

[W207] Applied Machine Learning Week 1

Professor Jason Anastasopoulos
ljanastas@princeton.edu

Princeton University
University of Georgia

January 8, 2018

About Me



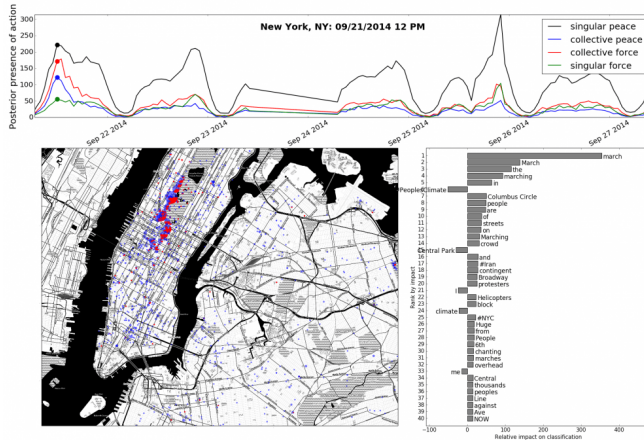
My full name has 27 letters.

About Me



I have a crazy 19 month old named Seth.

About Me

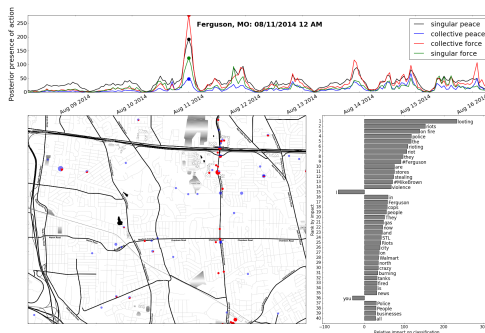


I do research on public policy, political science and machine learning.

Recent Machine Learning Projects

- 1 **Event detection** – Identifying violent and non-violent protest activity using streaming Twitter data (Naive Bayes).
- 2 Measuring violence in religious texts (support vector machines).
- 3 Fairness and bias in machine learning

[W207] Applied Machine Learning Week 1



Violent and non-violent protest activity in the wake of the Ferguson verdict.

Violent and non-violent protest activity during Hong Kong Occupy protests.



Measuring violence in religious texts

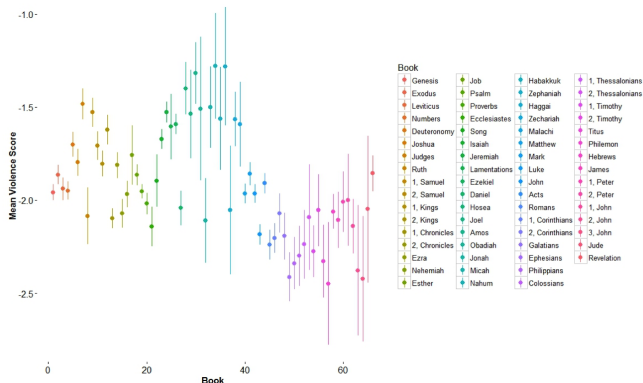
Bible	
Collective	"He arose and struck the Philistines until his hand was weary and his hand froze to the sword..." (2 Samuel 23:10)
Collective Promote	"I will set a fire in Egypt Sin shall be in great anguish..." (Ezekiel 30:16)
Interpersonal	"Even in the third year of Asa king of Judah did Baasha kill him and reigned in his place" (1 Kings 15:28)
Interpersonal Promote	"The daughter of any priest if she profanes herself by playing the prostitute she profanes her father she shall be burned with fire." (Leviticus 21:9)

Table 1: Samples of classified verses from training data.

Identification of different types of violence in verses with support vector machines.

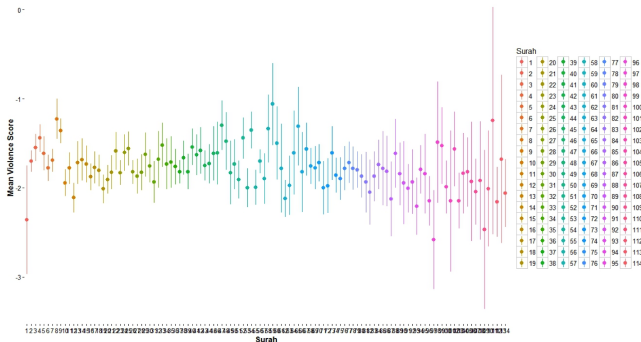
Trained on classified Quran and Bible verses.

Measuring violence in religious texts



SVM confidence scores to estimate patterns of violence in the Bible...

Measuring violence in religious texts



...and in the Quran.

About POLS 4150

- Course goals.
- Textbooks.
- R.
- Syllabus.
- Intro to Machine Learning

By the end of this course you will...

- Understand how to apply some of the most popular machine learning algorithms.
- Be able to read research papers discussing these machine learning algorithms.

Textbooks



A Course in Machine Learning

by Hal Daumé III

Machine learning is the study of algorithms that learn from data and experience. It is applied in a vast variety of application areas, from medicine to advertising, from military to pedestrian. Any area in which you need to make sense of data is a potential consumer of machine learning.

CIML is a set of introductory materials that covers most major aspects of modern machine learning (supervised learning, unsupervised learning, large margin methods, probabilistic modeling, learning theory, etc.). It's focus is on broad applications with a rigorous backbone. A subset can be used for an undergraduate course, a graduate course could probably cover the entire material and then some.

You may obtain the written materials by purchasing a (\$50) print copy, by downloading the entire book, or by downloading individual chapters below. If you find the electronic version of the book useful and would like to donate a small amount to support further development, that's always appreciated! You can get the source code for the book, labs and other teaching materials on GitHub. The current version is 0.99 (the "beta" pre-release). You can view v0.9 if you prefer.

Support and Mailing Lists:

If you would like to be informed when new versions of CIML materials are released, please join the CIML mailing list. If you find errors in the book, please fill out a bug report. If you're the first to submit an error, you'll get listed in the acknowledgments!

Hal Daume's: A Course in Machine Learning

Github



All materials will be on Github.

All assignments will be submitted on Github.

Course repository:

[https://github.com/ljanastas/
207-Applied-Machine-Learning](https://github.com/ljanastas/207-Applied-Machine-Learning)

Python and Jupyter



- I use Python 2.7+.
- Assignments will be completed in Jupyter (Python) notebooks.

Assignments



- 4 total assignments.
- 3 HW's due on Week 5, Week 9, Week 13 (Projects folder).
- Final project: Week 10 (Proposal); Week 12 (Discuss Project); Week 15 (Submit Project Notebook + Presentation).