

Computational Analysis of Religious and Ideological Linguistic Behavior

Seth Green, Megan Stiles, Leigh Harton, Samantha Garofalo



UNIVERSITY OF VIRGINIA
**DATA SCIENCE
INSTITUTE**

Advisor: Don Brown

Client: UVa Initiative on Religion, Politics, and Conflict (RPC)

Sponsor: Army Research Lab (ARL)

Collaborators

- Don Brown, Data Science Institute
- Peter Ochs, Jonathan Teubner, Essam Fahim, and Syed Moulvi, RPC
- Mohamed Al Boni, UVa Systems Engineering

Motivation

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

Motivation



Technical Approach



Results

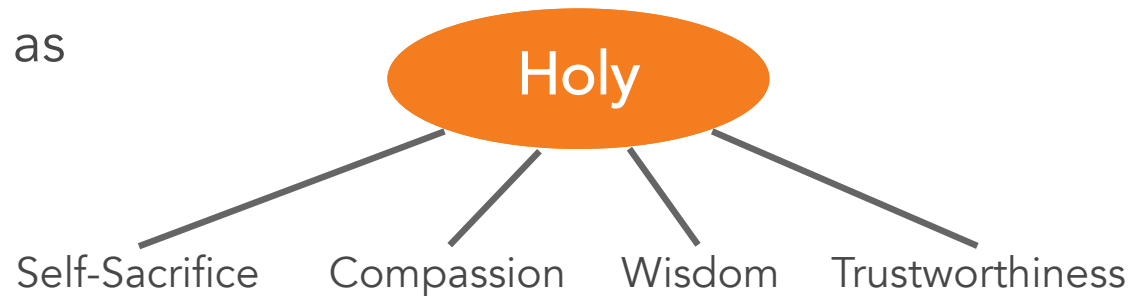


Conclusion

Linguistic Rigidity

1-9 Scale

- Number of definitions not the definition itself is what matters
- Use Manual Method as Ground Truth



Motivation

Technical Approach

Results

Conclusion

Related Work

Venuti et al. (2016)

Text Categorization and Topic Detection

- Gupta et al. (2009)
- Gabrilovich et al. (2013)
 - labeling text documents using outside corpora
- Random Forest
 - Palomino-Garibay et al. (2015)
 - Treeratpituk et al. (2009)
- SVM
 - Joachims (1998)

Motivation

Technical Approach

Results

Conclusion

About Our Data

Input

- Raw ideological text

Output

- Number on Linguistic Rigidity Scale

Sources:

- Venuti et al. (2016) corpus
- New Groups
 - Web Scraping
 - OCR for PDFs

Group	# Documents	Group Rank
ACLU	40	3
American Ethical Union	15	8
Bahai	73	6
Dorothy Day	774	4
Integral Yoga/Yogaville	59	6
ISIS	48	1
John Piper	579	4
Liberal Judaism	166	6
Malcolm X	15	2
Meher Baba	265	8
Pastor Anderson	228	1
Rabbinic	58	4
Sea Shepherds	606	2
Steve Shepherd	728	4
Unitarian	301	7
Westboro Baptist Church	422	1

Motivation

Technical Approach

Results

Conclusion

Technical Approach

- Context Vectors
- Keyword Selection
- Judgement Classification
- Pronoun Analysis

Context Vectors

sinners should be welcome. However, some churches are quick to boot people out if they don't live up to their standards. I heard about a preacher whose wife was in the hospital and dying of cancer. The elders of his church came to the hospital to see them. The preacher thought they were coming to pray for his wife, but instead, they came to fire him. I have no idea what that preacher did that was so bad, but those elders demonstrated no mercy at all. I have heard of other horror stories where people have been driven out of a church for one thing or another. No mercy was demonstrated, only rugged legalism. But Jesus said, "I desire mercy, not sacrifice." Matt. 23:23 "Woe to you, teachers of the law and Pharisees, you hypocrites! You give a tenth of your spices: mint, dill and cummin. But you have neglected the more important matters of the law: justice, mercy and faithfulness. You should have practiced the latter, without neglecting the former." The Pharisees of old were scrupulous or careful

Raw Text

Motivation

Technical Approach

Results

Conclusion

Context Vectors

sinner should be welcome. However, some churches are quick to boot people out if they don't live up to their standards. I heard about a preacher whose wife was in the hospital and dying of cancer. The elders of his church came to the hospital to see them. The preacher thought they were coming to pray for his wife, but instead, they came to fire him. I have no idea what that preacher did that was so bad, but those elders demonstrated no **mercy** at all. I have heard of other horror stories where people have been driven out of a church for one thing or another. No mercy was demonstrated, only rugged legalism. But Jesus said, "I desire mercy, not sacrifice." Matt. 23:23 "Woe to you, teachers of the law and Pharisees, you hypocrites! You give a tenth of your spices: mint, dill and cummin. But you have neglected the more important matters of the law: justice, mercy and faithfulness. You should have practiced the latter, without neglecting the former." The Pharisees of old were scrupulous or careful

