CSCI 599 AI in Society: Bias and Fairness in Data, Networks, and Algorithms

Units: 4

Term—Day—Time:

Fall 2019—MW— 10am-11:50am

IMPORTANT:
The general formula for contact hours is as follows:
Courses must meet for a minimum of one 50 minute session per unit per week over a semester.

Location: TBD – blackboard.usc.edu

Instructor: Kristina Lerman
Office: ISI 932
Office Hours: immediately before class
Contact Info: lerman@isi.edu, 310-448-8714

Instructor: Fred Morstatter
Office: ISI 938
Office Hours: immediately before class
Contact Info: fredmors@isi.edu, 310-448-9381

Teaching Assistant: TBD
Office: Physical or virtual address
Office Hours:
Contact Info: Email, phone number (office, cell), Skype, etc.

Catalog Description

Analytic methods for mining social data to understand how biases affect analysis of social data and their impact on fairness; Focus on hands-on experience with quantitative methods, including statistical analysis, machine learning, network analysis and linear algebra. **Recommended preparation**: Knowledge of at least one programming language (Java, C++, Python); undergraduate level training or coursework in linear algebra, basic probability and statistics; an undergraduate level course in Artificial Intelligence are helpful but is not required.

Course Description

Our society's rapid algorithmification is fueled by data, but the reliance on data raises important questions. What are the latent biases hidden in the collected data? If that data was used to train machine learning algorithms, how did these biases impact predictions made by algorithms and systems that depend on them? Are the algorithmic decisions fair, or do they perpetuate stereotypes and fortify discrimination? As we come to rely on AI to make decisions in our lives and allow for synergistic relationship with technology, we need to build trust in AI by improving algorithmic fairness, accountability, transparency and explainability.

The course will explore topics in the intersection of data, networks and algorithms with fairness and bias through quantitative analysis and hands on exploration.

Learning Objectives

Students will be introduced to a wide array of methods from disciplines ranging from mathematics to the social sciences, including graph theory, linear algebra, statistics, and machine learning. The course will survey recent research papers to examine how researchers apply these methods to large-scale social data to understand how biases in data and algorithms affect analysis of social data and the decisions of algorithms trained on this data.

Prerequisite(s): none
Co-Requisite (s): none
Concurrent Enrollment: none
Recommended Preparation: statistics, AI and/or machine learning, e.g., CSCI 561; knowledge of at least one programming language (Java, C++, Pythod)

Course Notes

The course will be run as a lecture class with student participation strongly encouraged. There are weekly readings and students are encouraged to do the readings prior to the discussion in class. All of the course materials, including the readings, lecture slides, homeworks will be posted online on Blackboard. The class project is a significant aspect of this course and at the end of the semester, students will present their projects in class.

Technological Proficiency and Hardware/Software Required

A basic understanding of programming that will allow you to manipulate data and implement basic algorithms, using any programming language, is required. Python is recommended, as it will be the “official” programming language of the class. We will use IPython Notebook as the environment to demonstrate algorithms and perform the analysis. Introductory statistics course or equivalent will help, and so will familiarity with linear algebra.

Required Readings and Supplementary Materials

Students will be given reading materials, such as research papers or online textbooks. Students are responsible for all assigned reading assignments. The reading material is based on recently published technical papers available via the ACM/IEEE/Springer digital libraries. All USC students have automatic access to these digital archives. Lecture slides will be placed on Blackboard and will be accessible to students before each lecture.

Recommended Python "cookbooks":

- Python Data Visualization Cookbook —by Igor Milovanović (ebook: $14)
- Learning IPython for Interactive Computing and Data Visualization —by Cyrille Rossant (ebook: $10)
- Learning scikit-learn: Machine Learning in Python —by Raúl Garreta and Guillermo Moncecchi (ebook: $10)

Description and Assessment of Assignments

Homework Assignments

There will be two homework assignments designed to give student basic proficiency with collecting, manipulating and analyzing social data

- **HW1: Research ethics** – This homework will expose the students to rules for responsible and ethical conduct in research. This will require completion of the "Responsibility in Research" module of the CITI training – a training required by Institutional Review Boards.
- **HW2: Hypothesis testing in social data** - This homework will introduce students to social media data analysis. Students will learn to download data and perform basic analysis to quantitatively investigate a scientific hypothesis.
- **HW3: Social network analysis** – Students will download networks in different but common network formats and conduct basic network analyses, such as identifying important individuals within a network, comparing centrality scores, identifying who the significant individuals are with respect to a given individual.
- **HW4**: **Text mining** – This will consist of several small exercises where students will write code to implement and test text mining algorithms. Particular attention will be paid to how algorithms can be "gamed" to yield biased results.

Course Project

An integral part of this course is the course project, which builds on the topics and techniques covered in the class, focusing on extending and evaluating methods to solve problems. Students will write a written proposal for the project, conduct the project, and then write a paper about the project, and present the project in class. Students are encouraged to identify a new problem, apply or extend the methods they learned in class to propose an approach to solve the problem. Emphasis is placed on quantitative evaluation of the approach. Working as a group is permitted if the project is large enough to justify this. A team can consist of no more than 3 persons.

Project Timeline:

- Aug 26 – Sep 29: Identifying team members and project topics
- Sep 29: Proposal due (team member, topics and milestone)
- Nov 3: Mid-term report due (data description, preliminary results)
- Dec 4: Project presentations (open to all faculty and students)
- Dec 6: Final report due (task and model description, major discovery, lessons learned)
-

Sample project:

"Twitter sentiment as a potential proxy for opinion polls:" the goal of the project is to explore whether Twitter data can reveal public opinions about controversial topics that are similar to results of opinion polls.

