Computational Analysis of Religious and Ideological Linguistic Behavior

University of Virginia, MS in Data Science Capstone

Samantha Garofalo, Seth Green, Leigh Harton, Megan Stiles

Advisor: Don Brown
Client: Peter Ochs
Sponsor: ARL

Problem: How to work with Value-Based Groups?

What is a Value-Based Group?

- *Any group motivated by a set of values*

Who needs to work with Value-Based groups?

- *US Government*
- *Humanitarian Groups*
- *Business Leaders*

Large-Scale Goals:

- *Real-time recommendations to organizations on the ground*

About Our Data

Input

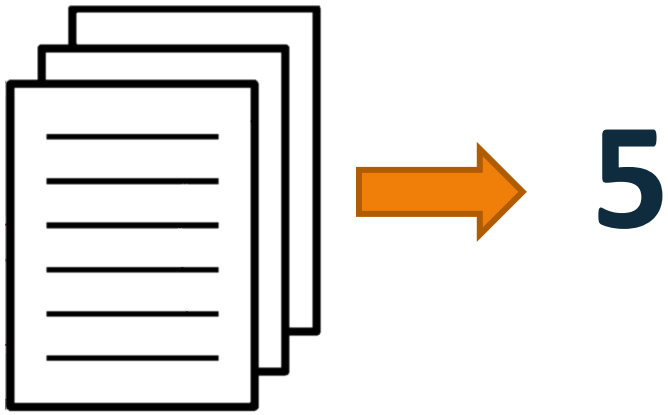
- *Raw ideological text*

Output

- *Number on Linguistic Rigidity Scale*

Sources:

- *Last year's corpus*
- *New Groups*
  - Web Scraping
  - OCR for PDFs



Data Wrangling

- *Stemming*
- *Part of Speech Tagger*
- *Appropriate Article Length*

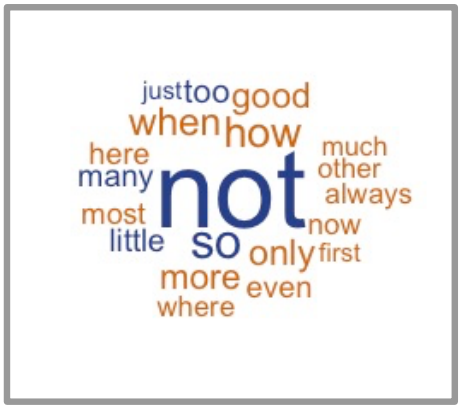
```
def cleanTokens(tokenList):
    #Convert all text to lower case
    textList=[word.lower() for word in tokenList]

    #Remove punctuation
    textList=[word for word in textList if word not in punctuation]
    textList=["".join(c for c in word if c not in punctuation) for word in textList ]

    #convert digits into NUM
    textList=[re.sub("\d+", "NUM", word) for word in textList]

    #Stem words
    textList=[stemmer.stem(word) for word in textList]

    #Remove blanks
    textList=[word for word in textList if word!= ' ']
    textList=[word for word in textList if word!= '']

    #Extract tokens
    return textList
```

TFIDF Matrix

	freq	idf	logidf	tfidf	logtfidf
thee	18	139.000000	4.934474	2502.000000	88.820531
prophet	12	139.000000	4.934474	1668.000000	59.213687
me	12	139.000000	4.934474	1668.000000	59.213687
we	21	75.818182	4.328338	1592.181818	90.895101
uniti	9	139.000000	4.934474	1251.000000	44.410265
fate	8	139.000000	4.934474	1112.000000	39.475791
muhammad	8	139.000000	4.934474	1112.000000	39.475791
decre	8	139.000000	4.934474	1112.000000	39.475791
christ	10	104.250000	4.646792	1042.500000	46.467919