Identify Keywords

Motivation

Technical Approach

Results

Conclusion

Context Vectors

sinners should be welcome. However, some churches are quick to boot people out if they don't live up to their standards. I heard about a preacher whose wife was in the hospital and dying of cancer. The elders of his church came to the hospital to see them. The preacher thought they were coming to pray for his wife, but instead, they came to fire him. I have no idea what that preacher did that was so bad, but those elders **demonstrated no mercy at all.** I have heard of other horror stories where people have been driven out of a church for one thing or another. No mercy was demonstrated, only rugged legalism. But Jesus said, "I desire mercy, not sacrifice." Matt. 23:23 "Woe to you, teachers of the law and Pharisees, you hypocrites! You give a tenth of your spices: mint, dill and cummin. But you have neglected the more important matters of the law: justice, mercy and faithfulness. You should have practiced the latter, without neglecting the former." The Pharisees of old were scrupulous or careful

Collect words in context vector

Motivation

Technical Approach

Results

Conclusion

Context Vectors

sinners should be welcome. However, some churches are quick to boot people out if they don't live up to their standards. I heard about a preacher whose wife was in the hospital and dying of cancer. The elders of his church came to the hospital to see them. The preacher thought they were coming to pray for his wife, but instead, they came to fire him. I have no idea what that preacher did that was so bad, but those elders **demonstrated no mercy at all.** I have heard of other horror stories where people have been driven out of a church for one thing or another. No mercy was demonstrated, only rugged legalism. But Jesus said, **"I desire mercy, not sacrifice."** Matt. 23:23 "Woe to you, teachers of the law and Pharisees, you hypocrites! You give a tenth of your spices: mint, dill and cummin. But you have neglected the more important matters of the law: **justice, mercy and faithfulness.** You should have practiced the latter, without neglecting the former." The Pharisees of old were scrupulous or careful

- Repeat for all occurrences of keyword
- Singular Value Decomposition to vector of length 100
- Calculate Average Semantic Density and Distributional Co-occurrence Score

Motivation

Technical Approach

Results

Conclusion

TFIDF Matrix

Term	Freq	IDF	TFIDF
servanthood	4	1,443	5,773
selfexalt	2	1,443	2,887
coquett	1	1,443	1,443
dumbfound	1	690	690
sabbath	11	47	515
manger	1	378	378
unti	1	353	353
childlik	2	169	338
personhood	1	305	305
jesus	31	10	302

Motivation

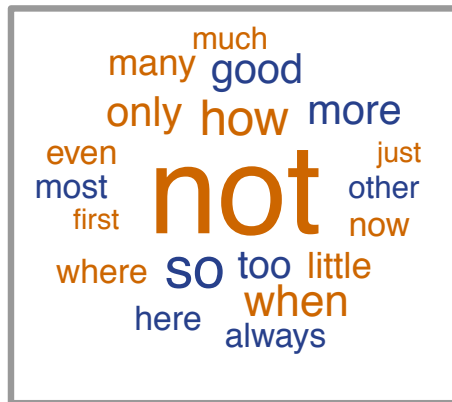
Technical Approach

Results

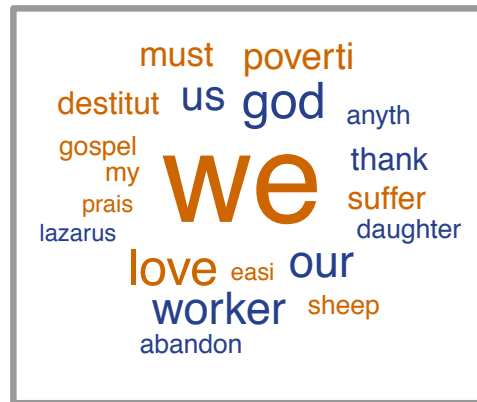
Conclusion

Value Words

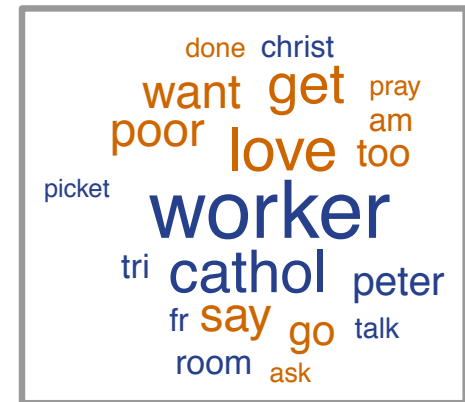
Venuti et al. (2016) Method



TFIDF Method w/Pronouns



TFIDF Method w/out
Pronouns



Judgment Classification

Venuti et al. (2016) Method:

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

New Method:

- Judgment = pronoun + keyword

The first priority is to perform hijrah from wherever you are to the Islamic State, from dārul-kufr to dārul-Islām. Rush to perform it as Mūsā ('alayhis-salām) rushed to his Lord, saying, {And I hastened to You, my Lord, that You be pleased} [Tāhā: 84]. Rush to the shade of the Islamic State with your parents, siblings, spouses, and children. There are homes here for you and your families. You can be a major contributor towards the liberation of Makkah, Madīnah, and al-Quds. Would you not like to reach Judgment Day with these grand deeds in your scales.

