Tom Stuckey
605.744 SP16
Semester Project Overview
19 March 2016

Classifying US Supreme Court Legal Opinions

1 Problem and Motivation

US Supreme Court legal opinions are interesting documents. Aside from the legalese, they present an interesting problem space with respect to text classification. The US Supreme Court's legal decisions are particularly interesting given the context the nine judges are appointed by the President and given the far reaching implications of their decisions. In 2016 with the combination of the US Presidential race and the recent passing of Justice Antonin Scalia, this project is focused on developing a model to classify these legal opinions and present the political distributions across each opinion.

Supreme Court opinions have two main areas: the Case and the Opinion. The Case refers to the topic in question. After the Supreme Court agrees to hear a Case, they issue an Opinion in a single document. The Opinion is the more interesting area as it has up to three parts. First, the Main Opinion is officially presented by a single Justice and the Disposition is one of three categories: 1. Affirm: uphold the lower court's ruling 2. Reverse/Void/Vacate: overturn the lower court's ruling 3. Remand: send it back to the lower court for retrial. Second, if any Justices agree but have a different legal grounds, they provide Concurring Opinions that align with the Main Opinion but present different rationale. Third, if any Justices disagree, they provide Dissenting Opinions.[1]

This project leverages *ground truths* from the bar association to help identify features to feed into a multi-classification Support Vector Model (SVM) to classify the component parts of the US Supreme Court's legal opinions to identify the Justices wherever possible. Separate, external lookup information is included to provide each Justice's appointing President and party.

2 Approach

This section provides an overview of the data, the classification methodology, and the overall statistics to be provided in the final project.

2.1 Data

The legal opinions are obtained in PDF form using the Supreme Court's RESTful web service[2], and they are converted from PDF to text via the Apache Tika utility.[3] For purposes of training

[1] American Bar Association. (2012, September). How to read a Supreme Court opinion. Retrieved March 19, 2016, from http://www.americanbar.org/publications/insights_on_law_andsociety/13/fall_2012/how_to_read_a_ussupremecourtopinion.html

[2] 2003-2011 opinions: http://www.supremecourt.gov/opinions/<2-digit year>pdf
2012-2015 opinions: http://www.supremecourt.gov/opinions/slipopinion/<two-digit-year>

[3] https://tika.apache.org/

the model, validating the model, and testing the model, the source set of legal opinions (~67MB of text) from 2003-2015 is being allocated 80%, 10%, and 10% respectively.

Seperately, a small lookup table has been built to map each of the thirteen Supreme Court justices involved over the period of the dataset (2003-2015) to their appointing President and political party.

2.2 Classification Methodology

The core of this project is the classification methodology for the Opinions. This project employs multi-class Support Vector Machines (SVMs). Leveraging the *ground truths* as established by the American Bar Association and other legal resources, features are used in conjunction with hand-written rules to identify, for each of the primary categories, **Main Opinion**, **Concurring Opinion**, **Dissenting Opinion** whether a given section of text *is* or *is not* a member of the category. For those sections of text identified a being a member of a the **Main Opinion**, there is another set of 2-class classifiers against the the categories of Disposition: ***Affirm***, ***Reverse/Void/Vacate***, ***Remand***.

Again this methodology is being employed on the data set with the following allocation. 80% to train the model, 10% to validate the model, and 10% to test the model.

2.3 Statistics

The following statistics are to be provided at the conclusion of this project across the entirety of the dataset:

- Number of Opinions
- Mean word size of Opinions
- Number of Unanimous Opinions
- Summary of each Justice:
    - % Main Opinion
    - % Concurring Opinions
    - % Dissenting Opinions

3 Relevant Legal and Scientific Research

A number of legal sources, existing semantic legal processing works, and Support Vector Machines are being leveraged to facilitate this project. These are enumerated in the final report.