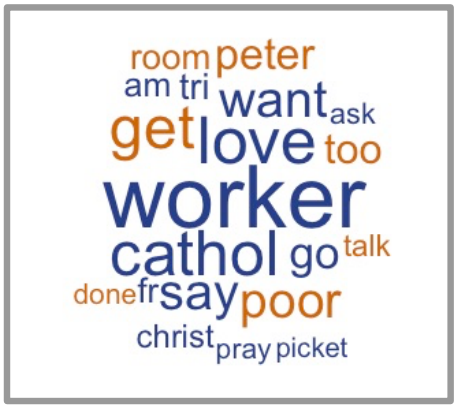
Value Words

Previous Method



TFIDF Method w/Pronouns

TFIDF Method w/out Pronouns



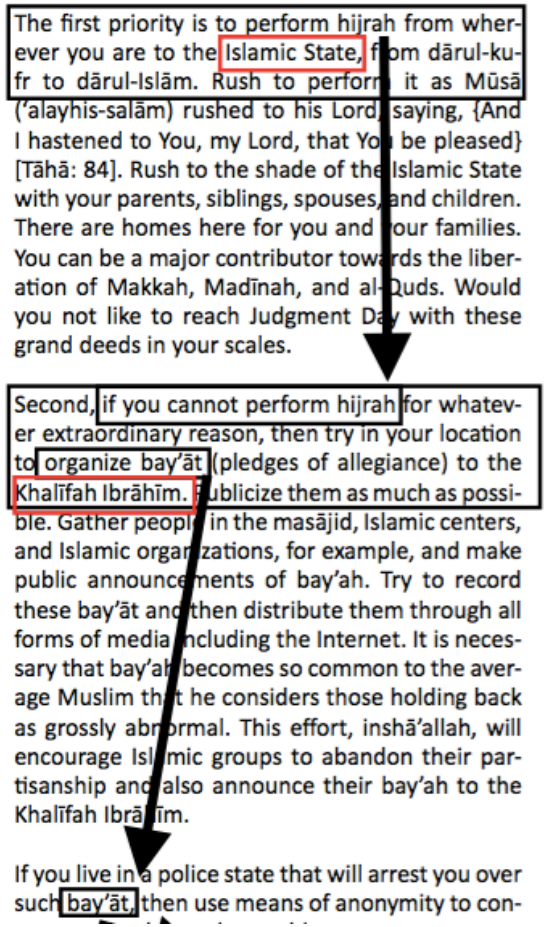
Judgment Classification

Previous Method:

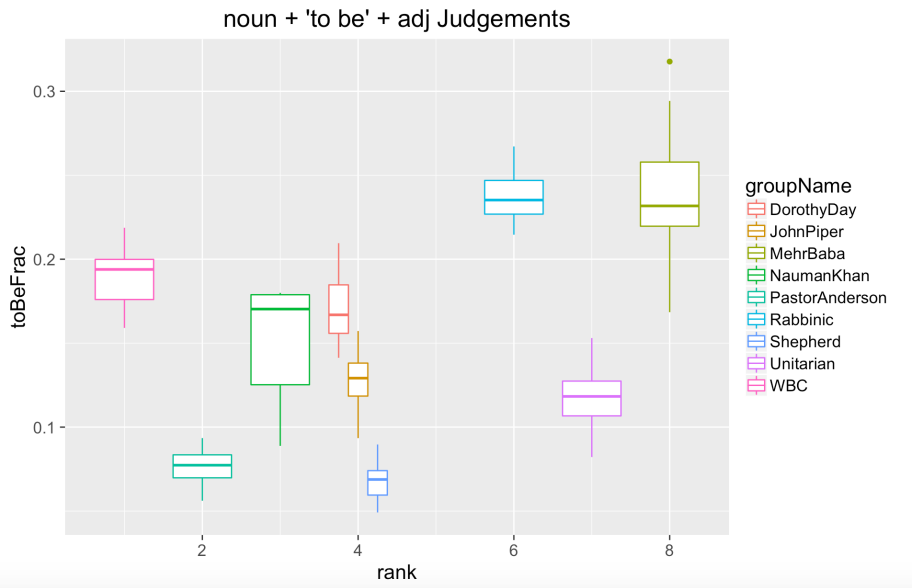
- *Judgment = (noun + "to be" verb + adj/adv)*