Second, if you cannot perform hijrah for whatever extraordinary reason, then try in your location to organize bay'āt (pledges of allegiance) to the Khalīfah Ibrāhīm. Publicize them as much as possible. Gather people in the masājid, Islamic centers, and Islamic organizations, for example, and make public announcements of bay'ah. Try to record these bay'āt and then distribute them through all forms of media including the Internet. It is necessary that bay'ah becomes so common to the average Muslim that he considers those holding back as grossly abnormal. This effort, inshā'allah, will encourage Islamic groups to abandon their partisanship and also announce their bay'ah to the Khalīfah Ibrāhīm.

If you live in a police state that will arrest you over such bay'āt, then use means of anonymity to con-

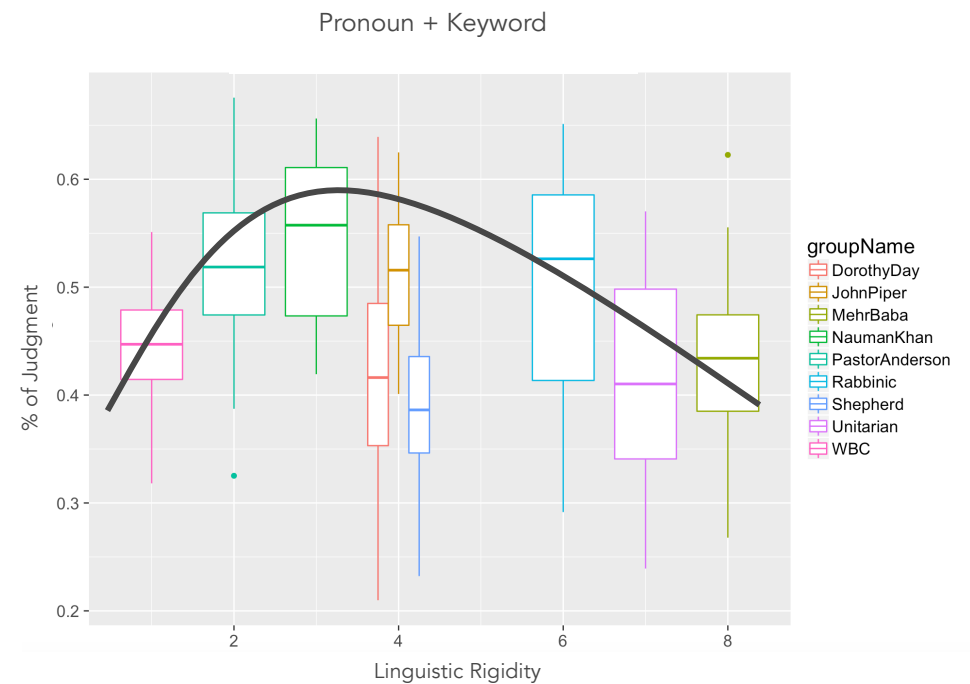
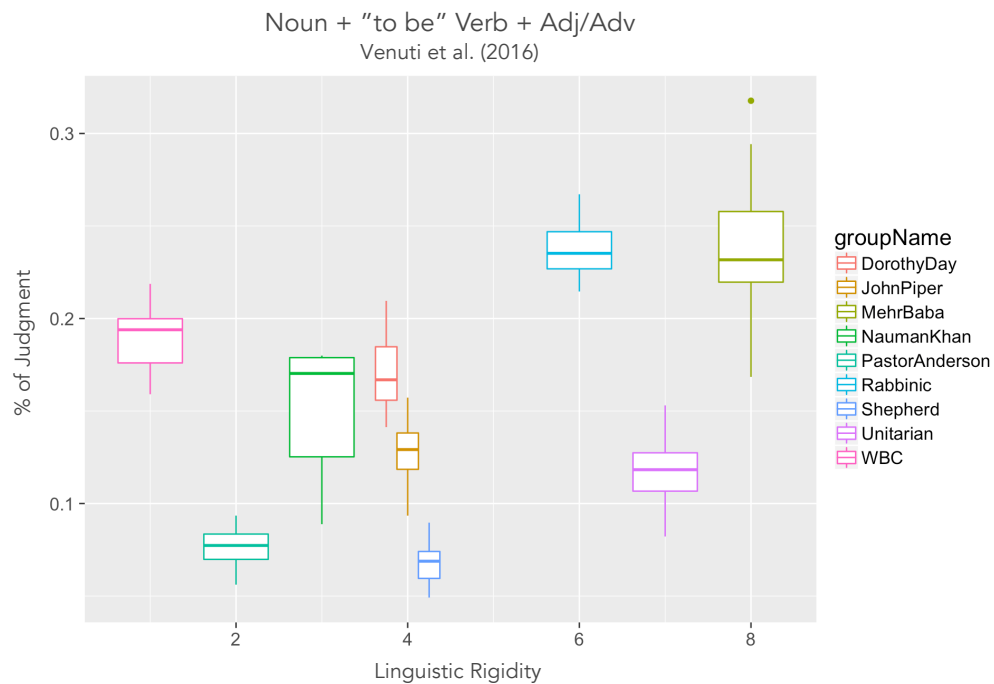
Motivation

