Bias and Fairness in Learning Systems

CS 7792 - Spring 2020

Thorsten Joachims

Department of Computer Science & Department of Information Science
Cornell University
Outline of Today

• Introduction
  – Thorsten Joachims
• Motivation of Class Topics
• Administrivia
  – Goals for the Class
  – Pre-Requisites
  – Credit Options and Format
  – Course Material
  – Contact Info
Bias and Fairness in ML

Google hit with record EU fine over Shopping service
By Leo Kelion
Technology desk editor

Machine Bias
There’s software used across the country to predict future criminals. And it’s biased against blacks.
by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

Wenn Maschinen kalt entscheiden
Algorithmen urteilen über Arbeitslose, Straftäter und Job-Bewerber. Eine Kommission der Bundesregierung will ihre Macht nun bändigen. Kann das gelingen?
Eine Analyse von Ann-Kathrin Nezik
ZEIT ONLINE

Amazon scraps secret AI recruiting tool that showed bias against women
Jeffrey Daskin
8 MIN READ
Two Motivating Examples

Example 1: Rankings in Recommender Systems

Example 2: College Admission
Ranking in Online Systems

Ranking function $\pi$ that ranks items for context $x$. 
→ Learning-to-Rank
Two-Sided Market

Online Retail

- Utility to Users: Customers find products they want
- Utility to Items: Sellers get revenue
Two-Sided Market

Music Streaming

• Utility to Users:
  Customers find music they enjoy

• Utility to Items:
  Artists get streaming revenue
Two-Sided Market

News

• Utility to Users: Readers find relevant articles

• Utility to Items: Writers get their voice out (and ad revenue)
Are Conventional Methods Fair?

Probability Ranking Principle:
• Rank documents by probability of relevance \( y^* \) [Robertson, 1977]
• For virtually any measure \( U \) of ranking quality
  \[ y^* = \arg\max_y [U(y|x)] \]
• Are rankings fair/desirable?

<table>
<thead>
<tr>
<th>Rank</th>
<th>Item</th>
<th>P(read)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Times 1</td>
<td>50.99</td>
</tr>
<tr>
<td>2</td>
<td>Times 2</td>
<td>50.98</td>
</tr>
<tr>
<td>3</td>
<td>Times 3</td>
<td>50.97</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>100</td>
<td>Post 1</td>
<td>49.99</td>
</tr>
<tr>
<td>101</td>
<td>Post 2</td>
<td>49.98</td>
</tr>
<tr>
<td>102</td>
<td>Post 3</td>
<td>49.97</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Fairness in Ranking

Endogenous Factors
How to allocate exposure based on merit in order to
• Satisfy legal requirements
• Shape marketplace dynamics (e.g. Spotify, superstar economics)
• Spam, Polarization

Exogenous Factors
How to estimate merit without biases like
• Position bias
• Trust bias
• Uncertainty bias
• Historical actions
• Stereotypes
Two Motivating Examples

Example 1:
Rankings in Recommender Systems

Example 2:
College Admission
College Admission

Problem

– 45,000 applications $\rightarrow$ 6,000 admissions
– Manual reading and evaluation by humans
$\rightarrow$ Noise, variability by reviewer, cost

How can this be (partially) automated to

– Make better decisions
– Reduce cost
Some Strawmen Proposals

Why not do the following?
• Sort students by high-school GPA.
• Sort students by SAT score.
• Sort students by SAT score for each legally protected class (race, color, religion or creed, national origin or ancestry, sex, age, disability, veteran status, genetic information, citizenship).
Broader Picture

Balance cost/benefit to
• Rejected students
• Accepted students
• Institution
• Society

How do we design a machine learning algorithm that can optimize a complex combination of goals?
Overall Goals for this Class

• Deeply explore bias and fairness research in ML.
  → Narrow focus.
  → Mostly original research papers.

• Practice being a successful academic.

→ Class targeted towards current PhD students with research interests in this area!
Pre-Requisites

- This is not an introductory Machine Learning class!
- You need to satisfy one of the following ML pre-reqs:
  - Successfully taken CS4780 “Machine Learning”
  - Successfully taken CS6780 “Advanced Machine Learning”
  - Successfully taken a comparable “Intro to ML” class (*)
  - Acquired the equivalent ML knowledge in some other way (e.g. strong background in Statistics + ML textbook) (*)
- You need to be a PhD student
- Currently doing or planning to do research in this area of ML
- Basic probability, basic statistics, general mathematical maturity

(*) means talk to me
Format of Class

• Lectures (by TJ)
  – Background material

• Research paper presentations (by students)
  – Explore current state of the art

• Peer reviewing
Research Paper Presentations

• Students present the paper in class
  – Slide presentation
  – Prepare discussion topics / group activity
  – Create critique, extended bibliography, examples, demo software, experiments etc. that help understand the paper
  – Prepare quiz

• Everybody reads the paper in preparation for class
  – Quiz about each paper

• All students give feedback afterwards.
Peer Reviewing

• Goals
  – Give presenter constructive feedback from audience.
  – Reviewer has to think through what works about a presentation.
  – Learn how to write reviews. Be constructive, respectful, and mindful of biases.

• Reviewing the reviewers
  – Presenter gets to give feedback on the reviews (both direct and confidential to me)
Credit Options and Grades

- **Pass/Fail**: Need to get at least 50% of points on each of following to pass.
  - paper presentation
  - in-class quizzes (lowest grades replaced by second lowest grade)
  - peer reviewing (lowest grades replaced by second lowest grade)
  - in-class participation
- **Letter grade**: not allowed
- **Audit**: not allowed, unless you have very good arguments
Course Material

• Reference Books
  – Barocas, Hardt, Narayanan. "Fairness and Machine Learning". (online)

• Background Reading
  – Shai Shalev-Shwartz, Shai Ben-David, "Understanding Machine Learning - From Theory to Algorithms", Cambridge University Press, 2014. (online)

• Slides, Notes and Papers
  – Slides available on course homepage or CMT
  – Papers on course homepage
Bidding on Dates to Present

• Use CMT bidding mechanism to reserve dates to presenters
  – If you are
    • enrolled via studentcenter,
    • filled out the paper sheet (no promise we still have space though)
      you will get email from me through CMT.
  – Place your bids on the dates by Friday night.
  – I’ll send you your assignment next week.

• Two weeks before your date
  – You get in touch with me and we finalize what you will be presenting.
  – I’ll have suggestions for papers, but I am happy to follow your suggestions as well.
How to Get in Touch

• Course Web Page
  – https://www.cs.cornell.edu/Courses/cs7792/2020sp/

• Email
  – Thorsten Joachims: tj@cs.cornell.edu

• Office Hours
  – Wednesdays 11:10pm – 12:00pm, 418 Gates Hall

• Piazza
  – https://piazza.com/cornell/spring2020/cs7792

• Peer reviewing platform
  – https://cmt3.research.microsoft.com/CS77922020