What Matters: Agreement Between Circuit Court Judges?

Xing Cui, Lanyu Shang, Junchao Zheng@ NYU Center for Data Science

NYU DATA SCIENCE

Contributions

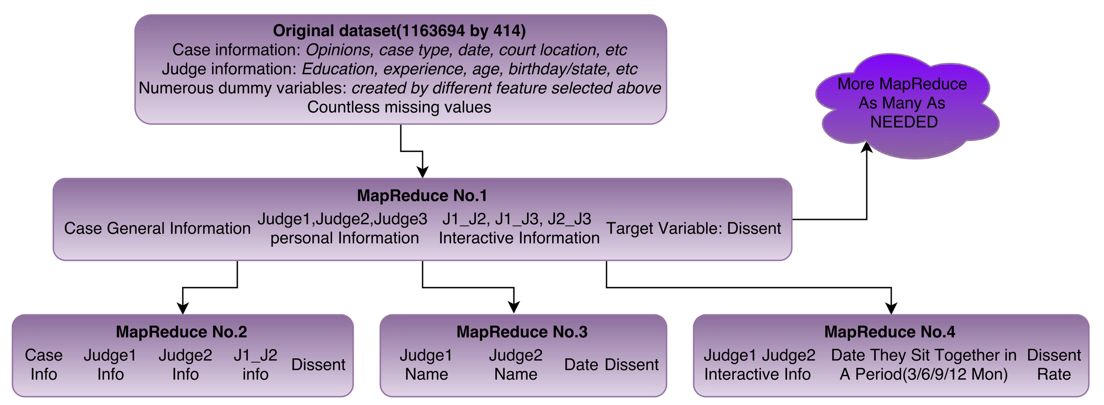
We present an application of Machine Learning study for the federal circuit courts. Our goal is to see what matters when a judge disagrees with another in the circuit and to predict how likely two judges would disagree with each other in the future.

Introduction

Today, the federal court system is formed by three basic levels that are one U.S. Supreme Court, 13 U.S. Courts of Appeals, and many U.S. District Courts. Federal circuit courts are the intermediate appellate courts of the federal court system. It is not like the Supreme court who hears less than 100 cases a year and is also not like District Courts who meets more than millions cases a year. Also, federal circuit has 179 judges who are nominated by the president, confirmed by the Senate, and authorized by the Congress.

