Text as Data: Poli 274

Course Information

Course Description	In this course we will introduce an approach for including text data in social science research design. We introduce how we can use text data to describe the prevalence of a social behavior or phenomenon and make inferences about its origins. We explain how the abundance of text and new statistical methods facilitate these inferences. The goal of inference in social science research is qualitatively different than the goals that have been often used to evaluate text analytic methods, which often focus on performing a predictive task. The focus on inference will push us to reconsider when and how some methods are useful, suggest new ways to evaluate methods, and will present new open questions in the practical and ethical use of text as data. The goal of the course is to provide students with an overview of the techniques for text analysis, examples of how it has been used in the social sciences, and a framework for understanding how it can be incorporated into social science research design. While the time scale does not permit a deep mathematical understanding of every approach, students will learn tools for analyzing texts quantitatively and a framework of how these methods can be used for social science inference. The course is designed to help students use text data in their own research. Assignments are tailored to students' individual research projects.
Instructor	Molly Roberts
IA/TA	Eddie Yang

Course Learning Outcomes

Upon completion of this course, students will be able to:
1. Define and identify the social science tasks discovery, measurement, causal inference, and prediction in applied social science research.
2. Explain (but not necessarily derive) the foundations of statistical text analysis methods, including bag of words approaches, word embeddings, clustering, topic modeling, and classification.
3. Implement the above statistical text analysis methods in statistical software.

4. Produce a final project that applies statistical text analysis to a social science project. Describe not only what is learned from the analysis, but also the limitations of the analysis.

Course Format

This course will be held completely virtually. Because three hours in a row of virtual lecture is not conducive to learning, students will watch ~4 short videos each week and participate in one hour of synchronous discussion.

Synchronous* Online Lectures:

Thursday, 10AM-11AM
Zoom links available on Canvas under the Zoom LTI PRO application.
*Those not able to attend the synchronous lectures should post a question and reply to a question on that week's discussion board to receive participation credit.

Asynchronous (Online) Course Elements:

Videos, readings, problem sets, and other resources available at: UC San Diego's Learning Management System: https://canvas.ucsd.edu. Readings are available through Perusall
Login: UC San Diego Active Directory credentials

A Typical Week in This Course

Monday-Thursday:

- Watch ~4 short recorded lecture segments and complete readings. These lectures will introduce the foundational concepts in the course that will help you complete the assignments and final project.
- Complete assigned readings and make one annotation per reading/chapter on Perusall (linked in Canvas)
- Optional Office Hours 9-9:50AM and 11-11:50AM Thursdays, Please sign up for office hours here: https://calendly.com/molly-e-roberts/poli274-officehours

Thursday:

- Assignments due before lecture, at 10AM.
- Synchronous discussion section, 10-11AM. In this discussion section, students will apply concepts from the lecture videos in hands-on code examples. Students will code with the professor.
- For those who cannot attend the discussion section synchronously, please post at least one question and a reply to a question on that week's discussion board.

Overall Course Expectations

What you can do to support your success in the course:	What I will do to support your success in the course:
Read the syllabus and stay current with course information	Be prepared and bring my enthusiasm for teaching to each session
Keep up with readings and assignments, as each one builds on the previous one.	Respond to emails within one working day, and provide timely feedback on assignments / submissions.
Contribute to the learning environment with fairness, cooperation, and professionalism	Establish a learning environment with fairness, cooperation and professionalism, and will take action if these principles are violated.
Treat your classmates, instructional assistants and myself honestly and ethically	Treat you honestly and ethically, and will address any concerns you might have
Commit to excel with integrity[1]. Have the courage to act in ways that are honest, fair, responsible, respectful & trustworthy.	Uphold integrity standards and create an atmosphere that fosters active learning, creativity, critical thinking, and honest collaboration.
Manage your time, so you can stay on track with the course and complete tasks on time	Only assign work that is vital to the course, and will work to meet the standard credit hour allotment for the course.
Communicate with me if you determine that a deadline cannot be met due to extenuating circumstances	Consider requests for adjustments and will make reasonable exceptions available to all students when approved

1. Please read UC San Diego's Policy on Integrity of Scholarship and take the integrity pledge!

Readings

Readings will be made available on the course website on Canvas through the Perusall application. Students are expected to make at least one comment on each reading/chapter as part of their participation grade.

Course Software

We will use R (http://www.r-project.org) to introduce the computational tools within the course. Students more familiar with Python are welcome to use Python for assignments and projects if they wish.

