were <sup>appointed</sup> ~~appointed~~ by Obama ~~use~~ use similar texts ~~compare to those~~ amongst each other?

③ How distinct are legal topics? Are there clear "abortion" cases and clear "bankruptcy" cases?

This question is important because it may allow me to identify ideological ~~the~~ differences (president-based clusters) within topics.

Oh but how about time within topics?

(It may not be so interesting to see that Reg. Dem. judges)

③ Within legal topics -- do differences in ideology due to time interfere meaningfully with differences in ideology due to ~~the~~ ideology?

So the core question is...?

What do I really want to know?

~~Are~~ Are Judges BIASED ~~or~~ ?

And if a set of decisions by "Obama judges" are similar to each other because of time ... what does it mean? It depends!

How is time relevant? If the decisions cluster because a certain issue was simply discussed more frequently during a certain era (i.e. transgender, first amendment, etc.) then they are more time-effects than ideology-effects.

## time-effects vs ideology-effects

### in own group variation

Time-effects: decisions that are "different" textually because they are simply from different eras

Ideology-effects: decisions that are "different" textually due to a judge's bias

Time-effects do impact ideology-effects in some ways because a person's bias may be affected by the time period in which they ~~grew~~ grew up. But within a certain year, i.e. 2004, the difference in texts would be based more clearly on ideology. It's not just because the issue was talked about more in 2004.

If not "bias", then what?

• Am I looking for "bias"? "Bias" implies prejudice to the law or unfairness.

A set of "Obama judges" may rule similarly ...

2/10/2019

$$\Pr(y_{i=1}) = \text{logit}^{-1} \left( \gamma_{k[i]} \left( \underbrace{\alpha_{j[i]}}_{\text{"discrimination"}} - \underbrace{\beta_{k[i]}}_{\text{"case position"}} \right) \right)$$



Then justice "j" is not likely to cast vote "i" on case "k". (Assume  $\gamma_{k[i]} > 1$ ).

$\Pr(y_{i=1})$ , the probability of a politically conservative outcome is low.

The purpose of the study is to try to use the text of <sup>judges</sup> options to identify ideology, ideal points. So are the liberal? Are they conservative? Are they something in between?

So Martha Minow uses the voting behavior of judges to place them on a liberal/conservative spectrum. Their votes are effectively reduced to a single dimension representing a latent variable that summarizes the judge's ideology.

So, the goal is to do the same with text. (AN ~~the text~~)  
I use the text to represent judges on a single dimension?

- 
- How should I handle different topics in different areas?
  - Maybe I should start with a couple decisions first. I'll have more questions once I start looking deeper at the data.
- 

Why am I doing this?

Ideal point models exist and they ~~will~~ place justices on a left-right spectrum in a way that seems to make sense (at least based on our familiarity with the courts). But what happens when we look at the text?

---

So what are you going to do with the text?

I want to do clustering (represent documents as vectors) ... i.e. reduce the dimensionality of the texts in order to represent judges in reduced-dimension space AND hopefully these reduced dimensions will do a good job of representing political ideology.

This may fail if ~~judges~~ judges systematically rule on different causes ... then, this latent trait may more accurately describe the time or the location of the justice rather than their ideology.

So I should compare within a circuit court or within the Supreme Court whatever has a larger sample size.

---

The purpose of my project is ideal point model. What is an ideal point model? Estimate judge's ideology using dimension reduction on text. With something like Martin-Oquin Scores, a vote expresses a judge's position. The text is the work but should be more subtle. ~~that may describe~~ considerations The "subtle" details within the text is foreign to me. It's a black-box approach ... and I don't like the idea of that too much...

It shouldn't be too much of a problem as long as I account for obvious ~~play~~ distortions. For example, the effects of the circuit effects. Within

In the 9<sup>th</sup> circuit, there are ~ 900 cases available per year. (29 justices currently serving).

Yours idea was to identify clusters of presidents then to see how these "president cohorts" match with clusters of judges. In effect this is strongly different "schools of thought" or ideology profiles using presidents then trying to see if that matches. An advantage of using presidents is that each judge has a president who nominated them. So when I evaluate whether clusters exist,

So ideal part and / ~~not~~ political ideology indicated  
from ~~not~~ groups of products.

⇒ Type of Party ~~→~~ ~~→~~

2/12/2019

- I'm not married to text analysis... I shouldn't do it unless I have good reason to suspect that the language / text of judge's decisions will reflect ideology.
- Also, I don't really understand Martin-Score methodology. I could explain ideal-point models on a high-level, but is ~~it necessary?~~ this even that necessary to know? Then I won't be using this ~~text~~ for text.

① Why would text be useful for reading ideology?

- ↳ Scalia, an originalist, might cite precedents that adhere more closely to originalist ~~text~~ interpretation.
- ↳ I should ① research this, ② look at examples from my law-related classes on examples where judges had different interpretations of the law.

2/15/2019

The unintroductory part about the new Clerk-based ideology measures of judge ideology is that it's just replication of a data linkage problem.

The ~~data~~ replication isn't totally mundane... it does require some "fuzzy matching" (I wonder if they use Levenshtein algorithm).

Anyway, linkage isn't too introducing.

But once the data have been linked, there are more interesting questions. (In a way, it's like how lawyer reported measures for each judge, where each "measure" is a law-clerk history.

Banica uses those histories to construct judge ideology. How exactly does Banica compute these values? ( $\rightarrow$  there a better methodology to do this?)

Could you also include votes to improve the computation of the ideology scores?

What else can I learn about judge ideology using clerk political contribution data?

