Machine Learning Final Project

Q. How do presidential candidates speak differently about trade within rust belt regions?

(1) Differences in political party

(2) Differences in before vs. after primary

(3) Differences in swing vs. non-swing states

What we did so far

* Downloaded the presidential election speech of every candidate (both before and after primary election) in 2008, 2012, and 2016
* Language Processing (stem words, stop words, crosstalk, laughter, boos…)
* Ran STM with 45 topics (Kuk, Seligsohn, Zhang 2017), with the focus on rust belt regions (10 states: NY, PA, WV, OH, IN, MI, IL, IA, WI, MI)
* Gathered swing state data

Result

**Trade related topic words [topic 10]:** trade, agreement, worker, manufactur, china, corpor, plant, enforce, industry, Youngstown

**Labor related topic words [topic 11]**: retir, senior, social, save, pension, secur, benefit, match, invis, account

**Labor related topic words [topic 28]**: union, card, labor, worker, corpor, compact, ceo, pension, bankruptcy

**Capital related topic words [topic 30]:** bank, growth, regulatori, financi, institut, capit, wall, street, rate, fed

**‘Promise by Money’** **topic words [topic 43]:** Obama, tax, senat, spend, promis, drill, dollar, rais, joe, spread

Next Step

* Comparison between swing state & rust belt and non-swing state & rustbelt regions
* Comparison between speech of a candidate before and after the primary election
* Extract the text that mentions the word ‘trade’ at least one time, and conduct positive and negative word analysis

Questions

* How can we define and extract party-neutral words?
* What do you think about hand-coding of documents that do not specify location in the title? (350 out of 1397 documents)
* Where can we get the dataset on the dictionary of positive and negative words?
* A foreign word in STM