Assignments, Projects, and Grading

Summary of Grade Criteria

Assignment	Weight	Due Date
Participation	20%	
Research Memos	45%	Thursdays, 10AM
Research Memo Peer Feedback	15%	Thursdays, 10AM
Final Project Presentation x 1	20%	Due Thursday, May 27, 11:59pm

Grading Scale

A = 93-100%
A- = 90-92.9%
B+ = 88-89.9%
B = 83-87.9%
B- =80-82.9%
C+ = 78-79.9%
C = 73-77.9%
C- =70-72.9%
D = 60-69%
F = 59%-below

Grading Procedure and Feedback

Students will be graded on an absolute scale over the course of the quarter. At the end of the quarter, the average of the top 5 scores in the class will become 100%, which will either keep all grades the same or move them up.

Attendance and Participation

Students will receive attendance and participation points by commenting on readings through Perusall, attending and participating in the weekly discussion section, and commenting on the discussion board.

Instructional Team: Who Are My Instructors?

Instructor

Professor Roberts

Margaret Roberts
Associate Professor
Political Science and HDSI
Margaretroberts.net
Virtual Office Hours:
Thursday 9-9:50AM, 11-11:50AM
Please sign up for office hours here:
https://calendly.com/molly-e-roberts/poli274-officehours

Teaching Assistants

Eddie Yang
Virtual Office Hours:
Thursdays 4-6PM
[redacted]

Resources for Support and Learning

Learning and Academic Support	
Ask a Librarian: Library Support *Chat or make an appointment with a librarian to focus on your research needs* **Course Reserves, Connecting from Off-Campus and Research Support** *Find supplemental course materials* **First Gen Student Success Coaching Program** *Peer mentor program that provides students with information, resources, and support in meeting their goals* **Office of Academic Support & Instructional Services (OASIS)** *Intellectual and personal development support*	**Writing Hub Services in the Teaching + Learning Commons** *One-on-one online writing tutoring and workshops on key writing topics* **Supplemental Instruction** *Peer-assisted study sessions through the Academic Achievement Hub to improve success in historically challenging courses* **Tutoring – Content** *Drop-in and online tutoring through the Academic Achievement Hub* **Tutoring – Learning Strategies** *Address learning challenges with a metacognitive approach*
Support for Well-being and Inclusion	
Basic Needs at UCSD *Any student who has difficulty accessing sufficient food to eat every day, or who lacks a safe and stable place to live is encouraged to contact:* foodpantry@.ucsd.edu | basicneeds@ucsd.edu | (858) 246-2632 **Counseling and Psychological Services** *Confidential counseling and consultations for psychiatric service and mental health programming*	**Community and Resource Centers** **Office of Equity, Diversity, and Inclusion** *As part of the Office of Equity, Diversity, and Inclusion the campus community centers provide programs and resources for students and contribute toward the evolution of a socially just campus* (858).822-.3542 | diversity@ucsd.edu **Get Involved** *Student organizations, clubs, service opportunities, and many other ways to connect with others on campus* **Undocumented Student Services**

Triton Concern Line *Report students of concern:* (858) 246-1111 **Office for Students with Disabilities (OSD)** *Supports students with disabilities and accessibility across campus*	*Programs and services are designed to help students overcome obstacles that arise from their immigration status and support them through personal and academic excellence*

Campus and Course Policies

Course Policies

Health and Well-Being Statement

Throughout your time at UC San Diego, you may experience a range of issues that can negatively impact your learning. These may include physical illness, housing or food insecurity, strained relationships, loss of motivation, depression, anxiety, high levels of stress, alcohol and drug problems, feeling down, interpersonal or sexual violence, or grief.

These concerns or stressful events may lead to diminished academic performance and affect your ability to participate in day-to-day activities. If there are issues related to coursework that are a source of particular stress or challenge, please speak with me, Professor Roberts, so that I am able to support you. UC San Diego provides a number of resources to all enrolled students, including:

- **Counseling and Psychological Services** (858-534-3755 | caps.ucsd.edu)
- **Student Health Services** (858-534-3300 | studenthealth.ucsd.edu)
- **CARE at the Sexual Assault Resource Center** (858-534-5793 | care.ucsd.edu)
- **The Hub Basic Needs Center** (858-246-2632 | basicneeds.ucsd.edu)

We care about you at UC San Diego, and there is always help available.