Technical Approach

Results

Conclusion

Judgment Classification



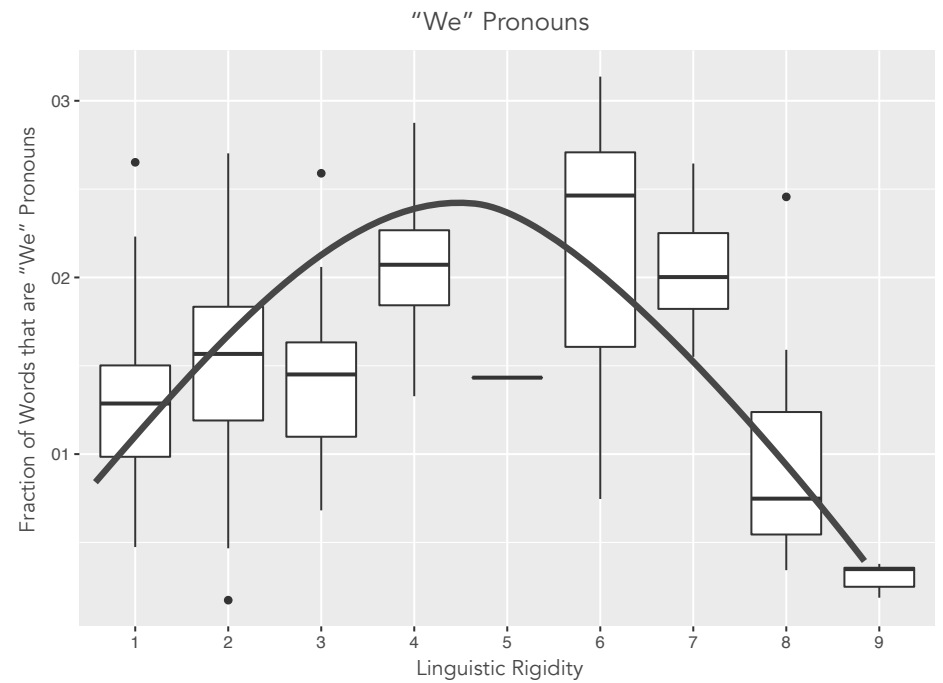
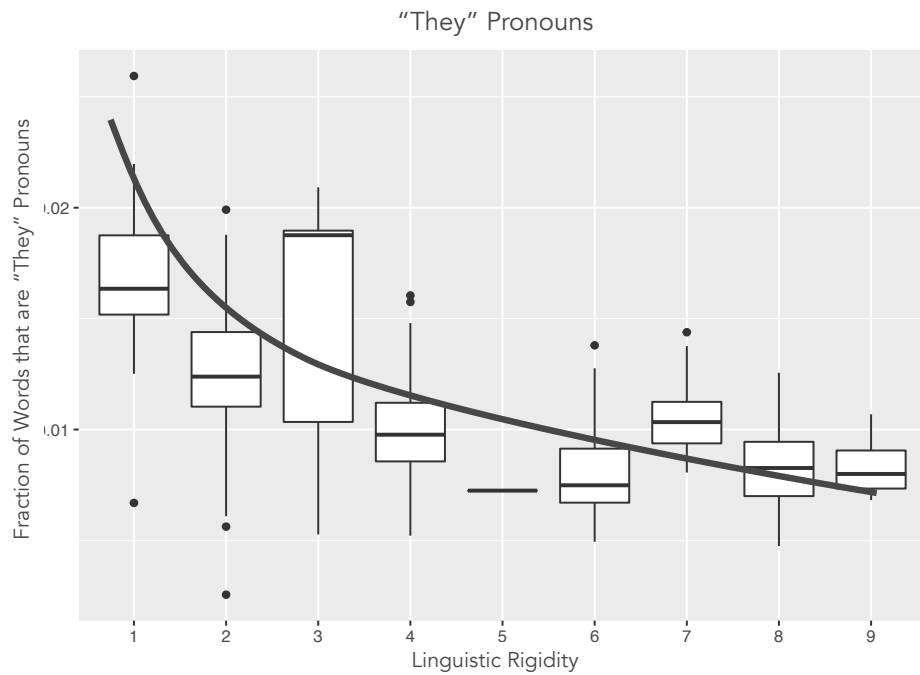
Motivation