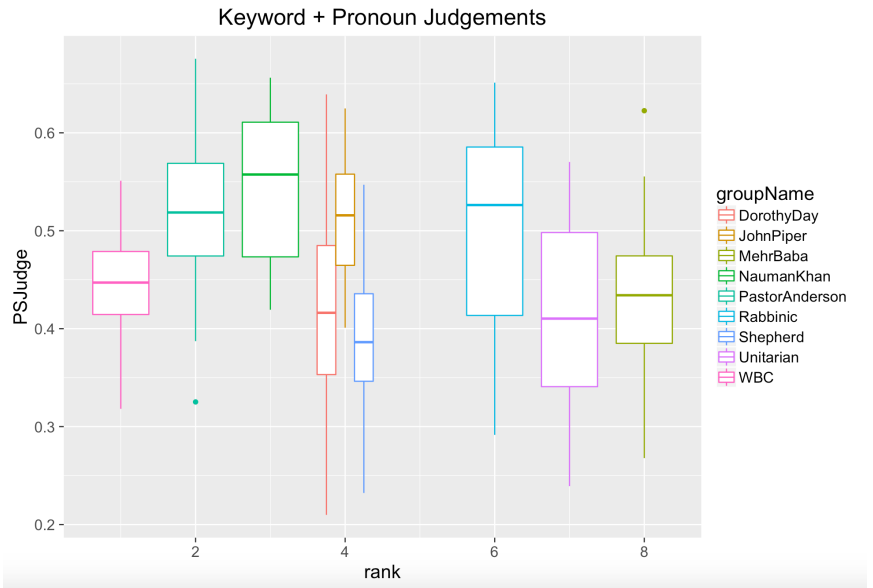
New Method:

