Week 5 - 6 Readings

Big Data's Disparate Impact

Barocas et al (2016)

How data mining discriminates

- **Defining target variable and class labels**
  Real-world problem must be parsed
  Sometimes target vars are pre-defined, other times they must be determined
- **Training data**
  If model is trained on prejudiced data, model will provide produced results
  Marginalized groups over-/under-represented
  Data must be pre-labeled in some cases
- **Feature selection**
  Must decide which features included in analysis
- **Proxies**
  Class membership can be determined using seemingly unrelated features
- **Masking**
  Forms of unintentional discrimination can be used intentionally
  Proxy variables used to determine a person's class membership

Title VII liability

Title VII legislation doesn't prevent data mining discrimination
If variables are relevant to the job then model is ok, but redundant encodings are still likely
Therefore data mining is good at both predicting job performance *and* reproducing bias
Which features are ok to use? Not ok?
Plaintiff's responsibility to find alternative to in a lawsuit

False Positives, False Negatives, and False Analyses
Flores et al

Written in response to an article by Angwin, et al (2016) titled “There is software that is used across the country to predict future criminals. And it is biased against blacks.”

This ProPublica study found that the COMPAS tool used to predict recidivism was inaccurate: of predicted future violent offenders, only 20% actually committed violent crime
Also found COMPAS more likely to incorrectly label a black person as future reoffender than white person

Racial disparities in American prison population are huge
Actuarial risk assessment instruments (ARAIs) have the potential to make non-biased sentencing decisions

Criticisms of Angwim et al

Used COMPAS on pretrial defendants, intended for use post-disposition
Didn’t use accepted standards to test for racial bias
COMPAS not intended for Y/N prediction, instead assigns people to low, med, and high risk categories

Flores et al conducted their own test of the COMPAS tool and found that it did not display and significant bias

A Human-Centered Approach to Algorithmic Services

Lee et al (2017)

412 Food Rescue is an organization that takes leftover food from donors and redistributes it to recipient organizations

Currently, a human manager makes all food allocation decisions, but the org is looking to adopt an algorithm

There are many **stakeholders** in 412 Food Rescue, and all have different ideas regarding “fair” food allocation.

- The **food manager** makes allocation decisions using efficiency and equity, believes alg would be fairer
- The **volunteers** mostly agrees with an equity model, but also thought alg could help with efficiency
- Most **donors** agreed with efficiency model; they want donations to stay local
- The **recipient organizations** believed in an equity model, although they admitted ranking based on need is nearly impossible
- The **clients** believed in an equity model
- **Everyday citizens** liked equity models, but had diverse ideas about how to measure need

Stakeholders have a wide set of ideas about what a fair algorithm would look like

Many believe in equity, but nobody could agree on how to rank orgs based on need

A potential alg could allow each stakeholder to set their own ideas of fairness, eg donors choose where food goes, volunteers choose which deliveries they make

Previous research has shown transparency in algs is crucial

People still need social interaction, donors especially

Algorithmic Bias in Autonomous Systems

Danks et al (2017)

Bias == any deviation from a standard
Not all biases are problematic, objectivity is not always good
Bias doesn't have a purely technological solution

A Taxonomy of Algorithmic Bias

- **Training data bias**
  If input is biased, model will be biased
  Sometimes, using manufactured biased training data can help create non-prejudiced model
- **Algorithmic focus bias**
  Deliberately not using certain features in model due to legal, moral standards
- **Transfer context bias**
  Using a model outside its intended context
- **Interpretation bias**
  Misinterpretation of algorithm's output by the user

Identifying Problematic Bias

Is bias ethically desirable, or does it need to be fixed?
The context of the model needs to be understood
May be an ethical obligation to use a certain bias
An alg may be biased or not depending on the viewpoint

Intervening on Problematic Bias

Redefine scope in which model is to be used
Algorithmic bias can be used to eliminate other biases
Nobody can objectively determine what is biased, which features are ok to use in a model

Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers

Lee et al (2015)

Algorithmic management used for drivers at Uber and Lyft: increases efficiency, less need for human managers on-site
Drivers get automatically assigned passengers, can either accept or reject
Drivers frustrated when receiving an assignment that not economical
Uber doesn’t disclose how it assigns passengers, drivers with more knowledge can game the system and get more economical rides
Drivers punished for rejecting rides, regardless of reason

Surge pricing model meant to incentivize drivers, mostly fails

New drivers anxious about their rating
Passengers’ rating isn’t always accurate

Drivers used online forums to share knowledge, help new drivers, vent

Conclusion

Drivers would trust Uber more if they made algorithms more transparent
Algorithms don’t allow for leniency with regard to driver ratings, acceptance rate
Many drivers had legitimate reasons for rejecting a ride, low rating from passenger
Algorithms, especially surge model, do not take into account the way people work and diverse motivations for driving for Uber

Fairness in Criminal Justice Risk Assessments: The State of the Art

Talks about 6 definitions of fairness and their incompatibility with each other and with accuracy

- **Overall accuracy equality**: seeks to equalize accuracy of both positives and negatives across groups.
- **Statistical Parity**: marginal distributions of predicted classes are the same.
- **Conditional procedure accuracy equality**: false positive and false negative rates are equivalent across groups
- **Conditional use accuracy equality**: probability of success and probability of failure is same across groups
- **Treatment equality**: ratio of false negatives and false positives is the same across groups
- **Total Fairness**: all of the above being achieved at once.

The Ambient Population and Crime Analysis

Defines two types of population:

- Census population: the number of people with residence in an area based on census data
- Ambient population: the estimated number of people actually in an area at any given time. For example, downtown will have a much higher ambient population during the work day than the census population would suggest.

Suggests using ambient population for more accurate population based predictions.

Fair, transparent and accountable algorithmic decision-making processes

Sources of discrimination:

- Poorly weighted input data
- The choice to use an algorithm at all
- Choosing the wrong model for a given context
- Biased training/testing data creating a negative feedback loop

Types of opacity:

- Intentional
- Illiterate
- Inherent

Survey of Clustering Algorithms

Traces the steps of properly utilizing a clustering algorithm
Mentions the importance of distance metrics/similarity metric and explains them in context
Algorithm types:

- Hierarchal
- Squared-error based (centroid and mediod)
- PDF estimation (i.e. Gaussian Mixture Model)
- Graph-theory based
- Combinatorial Search Techniques
- Fuzzy Clustering
- Neural Networks-Based
- Sequential
- Data-reduction for high dimensional data

Ultimate point is that no one clustering method is optimal for all situations

Battling Algorithmic Bias

By Keith Kirkpatrick (October 2018)

- They often inadvertently pick up the human biases that are incorporated when the algorithm is programmed
- Example was the use of risk scores used by criminal justice system
- Need to that identifying and eliminating cases of both explicit discrimination and implicit discrimination
- Solution:
    - Monitor algorithms to determine fairness (which is not clear cut)
    - the creation of a federal regulator that would oversee certain algorithms in an effort to help prevent unfairness or discrimination

Towards a Feminist HCI Methodology: Social Science, Feminism, and HCI
by Shaowen Bardzell and Jeffrey Bardzell

- HCI is increasingly engaging with matters of social change that go beyond the immediate qualities of interaction.
- The potential for feminist social science to contribute to and potentially benefit from HCI’s rising interest in social change.
- Some of the feminist HCI methodology
    - Connection to feminist theory
    - A simultaneous commitment to scientific and moral objectivities
- Explores the dynamic relations between values and different formulations of objectivity in the wake of post-positivist philosophy of science, stressing feminist contributions to these debates and arguing for their applicability to HCI today.