Subject to Change Policy

While I will try to adhere to the course schedule as much as possible, I also want to adapt to your learning pace and style. Therefore, the syllabus and course plan may change in the quarter. I always welcome feedback from you about what is working and not working for your learning in the course.

Campus Policies

Below are links to useful campus policies – principles of community, integrity, and code of conduct that are also policies within this course:

- **UC San Diego Principles of Community**
- **UC San Diego Policy on Integrity of Scholarship**
- **Religious Accommodation**
- **Nondiscrimination and Harassment**
- **UC San Diego Student Conduct Code**

Course Schedule

Below is a tentative course schedule. Please refer to the Modules section on canvas for the most up to date schedule.

Week	Activities, Assessments, and Due dates
1	Read: ● DiMaggio, Paul. "Adapting computational text analysis to social science (and vice versa)." Big Data & Society 2.2 (2015). ● Catalinac, Amy. "From Pork to Policy: The Rise of Programmatic Campaigning in Japanese Elections." The Journal of Politics. 2016; 78 (1) :1-18. Watch: ● Syllabus & class overview ● **Video 1:** Social Science Research Tasks ● **Video 2:** Example: Catalinac (2015) Do: ● Post to the Discussion Board to Introduce Yourself ● Reading Annotations on Perusall
2	Read: ● Grimmer, Roberts and Stewart, Chapters 3-7 Watch: ● **Video 1:** Selecting documents ● **Video 2:** Bag of words model ● **Video 3:** Multinomial Language Model ● **Video 4:** Vector Space Model & Similarity metrics Do: ● Reading Annotations on Perusall
3	Read: ● Grimmer, Roberts and Stewart, Chapters 8 and 11

	• Nelson, Laura K. "Computational grounded theory: A methodological framework." *Sociological Methods & Research* 49.1 (2020): 3-42. • Watch: • **Video 1:** Distributed Representations of Words • **Video 2:** Example: Bias in Word Embeddings • **Video 3:** Introduction to Discovery • **Video 4:** Discriminating Words Do: • **Research Question/Representation Memo Due** • Reading Annotations on Perusall
4	Read: • Grimmer, Roberts and Stewart, Chapters 8 and 11 • Nelson, Laura K. "Computational grounded theory: A methodological framework." *Sociological Methods & Research* 49.1 (2020): 3-42. Watch: • **Video 1:** K-means Clustering • **Video 2:** Interpreting Clusters • **Video 3:** Introduction to Topic Models • **Video 4:** Interpreting Topic Models Do: • **Research Question/Representation Memo Feedback Due** • Reading Annotations on Perusall
5	Read: • Grimmer, Roberts and Stewart, Chapters 15-17 • Gillion (2016), Chapter 2. Watch: • **Video 1:** Introduction to Measurement • **Video 2:** Word Counting • **Video 3:** Introduction to Supervised Classification • **Video 4:** Example: Gillion (2016) Do: • **Discovery Memo due** • Reading Annotations on Perusall
6	Read: • Grimmer, Roberts and Stewart, Chapters 19-20 Watch: • **Video 1:** Supervised Classification, Part I • **Video 2:** Supervised Classification, Part II • **Video 3:** Validation of Measurement Part I

	• **Video 4:** Validation of Measurement Part II Do: • **Discovery Memo feedback due** • Reading Annotations on Perusall
7	Read: • Grimmer, Roberts and Stewart, Chapters 22-23 Watch: • **Video 1:** Prediction versus Causal Inference • **Video 2:** Overview of Prediction • **Video 3:** Validation of Prediction • **Video 4:** Ethical Considerations in Prediction Do: • **Measurement Memo due** • Reading Annotations on Perusall
8	Read: • Grimmer, Roberts and Stewart, Chapters 24-25 Watch: • **Video 1:** Overview of Causal Inference Part 1 • **Video 2:** Overview of Causal Inference Part 2 • **Video 3:** Text as Outcome Overview • **Video 4:** Text as Outcome Example Do: • **Measurement Memo feedback due** • Reading Annotations on Perusall
9	Read: • Grimmer, Roberts and Stewart, Chapters 26-27 Watch: • **Video 1:** Text as Treatment Overview • **Video 2:** Text as Treatment Example • **Video 3:** Text as Confounder Overview • **Video 4:** Text as Confounder Example Do: • **Final Project Presentation due (5 minute video)** • Reading Annotations on Perusall
10	Watch: Video of presentations produced by students.