Technical Approach

Results

Conclusion

Pronoun Analysis



Motivation

Technical Approach

Results

Conclusion

Results

- Binned Document Analysis
- Single Document Analysis

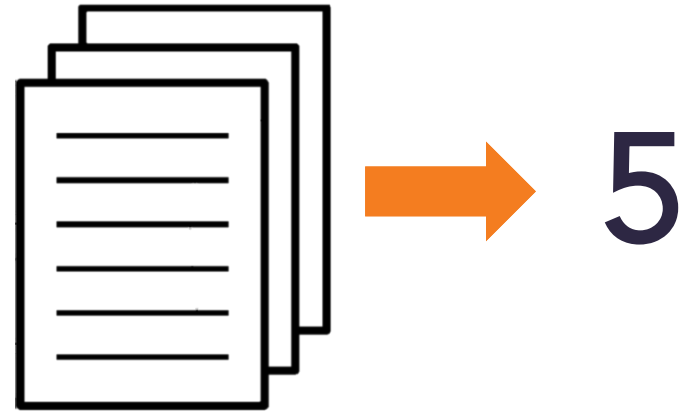
Binned Document Analysis

Why Binned Documents?

- Understand a group's baseline score
- Comparison to previous work

of Documents

- 4,568



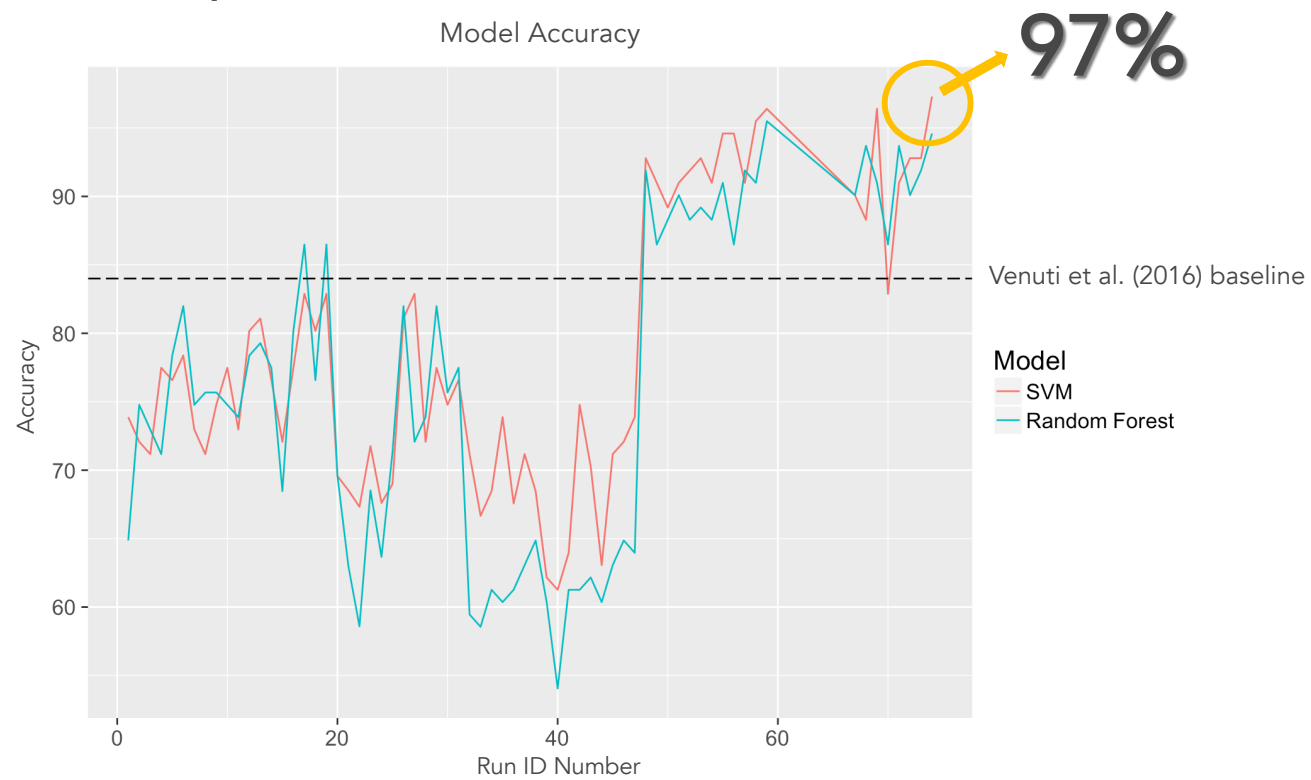
Motivation

Technical Approach

Results

Conclusion

Results – Group Scores



Motivation

Technical Approach

Results

Conclusion

Results – Group Scores

Compared to Venuti et al. (2016)