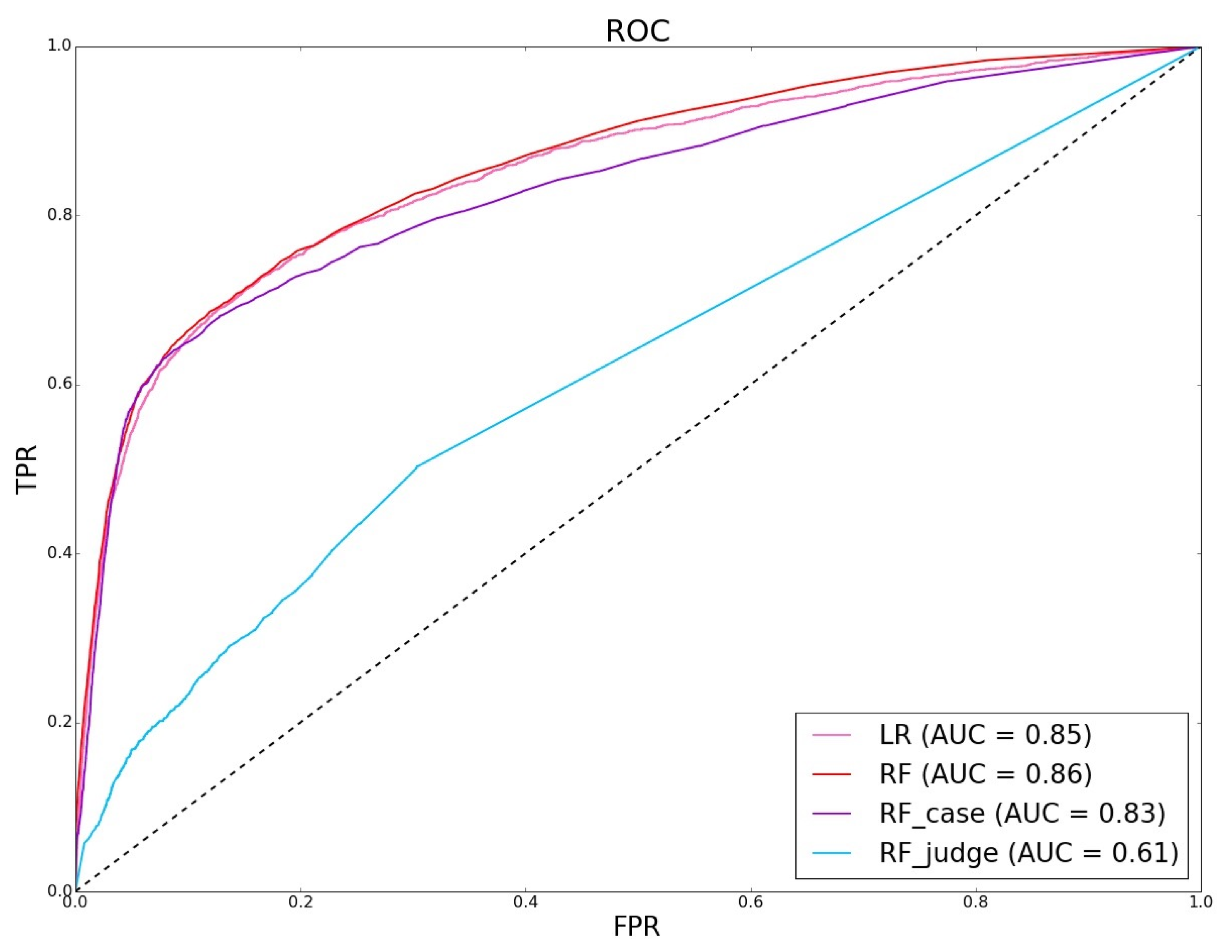
Data Interperation

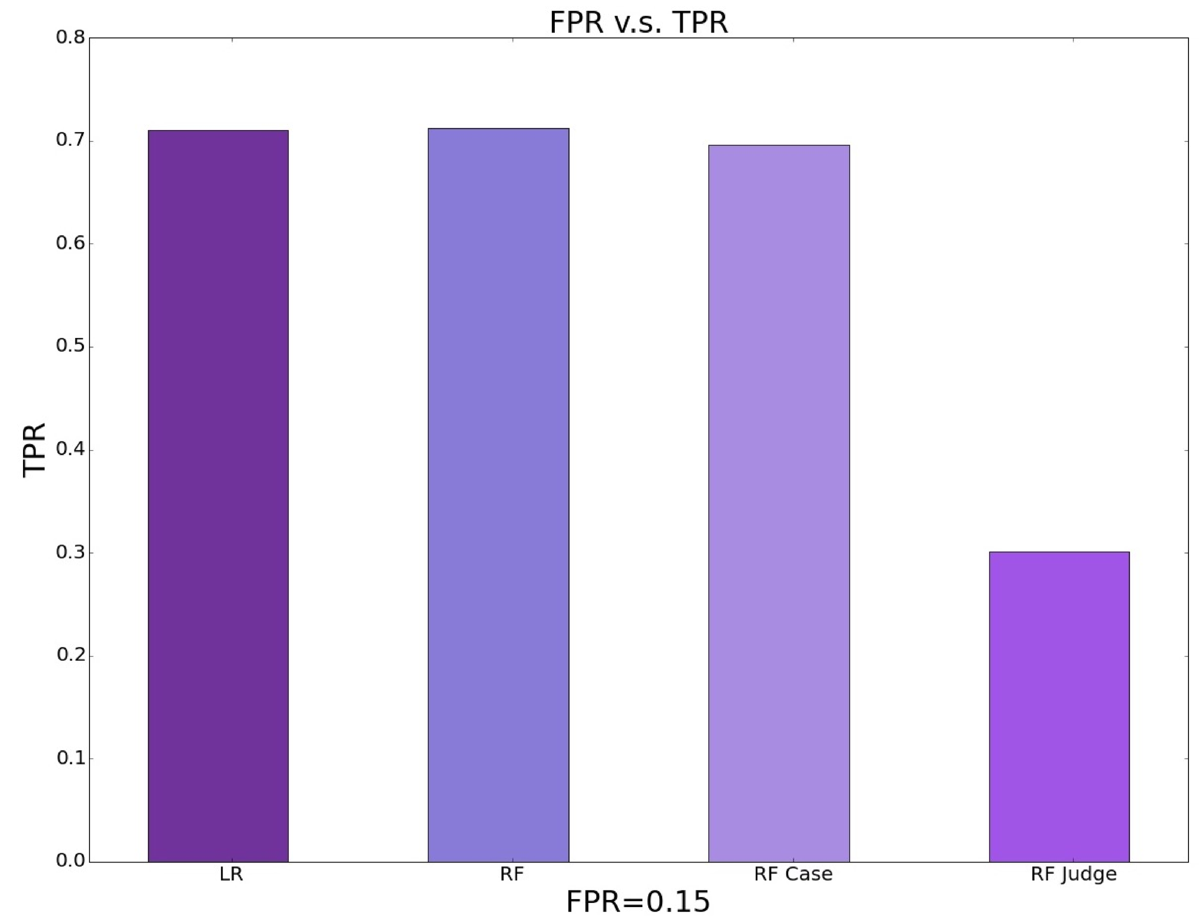
The 100VoteLevel_touse dataset has 1163694 data entries and each has 414 features. For each case, the dataset contains three data entries that have information of the case and three judges. We first fill reasonable values to missing fields. In order to have a durable and efficient dataset, we use MapReduce to reframe the dataset that merge three case entries into one with three judges personal information and interactive data separately. Therefore, we finally have case categories:



Finally we have the data from the first MapReduce with 868962 entries and 2215 features. From this dataset, we can whenever extract whatever information we need.

Model Fitting

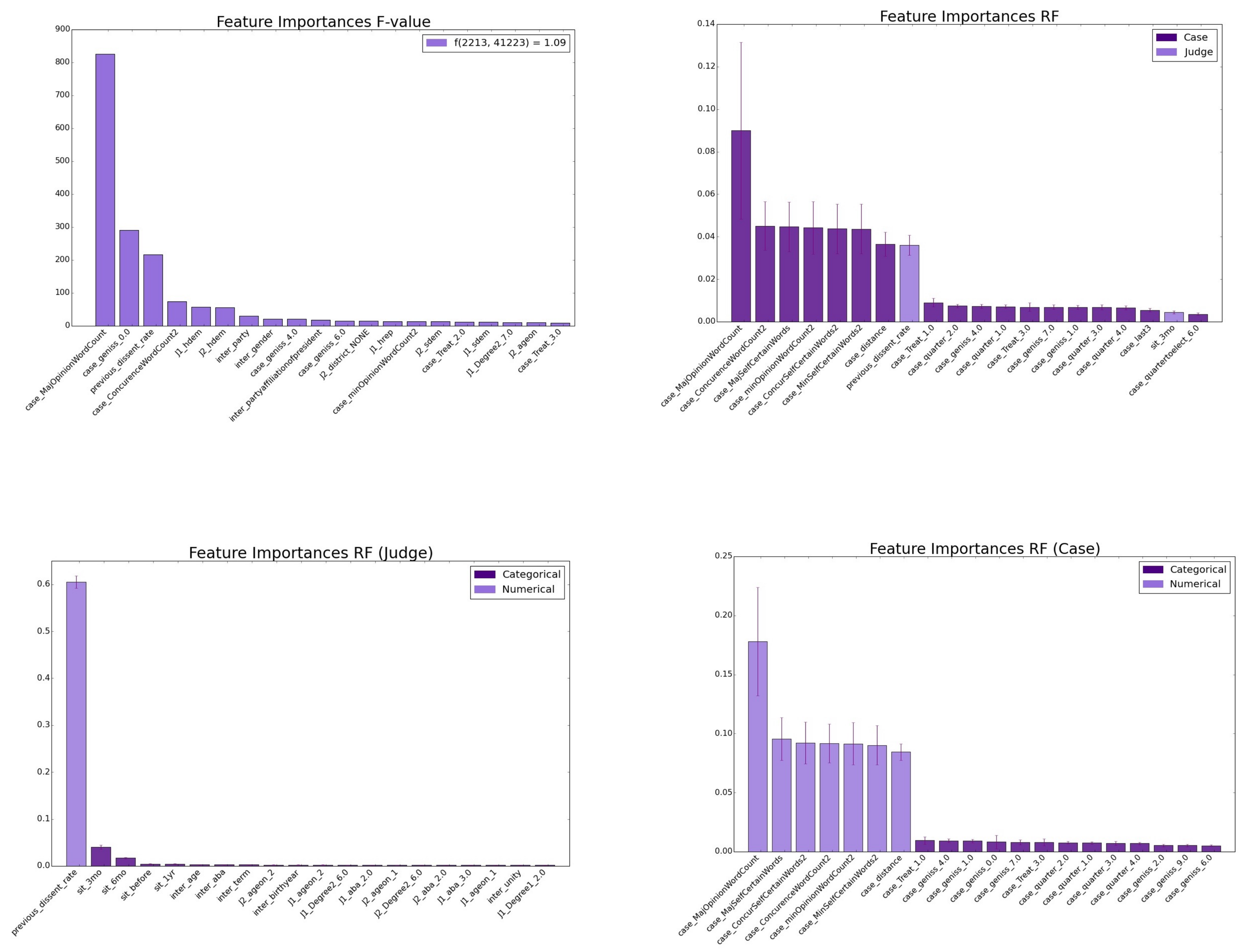




The dataset is highly unbalanced as there are only 5% target with value 1(disagree) while 95% with value 0(agree). We choose specificity(FPR) as our evaluation standard. As majority of courts reach an agreement, we focus more on what happens when two judges in a court disagree with each other, so we require a high confidence of prediction on negative results.

We choose Random Forest as the final model with parameters n_estimators = 500 and max_features = 'sqrt'. Since we have 2214 features, we generate 500 decision trees that each tree has 47 random features. Thus, each feature will appear in 10 or more trees. We could generate more trees but the running time will expand too much. We modify the threshold to find optimal TPR and FPR. We choose 15% FPR and 70% TPR as a good balance here.

Feature Ranking



In Random Forest, there may be one problem that most of the important features are features with numerical features. Considering the features are well normalized, we may draw a conclusion that numerical features may count more than binary features in model fitting, though we are not sure about the deduction.

In our F-test on Logistic Regression, any feature with a F-value larger than f(2213, 41223) = 1.09 should be a significant feature. We get 185 significant features.

case_MajOpinionWordCount: The number of words in major opinion which is presented to judges for their decision making.

case_geniss_0: The topic of this case. 0 refers to criminal case.

previous_dissent_rate: The historical disagreement rate when this two judges sit together.

case_ConcurenceWordCount2: The number of words in concurence which is presented to judges for their decision making.

Conclusion

1. We obtain an optimal Random Forest model and set specificity(FPR) as our evaluation standard. We achieve a FPR = 0.15 and a TPR = 0.70.
2. By applying F-test, we get 185 significantly important features over 2000 features, which include: *case_MajOpinionWordCount*, *previous_dissent_rate*, *case_geniss_0*, *inter_party*, *inter_gender*.
3. From the result we can roughly excogitate a conclusion: judges are righteus and impartial since decisions are mostly based on case information. Personal information like party they belong to or their age make slight influence on their judgement.

Future Work

So far we are running the model on 20% of the whole dataset due to a maintenance in NYU HPC server. Another problem is that due to the large dataset, running the model is very time consuming. We will try to speed up by running the model using Spark library on the whole dataset.

We can develop the accuracy more by generating more power features. For example, we can get more detailed features like the historical disagreement ratio on different court topics.

Acknowledgement

We would like to thank Dr. Daniel L. Chen for providing data source and guidance on the project and Professor David Rosenberg for instruction and efforts on Machine Learning course.

Reference

[1] The Judicial Research Initiative (JuRI)ă at the University of South Carolina *http://artsandsciences.sc.edu/poli/juri/sct.htm*

[2] Berdejó, Carlos and Chen, Daniel L. Priming Ideology: Electoral Cycles without Electoral Incentives Among U.S. Judges *2013*