- *Judgment = pronoun + key word*

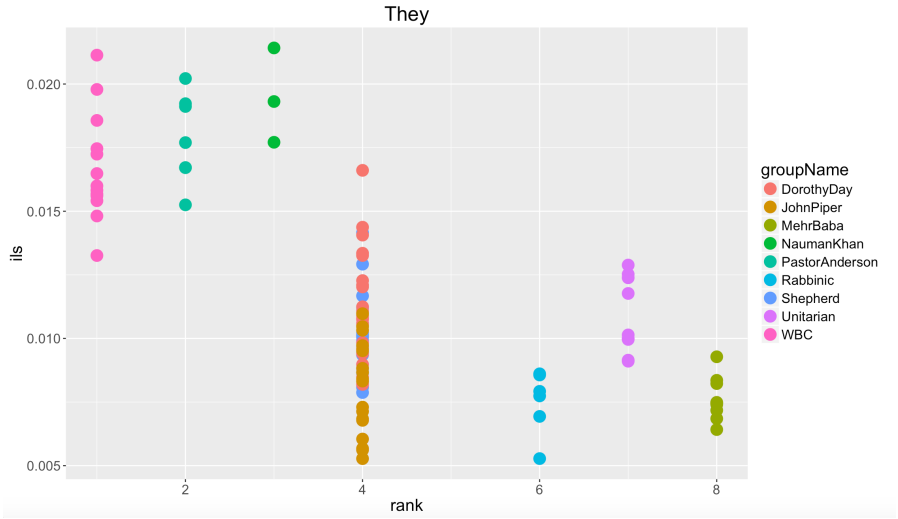


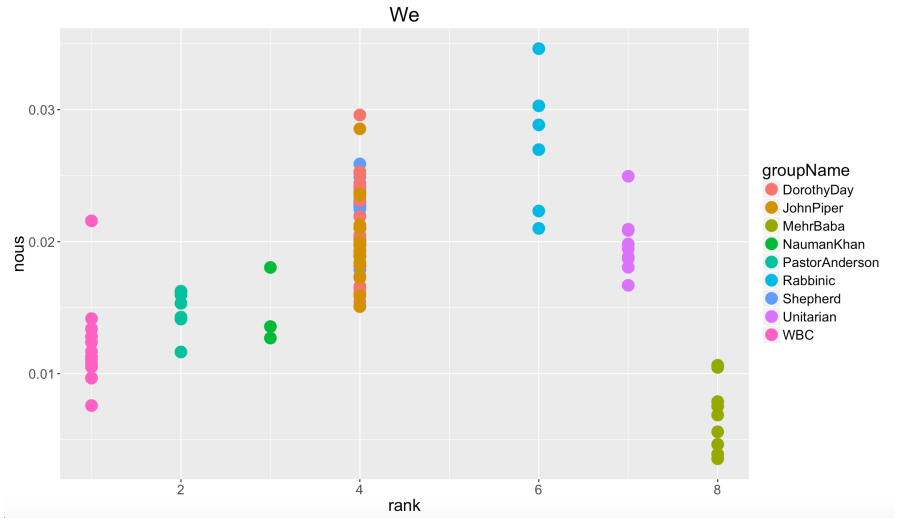
Judgment Classification



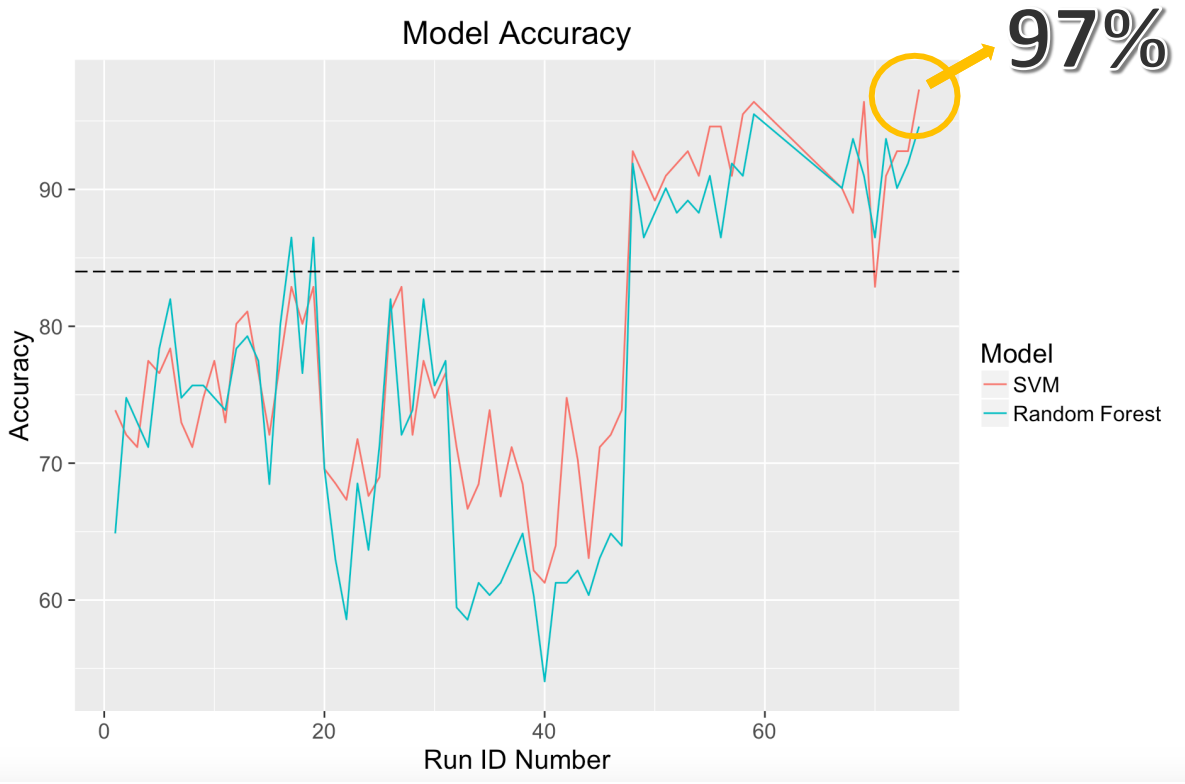
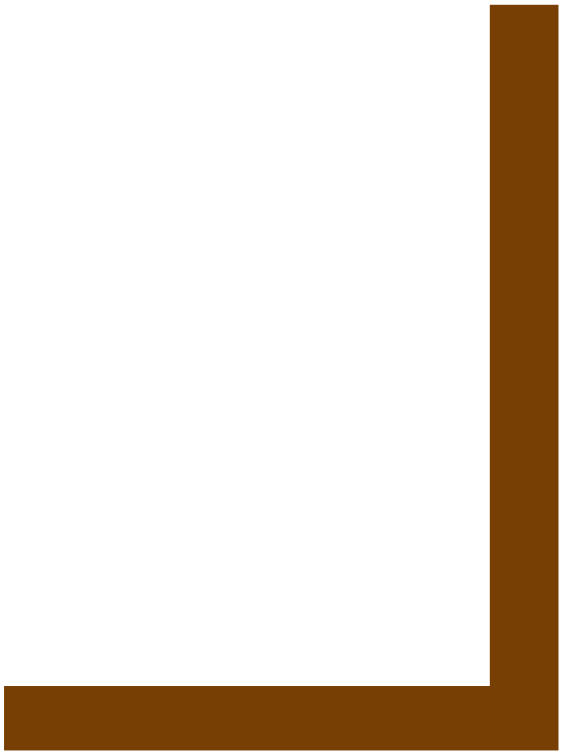


Pronoun Analysis





Preliminary Results



Questions?