Students can easily find resources available online, including twitter API and sentiment analysis tools. A project of this size usually consists of 2 persons. The team will work together on collecting the twitter data, examining the preliminary results, identifying one challenge in current sentiment analysis application, and providing a reasonable solution.

Grading breakdown of the course project:

- Proposal: Not Graded
- Mid-term report: 10%
- Final report: 25%
  Reports are 5 pages long, describing the goal, existing solutions to the problem and challenges, proposed approach, its evaluation and limitations.
- Presentation: 15%
  Presentations are 15-20 minutes long, depending on the number of projects.

Grading & Policy

Class participation and engagement are essential ingredients for success in your academic career, therefore during class turn off cell phones and ringers (no vibrate mode), laptops and tablets. The only exception to use laptops during class is to take notes. In this case, please sit in the front rows of the classroom: no email, social media, games, or other distractions will be accepted. Students will be expected to do all readings and assignments, and to attend all meetings unless excused, in writing, at least 24 hours prior. This is the (tentative) system that will be employed for grading:

Quizzes: There will be weekly quizzes based on the material from the week before. While there are 11 quizzes, only 10 best scores are used for grade calculation. There is no mid-term or final for this class.
Homework: There will be two homework assignments.
Project: Each student will team up with a classmate (max of 3 allowed in some cases) to do an independent project based on the topics covered in the class. Students will propose a novel project, do the research and build a proof-of-concept, write a report about the work, and present the work in class. A serious final paper will be expected. The report will be at least 2,500 words and will include appropriate figures and tables. The work should cover the following points:

1) Statement of the problem & Why the problem is important.
2) How the problem was faced —including a description of methodology and dataset(s);
3) Discussion of results, findings, and limitations of the study.
4)Related literature & Final remarks/conclusions.

Grading rubric: Projects will be graded on novelty, technical soundness, and the quality of evaluation. Reports and presentations will be graded according to the project grading rubric and the quality and clarity of presentation.
Class Participation: students are expected to attend every class and actively participate in the discussion

Assignment	Points	% of Grade
Homework	40	40
Class Participation	10	10
Project Proposal	0	0
Midterm Report	10	10
Project Report	25	25
Project Presentation	15	15

Assignment Submission Policy

Assignments are due at 11:59pm on the due date and should be submitted in Blackboard. You can submit assignments up to one week late, but you will lose 20% of the possible points for the assignment. After one week, the assignment cannot be submitted.

Topics Covered (to be removed after weekly lectures are defined)

- Data
  - Introduction & Ethical Data collection
    - IRB, copyright, privacy
    - Representativeness
    - Sampling
    - Specializing in Social Media (Olteanu survey)
  - Sampling issues
    - Representativeness
    - Limits on predictability and results
  - Modeling & Prediction
    - Features selection & Dimensionality reduction
    - Regression, decision trees
    - Issues: correlated features, correlated samples
    - Algorithmic decision making
      - Imputation
    - Other issues: class balance,
  - Bias in Data
    - Confounding
    - Simpson's paradox
      - *Hands on: Social data exploration, Including Simpson's paradox*
- Networks
  - Basics of networks
    - Degree, assortativity, clustering, homophily
    - Ranking, bias in ranking
  - Applications
    - Lecture 1
    - Searchability & bias (not every forwards the letter)
      - Social ties
    - Lecture 2
      - Communities (fairness example)
      - Visibility of minorities
  - Prediction in networks
    - Lecture 1:
      - Link prediction
    - Homophily, echo chambers, perceptions
    - Diffusion/Node prediction/Label propagation
  - Bias in Networks

        - Friendship paradox
        - *Hands-on: Twitter mining*
- Algorithms
    - Text mining & Sentiment analysis
    - Fair representations
        - Word embeddings
            - How word embeddings are biased by text
            - Tracing biased words to the data
        - (Maybe) Image processing
            - Digital camera example
            - Bias in captioning
    - Privacy & Fraud
        - Fraud & Manipulation (ebay, reviews)
        - Privacy in networks
            - Inferring user psychological states
            - Inferring user private attributes from friends
    - Algorithmic Bias
        - Cognitive biases, Position bias etc.
        - Biases (Ricardo BY CACM article)
        - Crowdsourcing
        - *Hands on: Gaming Sentiment analysis on Twitter, Hands on: Building a "fair" word embedding*
- Fairness
    - Definitions & fair machine learning
        - What is fairness? (Kleinberg)
        - Prediction with sensitive features
        - *Hands on: Xintao Wu's data + algorithms*
    - Case studies of discrimination
        - "Measuring discrimination in alg decision making"
        - COMPAS
    - Diagnosing fairness
        - Measures, methods, black box approaches
        - Dataset nutrition label
        - What are the solutions?
        - *Hands on: Investigating black box approaches to fairness*

Course Schedule: A Weekly Breakdown

Statement on Academic Conduct and Support Systems

Academic Conduct

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Section 11, *Behavior Violating University Standards* https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, http://policy.usc.edu/scientific-misconduct.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the *Office of Equity and Diversity* http://equity.usc.edu or to the *Department of Public Safety* http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us. This is important for the safety of the whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. *The Center for Women and Men* http://www.usc.edu/student-affairs/cwm/ provides 24/7 confidential support, and the sexual assault resource center webpage http://sarc.usc.edu describes reporting options and other resources.

Support Systems

A number of USC's schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the *American Language Institute* http://dornsife.usc.edu/ali, which sponsors courses and workshops specifically for international graduate students. *The Office of Disability Services and Programs* http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, *USC Emergency Information* *http://emergency.usc.edu* will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.