Model	Accuracy	Venuti et al. (2016)
Random Forest	96.5	86.0
SVM	97.2	84.0
Boosted Tree	96.2	N/A
Artificial Neural Net	N/A	80.0
Average	96.6	83.3

Using Extended Data Set

Model	Accuracy	Mean Absolute Error
Random Forest	96.1	0.239
SVM	96.3	0.222
Boosted Tree	95.4	0.246
Average	96.0	0.236

Motivation

Technical Approach

Results

Conclusion

Single Document Analysis

Why Single Documents?

- Identifying potential spikes or deviations from a groups baseline
- Moving towards real-time, on-the-ground analysis

of Documents

- 342



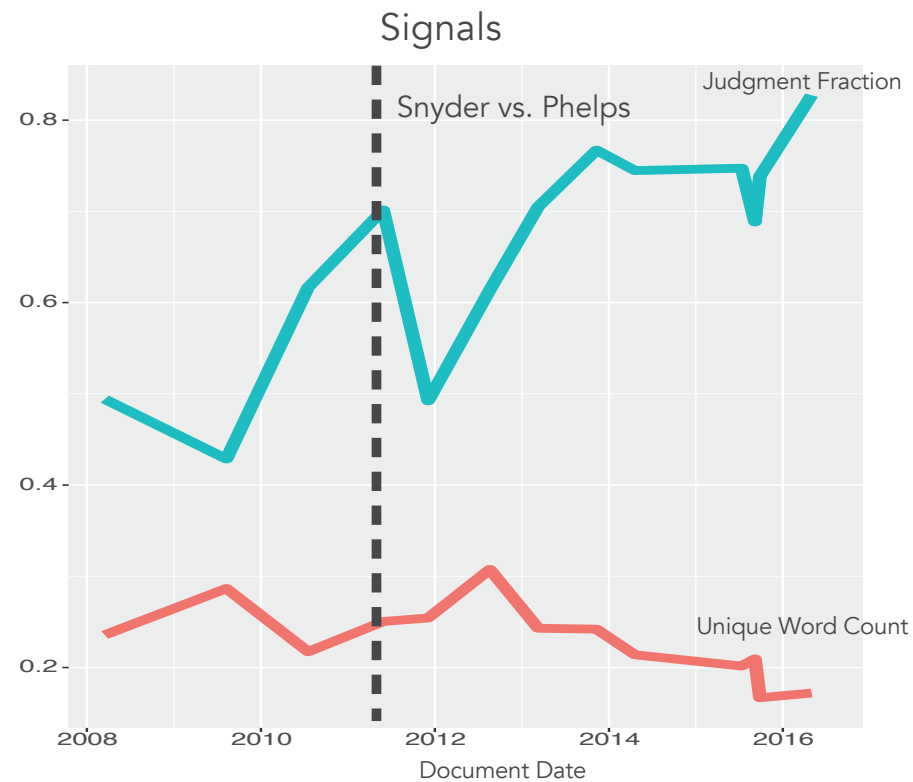
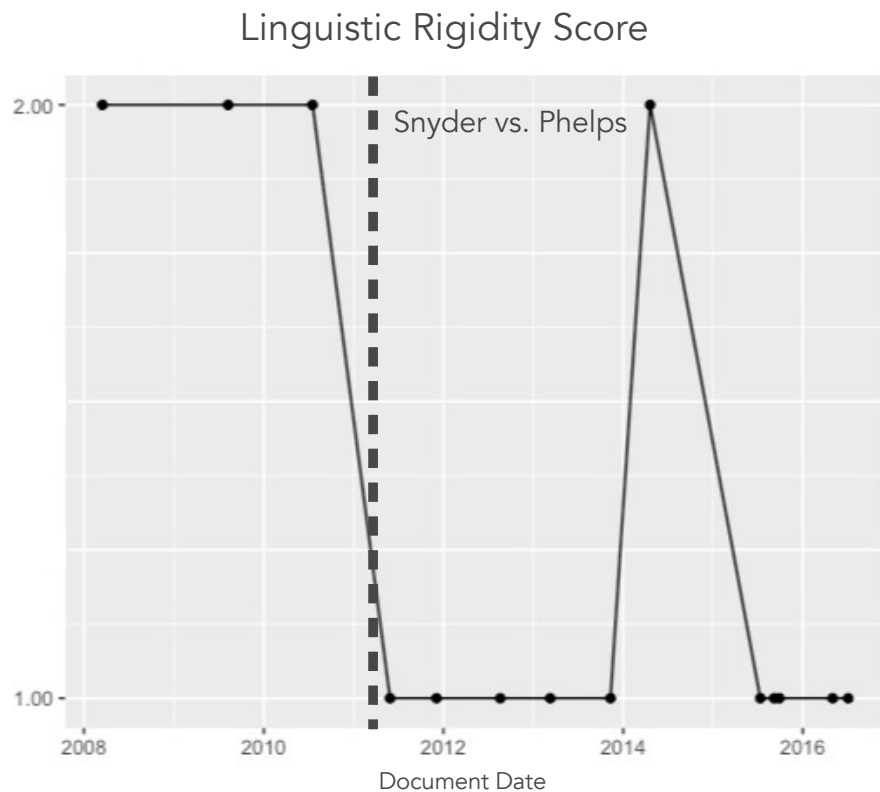
Motivation

