Affirm? Reverse?

WHY DO COURT DECISION REVERSALS HAPPEN? (FROM DISTRICT TO CIRCUIT COURT)

Amanpreet Singh and Sharan Agrawal

NYU

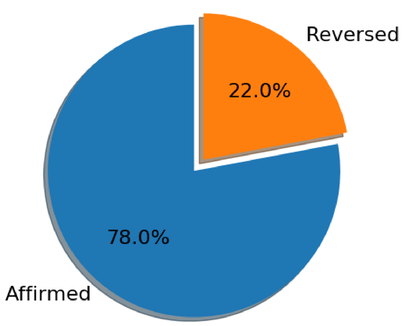
GOAL

Every year more than 300,000 civil and criminal cases are heard in the district courts all over the US. Less than 5% of these cases are appealed and heard in circuit courts. For most of the cases, the circuit court either affirms the decision of the district court or reverses it. In this project, we have attempted to predict reversal of the district court opinion by the circuit court using the summarized opinion of the district court.

DATA

Our data consisted of 2 separate data sets:

1. Raw records of approximately 750,000 U.S Circuit Court Opinions from 1880 to 2013 (Both published and unpublished). This complete and comprehensive dataset was cleaned, parsed and processed to find matching district court case using case name, circuit number and date published.
2. Raw text of approximately 280,000 District Court opinions from 1924 to 2013. This dataset, even though incomplete, is a strong representative of the decisions of the district court. A CSV file describing these district court cases containing case metadata. Ngrams were filtered out of these.
3. A DTA file containing information about 4096 judges. This information was used to filter out district court judges from the CSV file and match them with the details mentioned in this file.



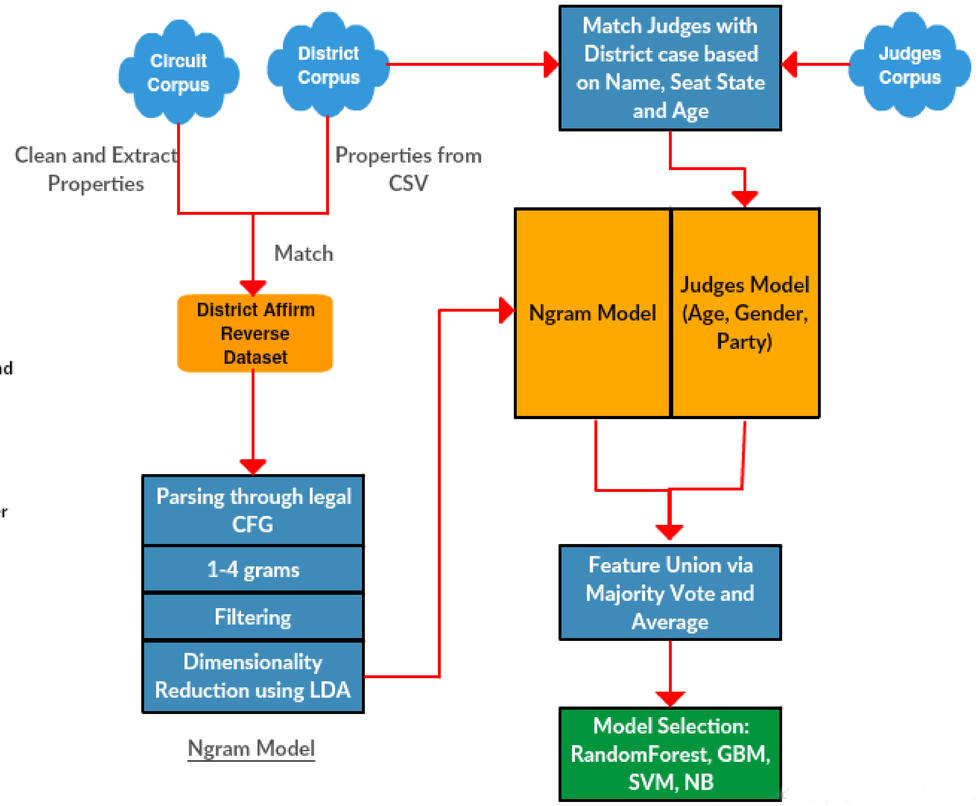
PIPELINE AND MODEL

DATA:
- Circuit Court data
- DIstrict Court Data
- Judge data

PREPROCESSING:
- Match circuit to district cases
- Match judges to district cases
- Filter ngrams and reduce number using frequency and dimensionality reduction
- Oversampling

FEATURES:
- Ngram
- Judges age, region, gender and party affiliation

MODEL:
- Train different kind of classifiers to find the best AUC and Accuracy

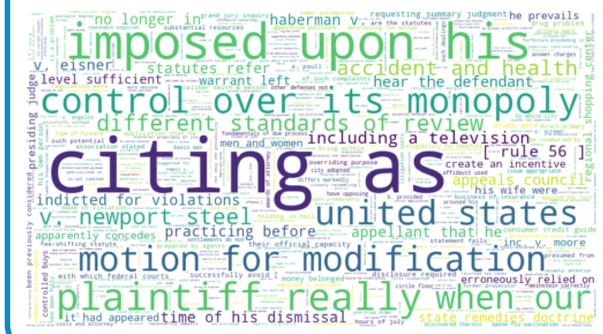


IMPORTANT WORDS

We did POS tagging on opinion texts and filtered out ngrams using a context free grammar purposed for legal language:

S	⇒	TWO | THREE | FOUR
TWO	⇒	A N | N N N | . . .
THREE	⇒	N N N | A A N | . . .
FOUR	⇒	N C V N | A N N N | . . .
A	⇒	JJ | JJR | JJS
N	⇒	NN | NNS | NNP | . . .
V	⇒	VB | VBD | VBG | . . .

A word cloud of important words is shown here. Size of word is scaled to the weight of importance it has.



FEATURE ENGINEERING

There were two main types of feature engineering involved:

1. Using memetic phrases (ngrams) from filtered raw text
   (a) Ngrams: To filter out insignificant ngram, we took $\frac{1}{8^{th}}$ most occurring and least occurring ngrams. Further, we applied LDA to reduce dimensionality.
2. Judges biography: After matching judges based on judge songername and surname to district court cases, we add extra features to our ngrams by following:
   (a) Seat State (Majority of the judges are in 'DC' and 'NY')
   (b) Gender (Most of the judges are male, ratio M:F is 7:1)
   (c) Party Affiliation

RESULTS

We trained a wide variety of classifiers discussed in the class on our data . We have taken the AUC score and accuracy as the performance measures for our classifiers. We used a train validation split of 10% and obtained the following results:

Classifier	AUC	Accuracy
Gradient Boosting (RT)	0.64	0.76
Multinomial NB	0.62	0.73
Random Forest Classifier	0.61	0.71
LinearSVC Classifier	0.59	0.68
Logistic Regression	0.58	0.70

We observed that GBM with regression trees obtains maximum scores. We applied oversampling and feature reduction techniques to improve the AUC score.

CONCLUSION & FUTURE WORK

We obtained best performance with a Gradient Boosting classifier using Regression Trees. However, model still has a lot of scope for improvement. More data can be collected from some of unpublished cases. Other algorithms like SMOTE can be used to handle imbalance of data for "Reversed" court cases. In case of feature, other judge features and citation (which we currently don't have can be added).

Future work on this will involve using representation learning (Word2Vec, Glove) for training classifier on deep neural nets as it might be true that opinion text of district case is not really a strong factor of reversal in circuit court. DNN will help us discover underlying relationships which we haven't discovered yet.