The accuracy, fairness, limits of predicting recidivism

by Julia Dressel and Hany Farid(2018)

- Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) vs random participants
- Human Assessment
  - In conclusion, there is no sufficient evidence to suggest that including race has a significant impact on overall accuracy or fairness. The exclusion of race does not necessarily lead to the elimination of racial disparities in human recidivism prediction.
- Algorithmic Assessment
  - In conclusion, the COMPAS is using nothing more sophisticated than a linear predictor or its equivalent. A classifier based only two features features—age and total number of previous convictions—performs as well as COMPAS
- When considering using software such as COMPAS in making decisions that will significantly affect the lives and well-being of criminal defendants, it is valuable to ask whether we would put these decisions in the hands of random people who respond to an online survey because, in the end, the results from these two approaches appear to be indistinguishable.

Why so many clustering algorithms – A Position Paper
by Vladimir Estivill-Castro

- A good clustering has biases because of the background of researchers and because of applications
- Some problems derived from the lack of explicit discussion of models and induction principles when presenting algorithm. It intends to illustrate that diversity is necessary and useful; but has lead to some confusion about the properties of algorithms. The effect of this lack of clarity and detail is a difficulty to compare algorithms.
- Some Recommendations:
    - Clusters are in the eye of the beholder
    - An algorithm designed for some universe of models has no chance if the data sets has a structure that is actually representable by a radically different family of models (*k*-Means can not find non-convex clusters)

Turkers, Scholars, “Arafat” and “Peace”: Cultural Communities and Algorithmic Gold Standards
by Shilad Sen and etc.

- RQ1: Do different cultural communities produce different gold standards?
    - RQ1a studies whether a gold standard dataset must be matched to the audience of the system it serves.
    - RQ1b studies shows some cultural communities will need more contributors than others to obtain a desired level of gold standard sample error
    - These finding have direct implications for practitioners and researchers collecting domain-specific gold standards
- RQ2: Do algorithms perform differently on gold standards from different cultural communities?
    - That broad concepts account for the most egregious scholar-expert errors
    - That algorithms typically under-predict the relatedness of broad concepts for scholar experts
    - These finding support for the algorithmic “communities matters”(AMT crowd workers and scholars) precept in contexts when domain expertise is relevant

Accountability in Algorithmic Decision Making
by Nicholas Diakopoulos

- It is time to think seriously about how the algorithmically informed decisions now driving large swaths of society should be accountable to the public
- Algorithmic Decision Making
    - Prioritizing
    - Classification
    - Association
    - Filtering
- Algorithmic Transparency Standard
    - Human Involvement
    - Data
    - The model
    - Inferencing
    - Algorithmic presence
- Work on machine learning and data-mining solutions that directly take into account provisions for fairness and anti-discrimination

Critical Questions from Big Data

- Big data “changes” definition of knowledge
- Claims to objectivity and accuracy are misleading
- Big data are not always better data
- Taken out of context, big data loses meaning
- Just because it is accessible does not make it ethical
- Limited access to big data creates new digital divides

Big Questions for Social Media Big Data

Methadological

- Only a few platforms are used overlooking the wider social ecology
- Their structural bias are sometimes ignored
- Denominator problem
- Selecting on dependent variable without requisite precaution

Inference

- User behavior sometime uninterpretable
- People alter behavior dependent upon if they think they are being observed

Investigating User's Understanding of Invisibl Algorithms and Designing Around It

- Proprietary culture prevents user transparency
- Users often lack understanding of algorithms process
- Facebook case study
- Algorithm aware design
- Actionable Transparency

Interacting Meaningfully with Machine Learning Systems: Three Experiments

- 1. Explored explanation paradigms (Rule – Based, Keyword – Based, Similarity –Based)
- 2. Users feedback
- 3. User implementation

- Concluded that design approach should favor a more informed, “hands-on” user

Interacting Meaningfully with Machine Learning Systems: Three Experiments

- 1. Explored explanation paradigms (Rule – Based, Keyword – Based, Similarity –Based)
- 2. Users feedback
- 3. User implementation

- Concluded that design approach should favor a more informed, “hands-on” user