Technical Approach

Results

Conclusion

Westboro Baptist Church



Motivation

Technical Approach

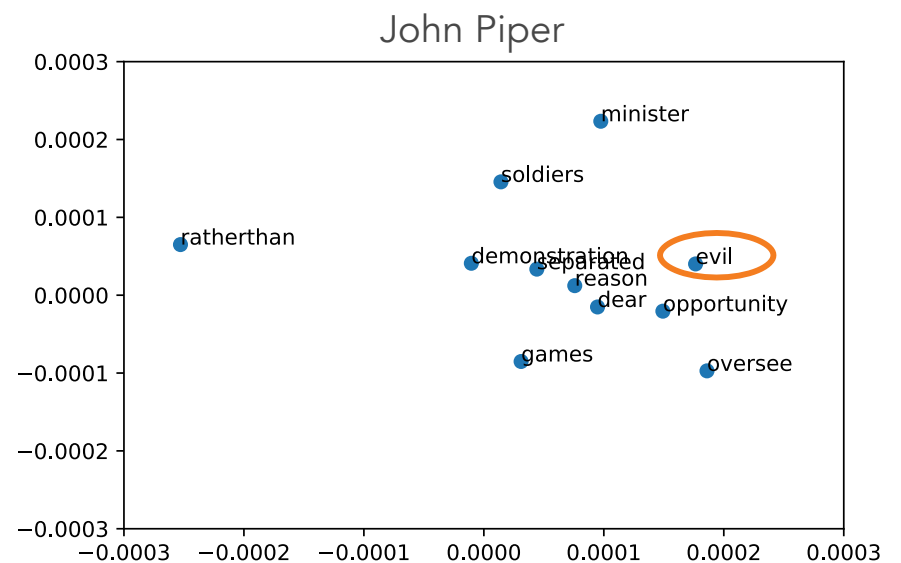
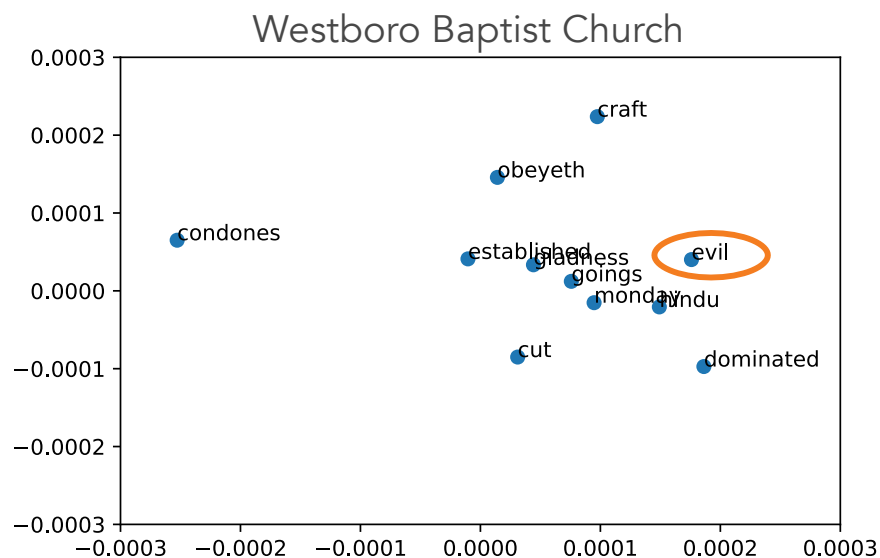
Results

Conclusion

Word Embeddings

What are Word Embeddings?

- Vocabulary mapped to vectors



Motivation

Technical Approach

Results

Conclusion

Results – Single Document Scores

Model	Accuracy	Mean Absolute Error
Random Forest	77.5	0.805
SVM	70.5	1.095
Boosted Tree	69.7	1.084
CNN with Embeddings	65.6	1.125

Motivation

Technical Approach

Results

Conclusion

Conclusion

Conclusion

In a global society, we want to work with Value-Based groups

- How a group uses language provides guidance for how they will interact

Computationally Efficient

- 250 hrs vs 1 hr

Real-time recommendations to organizations on the ground

- Understanding both Religious and Ideological groups

Motivation



Technical Approach



Results



Conclusion

Questions?

References

1. Consultation with researchers at the University of Virginia Center for Religion, Politics, and Conflict (RPC). 1540 Jefferson Park Avenue, University of Virginia, Charlottesville, VA 22904-4126.
2. Venuti, Nicholas, Sachtjen, Brian, McIntyre, Hope, et al. "Predicting the Tolerance Level of Religious Discourse Through Computational Linguistics," presented at the 2016 *IEEE Systems and Information Engineering Design Conference (SIEDS '16)*, Charlottesville, VA, 2016.
3. Wade-Benzoni, Kimberly A., Hoffman, Andrew J., Thompson, Leigh L., et al. 2002. "Barriers to resolution in ideologically based negotiations: The role of values and institutions." *Academy of Management Review* 27, pp. 41-57.
4. Lederach, John Paul and Appleby, Scott R. "Strategic Peacebuilding: An Overview," in Daniel Philpott and Gerard Powers, *Strategies of Peace* (Oxford: Oxford University Press, 2010), pp. 19-44.
5. "Algorithmia." 2016. <http://blog.algorithmia.com/introduction-natural-language-processing-nlp/>. Accessed: March 30, 2017.
6. "Expert System: Natural language processing and text mining." 2016. <http://www.expertsystem.com/natural-language-processing-and-text-mining/>. Accessed: March 30, 2017
7. Gupta, Vishal. and Lehal, Gurpreet S., 2009. "A survey of text mining techniques and applications." *Journal of emerging technologies in web intelligence*, 1(1), pp.60-76.
8. Gabrilovich, Evgeniy and Markovitch, Shaul. January 2007. "Computing semantic relatedness using wikipedia-based explicit semantic analysis." (In *IJCAI* Vol. 7), pp. 1606-1611
9. James, Gareth, Witten, Daniela, Hastie, Trevor, et al. 2013. *Introduction to Statistical Learning*, pp 303-372.
10. Palomino-Garibay, Alonso, Camacho-González, Adolfo T., Fierro-Villaneda, Ricardo A., et al. 2015. "A random forest approach for authorship profiling." *Cappellato et al. [8]*.
11. Treeratpituk, Pucktada and Giles, C. Lee. June 2009. "Disambiguating authors in academic publications using random forests." In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, pp. 39-48.
12. Joachims, Thorsten, 1998. "Text categorization with support vector machines: Learning with many relevant features." *Machine learning: ECML-98*, pp.137-142.
13. Özgür, Arzucan, Levent Özgür, and Tunga Güngör. "Text categorization with class-based and corpus-based keyword selection." *International Symposium on Computer and Information Sciences*. Springer Berlin Heidelberg, 2005.
14. Richardson, Leonard. *Beautiful soup documentation*, [Online] Available: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>.
15. Bird, Steven, Edward Loper and Ewan Klein (2009), *Natural Language Processing with Python*. O'Reilly Media Inc. Available: <http://nltk.org/book>
16. Pennington, Jeffrey, Richard Socher, Christopher D. Manning. "GloVe: Global Vectors for Word Representation." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. 2014.
17. Csardi, G. and Nepusz, T. *Package 'igraph'*. [Online]. Available: <http://igraph.org/python/>
18. Íñigo-Mora, Isabel. "On the use of the personal pronoun we in communities." *Journal of Language and Politics*. Volume 3, Issue 1, 2004, pages: 27–52.
19. Maitland, Karen and John Wilson. "Pronominal selection and ideological conflict." *Journal of Pragmatics*. Volume 11, Issue 4, August 1987, Pages 495-512.
20. Karatzoglou, Alexandros, Alex Smola, and Kurt Hornik, *Package 'kernlab'*. [Online] Available: <https://cran.r-project.org/web/packages/kernlab/kernlab.pdf>
21. Liaw, Andy and Matthew Wiener, *R Package 'randomForest'*. [Online] Available: <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
22. Chen, Tianqi, Tong He, Michael Benesty, Vadim Khotilovich, and Yuan Tang, *Package 'xgboost'* [Online] Available: <https://CRAN.R-project.org/package=xgboost>